Analysis and Modifications of DNA or RNA Sequences through the

TRICHROMATIC THEORY OF EQUILIBRIUM OF SYSTEMS

THE MYSTERIES OF DNA (First Part)

Relations between SARS-CoV-2, BAT, PANGOLIN, INFLUENZA A VIRUS, HIV-1 and the PARANTHRACIS and ANTHRACIS BACILLI

Author: ***Nunzio Bonaventura***
IT technician: ***Vincenzo Viggiano***

13 April 2021

INTRODUCTORY
INDEX

Abstract ..... page 5

Preamble ..... page 6

Reading Guide ..... page 8

Introduction ..... page 12

Reading Guide

CHAPTER I°: ***Basic Concepts of Molecular Biology***
Goals of Chapter I° ..... page 17
1.1 Molecular biology, amino acids and proteins ..... page 18

CHAPTER II°: ***Basic Procedure for Carrying Out a Blast Research and Its Main Goal***
Goals of Chapter II° ..... page 26
2.1 How to perform a basic blast research ..... page 27
2.2 How to use the results of a basic blast research ..... page 40

CHAPTER III°: ***From Bats to Human Beings through Pangolins: the RdRp Gene and the RNA Sequence "GTGAAATGGTCATGTGTGGCGG"***
Goals of Chapter III° ..... page 46
3.1 The most accredited hypothesis: from bats to human beings through pangolins ..... page 47
3.2 The RdRp Gene and the RNA sequence "*GTGAAATGGTCATGTGTGGCGG*" of SARS-Cov-2 ..... page 49

CHAPTER IV°: ***BLAST Research on the RNA Sequence "GTGAAATGGTCATGTGTGGCGG" of the RdRp Gene (RdRp_SARSr-F) (WH-Human_1\|China\|2019-Dec)***
Goals of Chapter IV° ..... page 56
4.1 BLAST research on the RNA sequence "*GTGAAATGGTCATGTGTGGCGG*" of the RdRp Gene (RdRp_SARSr-F) – (Hitlist size: 100) ..... page 57

CHAPTER V°: ***BLAST Research on two RNA Specific Sequences of the RdRp Gene***
Goals of Chapter V° ..... page 75
5.1 BLAST research on the RNA sequence

"*GTGAAATGGT**T**ATGTGTGGCGG*" of the RdRp Gene of the "Bat coronavirus BM48-31/BGR/2008, complete genome" – (Hitlist size: 100) ..... page 76
5.2 BLAST research on the RNA sequence "*GTGA**G**ATGGTCATGTGTGGCGG*" of the RdRp Gene of the "Bat SARS-like coronavirus isolate bat-SL-CoVZC45, complete genome" – (Hitlist size: 100) ..... page 95
5.3 BLAST research on the RNA sequence "*GTGA**G**ATGGTCATGTGTGGCGG*" of the RdRp Gene of the "Bat SARS-like coronavirus isolate bat-SL-CoVZC45, complete genome" – (Hitlist size: 1000) ..... page 122

CHAPTER VI°: ***Base Substitutions in the RNA Sequence "GTGAAATGGTCATGTGTGGCGG" of the RdRp Gene (RdRp_SARSr-F) and Subsequent BLAST Research***
Goals of Chapter VI° ..... page 126
6.1 BLAST research on the modified RNA sequence "*GTGAAATGGT**A**ATGTGTGGCGG*" of the RdRp Gene (Hitlist size 1000) page 127
6.2 BLAST research on the modified RNA sequence "*GTGAAATGGT**G**ATGTGTGGCGG*" of the RdRp Gene (Hitlist size 1000) page 129
6.3 BLAST research on the modified RNA sequence "*GTGA**T**ATGGTGATGTGTGGCGG*" of the RdRp Gene (Hitlist size 1000) page 131
6.4 BLAST research on the modified RNA sequence "*GTGA**C**ATGGTGATGTGTGGCGG*" of the RdRp Gene (Hitlist size 1000) page 132

CHAPTER VII°: ***Procedure for the Modification of the Original Sequence and the Generation of New Sequence***
Goals of Chapter VII° ..... page 135
7.1 Partial modification of an original sequence and generation of new sequences ..... page 136
7.2 The twenty-four "Non-Obvious Trends" ..... page 138
7.3 Procedure for partial modification of the original sequence ..... page 139

CHAPTER VIII°: ***Modifications of the RNA Sequence "GTGAAATGGTCATGTGTGGCGG" of the RdRp Gene (RdRp_SARSr-F) and Generation of New Sequences***
Goals of Chapter VIII° ..... page 149
8.1 Modifications of the RNA sequence "*GTGAAATGGTCATGTGTGGCGG*" of the RdRp Gene (RdRp_SARSr-F) and Generation of New Sequences ..... page 150

CHAPTER IX°: ***Blast Research and Discussion of the Results Obtained of the Significant Alignments of Sequence n° 1/3***
Goals of Chapter IX° ..... page 158
9.1 Sequence n° 1/3 (T=0; C=1; G=2; A=3) – (Hitlist size: 100) ..... page 160

9.2 Sequence n° 1/3 (T=0; C=1; G=2; A=3) – (Hitlist size: 1000) ..... page 167
9.3 The segment of Sequence n° 1/3
identified as corresponding to HIV-1 Virus ..... page 176
9.4 Prediction of the constitutive codons of the segment of the
Sequence n° 1/3 and discussion on their hypothetical products ..... page 183
9.5 Sequence of amino acids "**M A M T H**" ..... page 184
9.6 The modification of the sequence of amino acids
"**M A M T H**" in the sequence of amino acids "**M V M T H**" ..... page 190
9.7 "Audentes fortuna iuvat": the sequence "**ATGGTAATGACACATT**" ..... page 198
9.8 From the sequence of amino acids
"**M V M T H**" to the sequence of amino acids "**M V M T HF**" ..... page 220
9.9 Sequence of amino acids "**M A M T H K**" ..... page 227
9.10 Sequence of amino acids "**M A M T H K S Q**" ..... page 240
9.11 Sequence of amino acids "**M A M T H K S Q M K**" ..... page 246
9.12 Sequence of amino acids "**M A M T H K S Q M K PRQMAP**" ..... page 257

CHAPTER X°: ***Blast Research and Discussion of the Results Obtained of the Significant Alignments of Sequence n° 3/1***

Goals of Chapter X° ..... page 267
10.1 Sequence n° 3/1 (T=2; C=0; G=3; A=1) – (Hitlist size: 100) ..... page 270
10.2 The principal "products" identified through
BLAST research carried out on Sequence n° 3/1 – (Hitlist size: 100)
and the fundamental role played by the Serine Hydrolase ..... page 396
10.3 Sequence n° 3/1 (T=2; C=0; G=3; A=1) – (Hitlist size: 5000) ..... page 399
10.4 The principal "produtcs" identified through
BLAST researches carried out on Sequence n° 3/1 (Hitlist size: 5000) ..... pag. 433

CHAPTER XI°: ***Blast Research and Discussion of the Results Obtained of the Significant Alignments of Sequence n° 8/1***

Goals of Chapter XI° ..... page 437
11.1 Sequence n° 8/1 (T=1; C=0; G=3; A=2) – (Hitlist size: 100) ..... page 439
11.2 Sequence n° 8/1 (T=1; C=0; G=3; A=2) – (Hitlist size: 1000) ..... page 446

CHAPTER XII°: ***Blast Researches of other Modified Sequences of the RNA Sequence "GTGAAATGGTCATGTGTGGCGG" of the RdRp Gene (RdRp_SARSr-F)***

Goals of Chapter XII° ..... page 465
12.1 Sequences n° : 1/2, 2/2, 3/2, 3/3, 3/4, 3/5,
3/6, 4/1, 5/2, 6/1, 6/2, 7/1, 7/2, 9/1, and 10/1 ..... page 467

CONCLUSIONS

Final consideration ..... page 633

ABSTRACT

In this work, a sequence of 22 RNA bases of the *Rdrp gene* of *SARS-Cov-2* was modified through the Systems Equilibrium Trichromatic Theory and its computer program. This sequence is one of the RNA sequences used as positive control of rhinopharyngeal swab, along with other *SARS-Cov-2* gene sequences. The links between the **RNA** of ***SARS-Cov-2*** and the genome of two animals, the ***Bat*** and the ***Pangolin***, have been partially identified and clarified, among all the possible links identified by the research carried out on the DNA and RNA database, both considered implicated in the Covid-19 infection. However, BLAST research also reveals unexpected significant alignments (involving not only bases, but in some cases whole amino acids) corresponding to a sequence with new characteristics (*misc_feature*) ***Human Immunodeficiency Virus*** (***HIV-1***), several sequences of ***Influenza A virus*** and, above all, several sequences of ***Bacilli Parantrace*** (***Bacillus Paranthracis***) and ***Anthrax*** (***Bacillus Anthracis***), the latter a gram-bacterium positive infectious agent, the causative agent of ***anthrax in animals*** and ***anthrax in humans*** (the latter being considered in the highest priority category of infectious agents associated with a potential risk of being used as a biological weapon). To make possible the replication of the results obtained, a part of this document was dedicated to teaching the procedure used for the modification of DNA or RNA sequences. SARS-Rdrp gene RNA sequence modificationsCov-2 and the methodology proposed in this document open up different perspectives for initiating targeted studies and innovative treatments of the Covid 19 pandemic and for a more functional reorganization of DNA and RNA databases.

PREAMBLE

About five months ago I started writing this document and I never thought I would embark myself on an enterprise of this magnitude, even less about completing it, in such a short time. Unfortunately, the pandemic forces me to publish a paper, unlike what I am used to doing, partly incomplete, despite its lenght, justified by the need to copy all the most important original data into the document and make them immediately available. I apologize for these limitations, but I am sure you will understand why. However, to make the reading easier and more understandable, at the beginning of the chapters I have summarized the topics covered within them and what the purpose of their discussion is. This work, while using the same methodology as my previous studies on the analysis and modification of DNA sequences (studies published on the website www.ttesystems.eu/applicazioni.php, in https://www.linkedin.com/in/dr-nunzio-bonaventura-41723660/ e in https://www.facebook.com/Dott.NunzioBonaventura), assumes a totally autonomous character. My initial intent was to modify some SARS-CoV-2 RNA sequences, by availing from my own Theory of Systems (www.ttesystems.eu) and its computer program. The ultimate goal should have been to better identify and clarify, among all the possible links identified by the searches carried out in the DNA and RNA database, the links between the ***SARS-CoV-2 RNA*** and the genome of two animals, the ***Bat*** and the ***Pangolin***, both considered implicated in the Covid-19 infection. This goal was largely achieved, but to my extreme surprise, during the modification of a RNA sequence of a SARS-CoV-2 Gene, I came across a significant alignment corresponding to a sequence with new characteristics (*misc_feature*) of the ***Human Immunodeficiency Virus (HIV-1)*** and others corresponding to different sequences of the ***Influenza A Virus*** and, especially, ***Parantrax Bacillus (Bacillus Paranthracis)*** and ***Anthrax (Bacillus Anthracis)***, that later an infectious gram-positive bacterium, the causative agent of ***anthrax in animals*** and ***anthrax in humans*** (https://it.wikipedia.org/wiki/Bacillus_anthracis).

To avoid any kind of misunderstanding immediately, **this work does not state that the *HIV-1 Virus* or the *Parantrax and Anthrax Bacillus* are totally present in the SARS-CoV-2 RNA sequence. Rather, it is noted that the modifications of this SARS-CoV-2 RNA sequence, through the methodology that will be proposed and taught here, allow to generate (strictly respecting the "Non-Obvious Trends" of the SARS-CoV-2 RNA sequence) some new sequences which (**in the searches carried out in **the BLAST database) identify, by very high similarity (similarity not only of the individual bases, but in some cases also of some amino acids), a sequence of the *HIV-1 Virus*** (for further information see the Chapter IX) and **various sequences of the *Anthrax Bacillus*** (for further information see the Chapter X). It is also possible to perform the **reverse procedure**, that is, by modifying the **new sequences generated**, to identify, by very high similarity, **the original sequence**, the previously mentioned **SARS-CoV-2 RNA sequence**. Since the ***HIV-1 Virus*** is already known for its lethality and **anthrax** is considered in the category A (i.e. of highest priority) of infectious agents associated with a

potential risk of being used as a biological weapon, it is understandable that I was very impressed by the search result in the database. In truth, despite the surprise I had at the discovery of the numerous significant alignments between the **SARS-CoV-2 RNA** sequence and the ***Anthrax Bacillus***, I had recently read about studies and experiments concerning the association between **anthrax** and **coronavirus**. In particular, I am referring to the research protocol made by researchers at the **University of Pittsburgh** to identify a potential candidate for the vaccine against the Sars-Cov-2 coronavirus that causes Covid-19 disease (see the document presented below).

Statement on Bacillus Anthracis SARS-CoV-2 Research

The goal of this research protocol is to identify a potential vaccine candidate against the SARS-CoV-2 coronavirus that causes the disease COVID-19. The research will involve expressing a single SARS-CoV-2 protein in an attenuated (weakened) strain of the bacterium Bacillus anthracis. This type of bacteria has not been shown to cause disease in humans or animals and is commonly used as a vaccine in farm animals to prevent Anthrax, a serious disease caused by a different type of bacteria. This attenuated strain does not produce toxins that can cause harm, but is useful for delivering viral proteins to cells in laboratory mice to produce immunity to protect against COVID-19.

This study will provide insights into how to manufacture an effective COVID-19 vaccine for use in humans, but is not intended to directly produce or test a vaccine. The same SARS-CoV-2 protein will be introduced into the cells that multiple laboratories around the world are using to learn more about COVID-19 and develop vaccine candidates and other potential treatments. This work will be performed at an appropriate laboratory biosafety level (BSL-2) by an established microbiologist.

Safety is always a top priority for the University of Pittsburgh, and we are confident that the research in question meets our stringent standards. The investigator's proposed experiments were reviewed and approved by the Institutional Biosafety Committee, which includes expert virologists, microbiologists and the University's Biosafety Officer. Prior to approving this research, the committee ensured the investigator demonstrated that this attenuated strain of bacteria did not produce toxins and that the bacteria used in the studies are not contaminated with any non-attenuated bacteria that could cause disease in humans.

(https://www.pittwire.pitt.edu/statement-bacillus-anthracis-sars-cov-2-research?fbclid=IwAR0KfslAd-xM2TEAx7C6xqtUHdmtdnxeUPVD8pBRLn9xgbhy_wXSAAC3YOM)

As reported on the University's own website, "... the research involves the expression of a single SARS-CoV-2 protein in an **attenuated (weakened) strain** of **Bacillus *anthracis*** bacterium .... this weakened strain does not produce toxins that can cause damage, but it is useful for delivering viral proteins to cells in laboratory mice to produce immunity and to protect themselves from Covid-19…. ". On the same page of the website, it is written that "**This study will provide insights into how to produce a Covid-19 vaccine that is effective for use in humans, but is not intended to directly manufacture or test a vaccine** .... The same SARS-CoV protein -2 will be introduced into the cells that more

laboratories around the world are using in order to learn more about Covid-19 and develop vaccine candidates and other potential treatments… ".

Anthrax infection occurs in *three forms*, but the **pulmonary form** (very rare and present only in risk categories), contracted by inhalation, is **considerably more serious** and has been associated with the widespread fear of a biological warfare (https://it.wikipedia.org/wiki/Bacillus_anthracis).

The *pulmonary form* of **anthrax infection begins with the typical symptoms of a flu (sore throat, fever, myalgia); followed by cough, chest pain and difficulty in breathing. This form is, from an anatomical-pathological point of view, hemorrhagic lymphadenitis-mediastinitis, rather than actual pneumonia, and does not involve the presence of bacilli in the sputum** (https://db0nus869y26v.cloudfront.net/it/Bacillus_anthracis).

My discovery of a relationship between **SARS-CoV-2 *RdRp Gene*** *(RdRp_SARSr-F)* **RNA** (gene currently used, along with others, to diagnose Covid-19) and **Human Immunodeficiency Virus DNA (HIV-1)** and that of **Bacillus Anthracis**, currently for me it remains inexplicable, but it deserves further investigation and clarification by the whole scientific community. In any case, this discovery completely upset the type of document which I had planned to write and publish. In my initial intent, this should have been a hyperspecialist work, less extensive and intended only for scholars of the sector (molecular biologists, biotechnologists, geneticists, etc.).

Contrary to what was expected, **now my main intentions are two**.

The **first** is to actively contribute to the fight against the pandemic by providing **specialists in the sector** with useful and unpublished information to carry out further scientific studies and continue the study aimed at creating *monoclonal, antiviral, antibiotic antibody therapies and specific anti-Covid 19 intervention protocols.*

The **second** aim is to reveal for the first time and to everybody the **mysteries inherent with the "non-obvious trends" present in all DNA or RNA sequences**, to spread their revolutionary knowledge and to teach, with the simplest and most accessible language (as much as it is allowed by the complexity of the subject), the scientific methodology with which I reached the conclusions summarized in the summary of this study. In this way I will **allow all scholars**, not only specialists in the field, but anyone with an average ability to manipulate data, who is interested and willing to do so, to **autonomously replicate my results**. For this purpose, it will be necessary to acquire the basic concepts of molecular biology and the necessary practice to carry out research on the *significant alignments* of a DNA or RNA sequence in a database of DNA and RNA (these teachings will be clearly highlighted as reserved for beginner readers), as well as a specific methodology for the modification of gene sequences, the latter of interest also and above all to the expert reader. Of course, the opportunity offered to replicate the results implies the **full publication of the algorithms used in this work and in those previously published.** Convinced that researchers will be able to treasure this information, I am now available for any clarifications or explanations and I wish everyone a good reading and a fruitful in-depth work.

READING GUIDE

INTRODUCTION

The Introduction is recommended for all readers. It explains very briefly the complete procedure to analyze and modify a (DNA or RNA) sequence. The in-depth information on the contents covered in the Introduction can be accessed through links in this same document or in other previously published documents.

CHAPTER I

Chapter I is reserved for beginners and has the general objective of allowing the reader to acquire some fundamental principles of *Molecular Biology* necessary to understand the contents covered in the following chapters. In particular, the concepts of *Molecular Biology*, *DNA* (Coding and Non Coding), *RNA* (Messenger RNA, Transfert, Ribosomal and Non Coding), *Nitrogenous Bases*, *Codons*, *Amino Acids* and *Protein Synthesis* are defined.

CHAPTER II

Paragraph 2.1 of Chapter II is reserved for beginners. It explains the basic procedure to carry out a search in the **BLAST** database and compare the *nucleotide sequences* or the *protein sequences* of DNA or RNA molecules. Instead, **paragraph 2.2 is recommended to all readers**, because the **BLAST research** is explained and integrated with the other *Phases* of the complete procedure proposed in my model of study of DNA (or RNA) and constitutes a preparatory phase. Furthermore, the example used to explain the fundamental concepts refers to the results of some *BLAST* researches concerning the of **SARS-COV-2 RNA** sequence which is object of this study.

CHAPTER III

Chapter III is recommended for all readers. In paragraph 3.1, bibliographical research is presented which investigates the hypothesis that there may be a report of transmission of *SARS-CoV* from *bats* to *pangolins* and from *pangolins* to *humans*. The next paragraph, on the other hand, describes the importance of the **RdRp gene** *of SARS-CoV-2*, clarified why the sequence of **22 RNA bases "GTGAAATGGTCATGTGTGGCGG"** was selected and explained the reasons for its multiple modifications.

CHAPTER IV

This chapter is recommended for all readers who have acquired from the previous chapters the skills necessary to perform a complete Blast Search and to interpret its results. It shows

the results of the complete *Blast Research* performed on the sequence of **22 RNA bases "GTGAAATGGTCATGTGTGGCGG"**, a sequence taken from the **RdRp Gene** (*pSARS-F*). As **RNA dependent RNA polymerase** (**RdRp**) is the engine of the *SARS-CoV-2* **transcription** process, understanding the results of this chapter is critical to interpreting the results that will be presented in the following chapters.

CHAPTER V

In this chapter (the "Goals of Chapter V" section is recommended for all readers) BLAST searches were carried out on the sequences **"GTGAAATGGTTATGTGTGGCGG"** and **"GTGAGATGGTCATGTGTGGCGG"**. These two sequences are important because they are the two sequences of the **RdRp gene** (*RNA dependent RNA polymerase*) whose significant alignments of BLAST research have identified the ***Bat*** (the only "organism" that the two sequences have in common) and ***Pangolin***, both of which are currently considered to be involved in the transmission of *SARS-CoV*.

CHAPTER VI

In Chapter VI (the "Goals of Chapter VI" section is recommended to all readers), in order not to overlook the impact of any other variations of nitrogenous bases in the same positions highlighted in blue in the sequence **"GTGAAATGGTCATGTGTGGCGG"** and identify any other "organisms" relevant for the purposes of this study, *all possible base replacements* were performed, resulting in *4* other *sequences*. These sequences have been the subject of BLAST researches.

CHAPTER VII

This chapter is recommended for all readers, but it **is indispensable** for those among them who have a strong motivation to understand the important and revolutionary topic of *DNA or RNA sequence analysis and modification* of based on the principles of **TRICHROMATIC THEORY OF EQUILIBRIUM OF SYSTEMS** (**T.T.E.S.**): the **procedure** through which one can **modify the Original Sequence** and **generate new DNA or RNA sequences**.

CHAPTER VIII

This chapter (the "Goals of Chapter VIII" section is recommended for all readers) is dedicated to **modifications** to the *RNA* sequence **"GTGAAATGGTCATGTGGCGG"** of the *RdRp Gene (RdRp_SARSr-F)*. The modifications of these **sequence and the methodology proposed in this document open up different perspectives to start targeted studies and innovative treatments of the Covid 19 pandemic**.

CHAPTER IX

This chapter (the "Goals of Chapter IX" section is recommended for all readers) contains the results of BLAST research carried out on **Sequence n° 1/3**. This sequence is one of many generated by the **Original Sequence**, the sequence "**GTGAAATGGTCATGTGTGGCGG**" of the RdRp Gene. Several "organisms" have been identified, including two ***Pipistrellus pipistrellus*** sequences and one ***HIV-1 virus*** sequence. Identified ***HIV-1 virus*** sequence was studied and the results were surprising. Relations have been found with **three** species of **bats**, **two** species of **pangolins** and various "organisms" involved in mainly **infectious**, **febrile** and **pulmonary diseases**. Also to be highlighted are the relationships with ***Influenza A virus***, a sequence that refers to the ***N-terminal domain of RNA polymerase directed by DNA*** and, above all, a significant number of sequences of the ***Bacillus anthracis***, a **deadly bacterium** that causes **lung damage**.

CHAPTER X

This chapter (the "Goals of Chapter X" section is recommended for all readers) contains the results of BLAST research carried out on **Sequence n° 3/1**. This sequence is one of many generated by the **Original Sequence**, the sequence "**GTGAAATGGTCATGTGTGGCGG**" of the RdRp Gene. Several "organisms" have been identified and, among them, **Manis pentadactyla** (*Chinese pangolin*), six sequences of **Pipistrellus pipistrellus**, four sequences of the **Influenza A Virus** and, above all, a considerable number of sequences of **Bacilli Paranthracis**, **Anthracis** and **Bacillus Cereus**. **Bacillus Anthracis**, due to its **lethality** and its massive involvement in **lung damage**, will be the subject of extensive discussion.

CHAPTER XI

This chapter (the "Goals of Chapter XI" section is recommended for all readers) contains the results of BLAST research carried out on **Sequence n° 8/1**. This sequence is one of many generated by the **Original Sequence**, the sequence "**GTGAAATGGTCATGTGTGGCGG**" of the RdRp Gene. Several "organisms" have been identified and, among them, also two species of bats (*Rhinolophus ferrumequinum* and *Pipistrellus pipistrellus*) and a considerable number of sequences of **Bacilli Paranthracis** and **Anthracis**.

CHAPTER XII

This chapter (the "Goals of Chapter XII" section is recommended for all readers) contains the results of BLAST research of other Modified Sequences of the RNA Sequence "GTGAAATGGTCATGTGTGGCGG" of the RdRp Gene (RdRp_SARSr-F). Modified sequences are sequences n° : **1/2**; **2/2**; **3/2**; **3/3**; **3/4**; **3/5**; **3/6**; **4/1**; **5/1**; **5/2**; **6/1**; **6/2**; **7/1**; **7/2**; **9/1**; **10/1**.

INTRODUCTION

Many metodological aspects of *DNA or RNA sequence analysis and modification* proposed in this paper are based on the principles of **TRICROMATIC THEORY OF EQUILIBRIUM OF SYSTEMS** (**T.T.E.S.**), a theory of systems through which it is possible to observe, analyse, control and modify the state of every system (http://www.ttesystems.eu/index_en.php; http://www.ttesystems.eu/application.php).
With appropriate adaptations of the basic **T.T.E.S.** software, it is possible to analyse and modify a DNA (or RNA) sequence in an *innovative way*.

The **complete procedure** for analyzing and modifying a DNA (or RNA) sequence involves **NINE PHASES** (http://www.ttesystems.eu/Cms/files/Pdf/Introduzione%20DNA%20ENG.pdf).

Due to the short time available for the drafting of this document and the huge amount of data presented, the ***Phases*** considered in this writing are only those highlighted below in **red** (and not all have been completed as required by the procedure):

First Phase: Analysis of the Original Sequence
Second Phase: Modification of the Original Sequence
Third Phase: Comparison of Graphic Representations
Fourth Phase: Blast research
Fifth Phase: Discovering and Highlighting Common Organisms
Sixth Phase: Specification of the Product of Significant Alignment
Seven Phase: Bibliographic Research focused on a single Organism
Eighth Phase: Bibliographic Research focused on two or more Organisms
Ninth Phase: Concluding Considerations and Pragmatic Choices

First Phase: Analysis of the Original Sequence

The **acquisition of the sequence** and its **4 principal graphic representation** constitute the FIRST PHASE of the study, the **ANALYSIS OF THE ORIGINAL SEQUENCE** (see http://www.ttesystems.eu/Cms/files/Pdf/Analysis%20and%20modification%20of%20DNA%20or%20RNA%20sequences%20with%20T.T.E.S.%20(Chapter%20II%C2%B0).pdf e http://www.ttesystems.eu/Cms/files/Pdf/Introduzione%20DNA%20ENG.pdf).
The acquisition of the DNA (or RNA) sequence to be analysed and modified can be obtained directly from the *organism* studied, from the website of the **NCBI** [National Center for Biotechnology Information (1)] or from any other source.

(1) National Center for Biotechnology Information (**NCBI**)[Internet]. Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information; [1988]. Available from: https://www.ncbi.nlm.nih.gov/

Second Phase: Modification of the Original Sequence

Subsequently (and at scientists' discretion), the SECOND PHASE**,** that is to say the **MODIFICATION OF THE ORIGINAL SEQUENCE (**and possibly the next phases) begins.
The **T.T.E.S.** software offers also the possibility to modify the **original sequence** and to generate numerous and different **new** DNA or RNA **sequences** that faithfully respect the numerous and different "*non-obvious trends*" of the original sequence.
The "**non-obvious trends**" are *trends*, not clearly identifiable, present in all DNA and RNA sequences (for further information see Chapters VIII and IX in this document and http://www.ttesystems.eu/Cms/files/Pdf/Introduzione%20DNA%20ENG.pdf).

Third Phase: Comparison of Graphic Representations

The afore-mentioned account is followed by a THIRD PHASE, the one where it is possible to **COMPARE THE GRAPHIC REPRESENTATIONS** of an **original sequence** with those of the **new sequences**.
From the *comparison of the four principal graphic representations* of an **original sequence** with those of its **new sequences**, important *similarities* and *differences* can emerge.

Fourth Phase: Blast research

The next phase, la FOURTH PHASE, is that of **BLAST RESEARCH** (*Basic Local Alignment Search Tool* (2)), respectively on the **original sequence** considered and on all its **new sequences** [for further information see Chapter II in this document and see all results presented in: DNA or RNA Sequence Analysis and Modification through T_T_E_S_ (Chapter I° - Second Part)].
The purpose of this Phase is to identify, for the **original sequence** and for the new **sequences**, all *significant alignments* with the sequences and with the *organisms* stored in the database.

Fifth Phase: Discovering and Highlighting Common Organisms

The FIFTH PHASE is this is dedicated to **DISCOVERING** and **HIGHLIGHTING** those "organisms" (and consequently the *Dna or Rna bases*) that are *in common* between the **original sequence** and its **new sequences** (for further information see paragraph 2.2 of Chapter II and the various examples presented in the following chapters in this document).

(2) Altschul S. F., Madden T. L., Schaffer A. A., Zhang J., Zhang Z., Miller W. and D. J. Lipman. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res., 1997, 25 (17) :3389-3402. PMID: 9254694. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC146917/

In http://www.ttesystems.eu/Cms/files/Pdf/Introduzione%20DNA%20ENG.pdf are the suggested **criteria** for performing the *highlighting* procedure in a relatively simple manner.

Sixth Phase: Specification of the Product of Significant Alignment

The SIXTH PHASE **SPECIFICIES**, for all "organisms" that are *in common*, the "**Product**" of the *significant alignment* and which *DNA* (*or Rna*) *modified bases* they have in common. An example of this last procedure is presented in Chapter I – *First Part* at page 33 [DNA or RNA Sequence Analysis and Modification through T_T_E_S_ (Chapter I° - First Part)].
Overall, the results of this phase are very important (along with the results of the SUCCESSIVE PHASES). It is indeed on them that will depend the eventual choice of intervening concretely in the modification of the **biological material** of the analysed sequences for the realization of the various and possible practical purposes.

Bibliographic Research focused on a single Organism

In the SEVENTH PHASE a **BIBLIOGRAPHIC RESEARCH FOCUSED** on a single organism is carried out in order to confirm the existence of important relationships between the characteristics (e.g., the "products" highlighted in the previous phase) of one of the *organisms* [e.g., *Pseudomonas*. For further information, see pages 37 - 46 of DNA or RNA Sequence Analysis and Modification through T_T_E_S_ (Chapter I° - First Part)], that has been identified by the Blast research performed on one of its **new sequences**, and some functional characteristics of the **original sequence**.

Eight Phase: Bibliographic Research focused on two or more Organisms

The researcher, at their discretion, and in order to reach specific goals, can carry out an EIGHTH PHASE, that is to say, a **BIBLIOGRAPHIC RESEARCH FOCUSED** on two or more organisms in order to confirm the existence of important relationships between the characteristics (e.g., the "products" highlighted in the previous phase) of **two** or **more** *organisms* [e.g., *Pseudomonas* and *Heligmosomoides polygyrus*. For further information, see DNA or RNA Sequence Analysis and Modification through T_T_E_S_ (Chapter I° - First Part)], identified by the Blast research performed on a **new sequence**, and some functional characteristics of the **original sequence**.

Ninth Phase: Concluding Considerations and Pragmatic Choices

In this LAST PHASE, it is possible to draw the **CONCLUSIONS** of the work performed in the previous phases and to make the necessary **PRAGMATIC CHOICES**, such as, to use the *acquired knowledge and/or the new* (partial or integral) *DNA or RNA sequences* for scientific, industrial, food, etc. research purposes, or to intervene concretely in the modification of the **biological material** (DNA or RNA) of one or more *sequences analysed* for the realization of the different possible purposes.

CHAPTER I

BASIC CONCEPTS OF MOLECULAR BIOLOGY

GOALS OF CHAPTER I

This chapter is reserved for beginners and has the general goal of allowing the readers to acquire some fundamental principles of Molecular Biology, which are necessary for understanding the contents covered in the following chapters.
In particular, the concepts of *Molecular Biology*, DNA and RNA (Coding and Non-Coding), *Amino Acids* and *Protein Synthesis* are here defined.
The *Nitrogen Bases* are considered with particular attention for the role they play in the coding of amino acids; indeed, the **order in the sequential arrangement of nucleotides** (of which the nitrogenous bases are the main constituents) constitutes the heart of the *genetic information*.
Also important are the concepts of *messenger* RNA, *Transfert* and *Ribosomal*, involved in the processes of *translation* and *transcription* and essential for the understanding of *protein synthesis*.
Two Codon Tables (codon or triplet) are attached at the end of the chapter for it is important to learn how to use to be able to translate a genetic code into a sequence of amino acids.

1.1 MOLECULAR BIOLOGY, AMINO ACIDS AND PROTEINS

Molecular biology is the branch of *biology* (the science that studies the physical and chemical processes of living systems) that deals with living beings at the level of *molecular* mechanisms (i.e., *molecules*, fundamental units of *substances*, composed of two or more *atoms* , of similar or different *elements*, joined together) at the basis of their *physiology* (of their functioning as living beings), focusing in particular on the interactions between *macromolecules* (molecules of very large size and very high molecular weight), i.e. *proteins* and *acids nucleic* (DNA and RNA) (https://en.wikipedia.org/wiki/Molecular_biology).

In living organisms, there are **two types of nucleic acids** (macromolecules responsible for the conservation and transport of genetic information): **DNA** and **RNA**.

DNA (or Deoxyribonucleic Acid) is a nucleic acid that contains the genetic information that are necessary for the *biosynthesis* (the chemical synthesis that occurs within a living organism) of **RNA** and **proteins**, molecules essential for the development and correct functioning of most living organisms (https://en.wikipedia.org/wiki/DNA).

RNA (or Ribonucleic acid) is involved in various biological roles, such as those of coding, decoding, regulation and expression of genes [fundamental hereditary units of living organisms, i.e. *nucleotide* sequences (see below) of DNA or more rarely, in certain **viruses**, of RNA] (https://en.wikipedia.org/wiki/RNA).

Chemically, **DNA** has a **double-stranded** structure (known as a double helix) whose constituent elements are called nucleotides. *Nucleotides* are made up of three fundamental components: a nitrogenous base (known also as nucleobase), a five-carbon sugar (ribose or deoxyribose), and a phosphate group consisting of one to three phosphates.

One of the **fundamental concepts** that must be memorized in order to understand the contents of this document is the following:

The **nitrogenous bases** that make up nucleotides are **four**:

adenine (A);
thymine (T);
cytosine (C);
guanine (G).

Instead, in the place of t**hymine**, in **RNA** is found uracil **U**.

Attention:
As a rule, in DNA and RNA databases, the *uracil* nucleobase is always defined with the letter **T** (it is good to remember it as it can cause confusion).

As shown in Fig.1.1, **DNA** is a *double polynucleotide chain* (in which there are numerous nitrogenous bases **A**, **T**, **C** and **G**), that is to say that it is formed by several nucleotides, which can be globally defined as : *antiparallel* (because the two nucleotide chains run in an inverse position with respect to each other), *oriented* (because the nucleotides organize themselves according to precise chemical laws that specify their orientation in space), *complementary* (because one chain is the mirror of the other), *spiralized* (because the chain twists in a double helix on itself), *informational* (because it carries genetic information within it) (https://en.wikipedia.org/wiki/DNA).

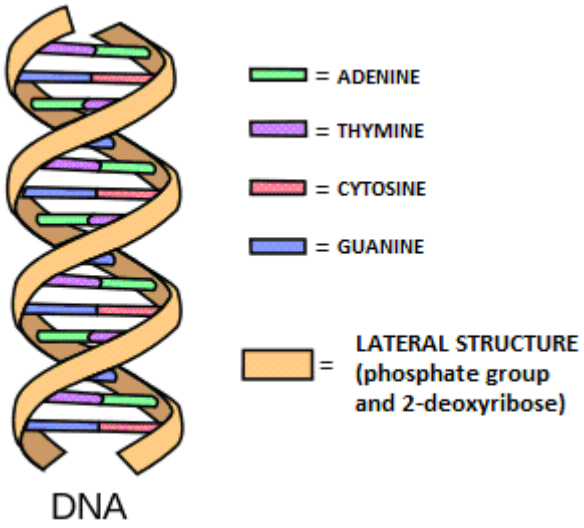


Fig. 1.1 [Double helix structure of DNA. Couplings among the four nitrogenous bases are here highlighted. It is **essential to remember** that **Adenine does always match** with **Thymine** and **Cytosine does always match** with **Guanine** (image taken and edited by https://it.wikipedia.org/wiki/DNA)]

In the **DNA lateral structure**, each phosphate molecule forms a molecular bridge, connecting, **carbon** of one deoxyribose molecule at **3 ′** with that at **5 ′** of the adjacent sugar ring (see Fig. 1.2 on the next page). As a result of these asymmetrical bonds **each DNA strand has a directionality**, **determined by the direction of the molecular bridges** (*phosphodiester bonds*).
In a double helix, the *direction of one strand is opposite to that of the complementary strand*. For this reason, the two filaments that make up a double helix are called **antiparallel**. The asymmetric ends of a DNA strand are defined as 5 ′ **end** and 3 ′ **end** and, *in vivo*, the only way to produce a synthesis of a *nucleic acid* follows the 5 ′ → 3′ direction, which is why the DNA and RNA sequences are written in the sense 5 ′ → 3′ (https://it.wikipedia.org/wiki/DNA).

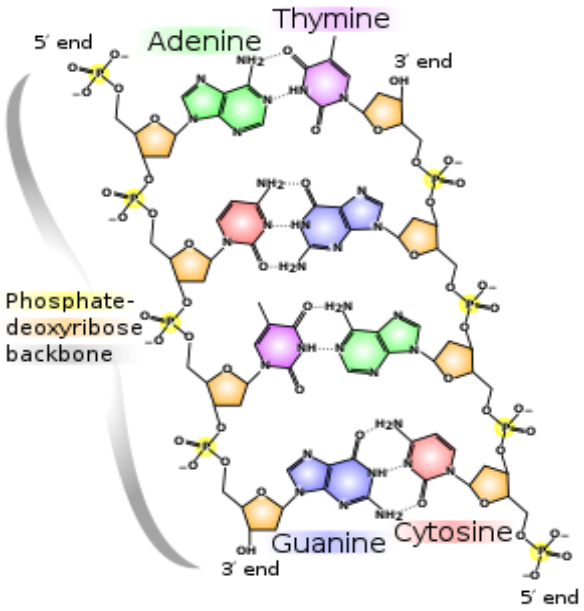


Fig. 1.2 [Chemical structure of DNA. **It is essential to remember that the only way to synthetize a nucleic acid follows the 5 '→ 3' sense** (image taken from https://en.wikipedia.org/wiki/DNA)].

Like DNA, **RNA** is assembled as a *nucleotide chain*, but **unlike DNA**, it is more common in nature as a **single strand folded over on itself,** rather than a paired double strand (https://en.wikipedia.org/wiki/RNA).
Many viruses [*biological entities* with characteristics of obligate parasites, as they replicate exclusively inside the cells of organisms https://en.wikipedia.org/wiki/Virus] encode their genetic information, by using RNA genome.

Another **fundamental concept** that must be memorized in order to understand the contents of this document is that the **order in the sequential arrangement of the nucleotides** constitutes the *genetic information*.

Since the nucleotides are substantially constituted and represented by the numerous nitrogenous bases **A**, **T**, **C** and **G**, it follows that the **sequential order assumed by the nitrogenous bases A**, **T**, **C** and **G** is the heart of the genetic information.

Let's make an example.

The following sequence of 22 nitrogenous bases

T A T C C C A T T A C A T G A T A T T A T T

will constitute a body of genetic information different from the following sequence of 22 nitrogenous bases

G T G A A A T G G T C A T G T G T G G C G G

except in the *area highlighted in gray*, where the information is the same because the bases are the same and are placed in the same order.

In **bioinformatics** this genetic information is treated in a **binary way**, that is, as **a series** of **0s and 1s**, and transformed into **algorithms.**

The *genetic information* is translated through the genetic code. The genetic code is the set of rules by which the information encoded in the *nucleotides* constituting the *genes* for the synthesis of *proteins* in cells is translated through the corresponding *amino acids* (organic molecules constitutive of proteins).
The amino acid sequence produced forms proteins, biological macromolecules made up of chains of *amino acids* that perform, in living organisms, a wide range of functions, including *metabolic* functions (the chemical transformations of macromolecules that occur in cells), synthesis functions such as DNA *replication* (the biological process of producing two identical replicas of DNA from an original DNA molecule), as well as functions of *cell signaling* (a process that makes cells able to perceive and respond correctly to their microenvironment) and the *transport* of molecules from a place to another of the organism (https://en.wikipedia.org/wiki/Protein).

The following are **fundamental concepts** that need to be memorized in order to understand the contents of this document. There are **three different types of RNA**:

- ***messenger RNA (mRNA),*** necessary to *transcribe* the amino acid sequence for the synthesis of proteins (https://en.wikipedia.org/wiki/Messenger_RNA);

- ***Transport RNA*** or ***Transfer RNA (tRNA),*** made up of the complementary triplet or anticodon that identifies an amino acid in the mRNA. It transfers a specific amino acid of o the protein synthetic machinery of a cell (ribosome), and as suchi s a necessary component of translation (https://en.wikipedia.org/wiki/Transfer_RNA);

- ***- Ribosomal RNA or rRNA,*** hooks the *messenger* RNA and scrolling it reads (translates) the triplets (or Codons. For further details see the next page) in interaction with the *tRNA* (which carries the amino acid). rRNA does not directly encode proteins, but is the essential component of ribosomes (which assemble proteins) present in all living cells (https://en.wikipedia.org/wiki/Ribosomal_RNA).

Cellular organisms use messenger RNA (**mRNA**) to transmit genetic information through the nitrogenous bases guanine (**G**), uracil (**U**), adenine (**A**) and cytosine (**C**), which direct the synthesis of specific proteins
(https://en.wikipedia.org/wiki/RNA).

The process of **genetic translation** is possible only in the presence of RNA, which is generated by *complementarity with the four bases of the DNA nucleotides* in a process known as **transcription** (https://en.wikipedia.org/wiki/Transcription_(biology)). This process not only generates RNA strands intended for translation, but also fragments that are already able to perform various biological functions (for example within ribosomes, where RNA has a structural function).

During **transcription**, the *double-stranded DNA* produces *RNA* starting from the so-called *antisense strand*; the other complementary filament is called *sense*. With the term **antisense RNA** (https://en.wikipedia.org/wiki/Antisense_RNA), we mean a single stranded RNA that is **complementary** to a protein coding messenger RNA (mRNA). In some organisms, the presence of an **anti-sense RNA** can inhibit genetic expression through pairing with specific RNAs.

Protein synthesis (the process of *genetic translation* that leads to the formation of proteins) is possible thanks to the existence of a specific **three-letter language** that associates *codons* [**3 nucleotides on RNA** or **triplets**, for example **AUG** (which in the database of Blast is written ATG, because U = T)] with *amino acids*, the constituents of proteins [for example, the codon AUG (or ATG) encodes the amino acid **Methionine** (**Met / M**)]. The two **Codon Tables** shown on the next page (Fig. 1.3 and Fig. 1.4) are used to translate a genetic code into an amino acid sequence.

The **codons** are present along the mRNA (*messenger RNA*), a type of RNA that **encodes** the information for the insertion of a specific *amino acid* during *protein synthesis* or for the **termination** of this process [such a codon is defined as *stop codon* (https://en.wikipedia.org/wiki/Stop_codon) or *termination* codon or *nonsense* codon. This type of codon is made of a triplet of bases which **does not encode** any amino acid (https://en.wikipedia.org/wiki/Non-coding_DNA) and therefore blocks the translation of the mRNA strand into proteins. There are **three stop codons**: **UAA**, **UGA** and **UAG**]. There are also **three start codons**, and the most common is **AUG** [which is read as the amino acid *Methionine* (**Met / M**) or, in bacteria, as *Formylmethionine*], while the other two are **GUG** and **UUG**, which normally encode respectively the amino acids **Valine** (**Val / V**) and **Leucine** (**Leu / L**), but as initial codons they are translated as *Methionine* or *Formylmethionine*.

The **codon** is at the basis of the genetic code (https://en.wikipedia.org/wiki/Genetic_code). Using **4 letters** (**A**, **C**, **G** and **T** or **U**) in groups of **three letters** (for example, **AUG**) you can have up to **64 different combinations** (4 x 4 x 4), able to *encode* the existing **twenty different amino acids**. Since there are **64 possible triplets** and only **20 amino acids**, the *genetic code* is said to be *redundant* because some amino acids can be encoded by **several different triplets**. For example, the amino acid **Valine** (**Val / V**), can be encoded indifferently by the codons **GUU**, **GUC**, **GUA** and **GUG**. However, the opposite is not true: each triplet will correspond to only one amino acid.

Before concluding, it is important to know that only a small fraction of the total sequence of a genome can be *transcribed* and *translated* for *protein synthesis*, while most of the genome is constitute by repeated sequences of non-coding DNA (https://en.wikipedia.org/wiki/Non-coding_DNA). Although the reasons for this are not yet fully understood, DNA sequences that do not encode a protein can be transcribed into **non-coding RNA** (https://en.wikipedia.org/wiki/Non-coding_RNA), which is involved in the regulation of gene expression.

Amino-acid biochemical properties	Nonpolar †	Polar †	Basic ‡	Acidic ↓

Termination: stop codon

Standard RNA codon table

Standard genetic code

1st base	2nd base								3rd base
	U		C		A		G		
U	UUU	(Phe/F) Phenylalanine †	UCU	(Ser/S) Serine †	UAU	(Tyr/Y) Tyrosine †	UGU	(Cys/C) Cysteine †	U
	UUC		UCC		UAC		UGC		C
	UUA	(Leu/L) Leucine †	UCA		UAA	Stop (*Ochre*)[note 2]	UGA	Stop (*Opal*)[note 2]	A
	UUG		UCG		UAG	Stop (*Amber*)[note 2]	UGG	(Trp/W) Tryptophan †	G
C	CUU		CCU	(Pro/P) Proline †	CAU	(His/H) Histidine ‡	CGU	(Arg/R) Arginine ‡	U
	CUC		CCC		CAC		CGC		C
	CUA		CCA		CAA	(Gln/Q) Glutamine †	CGA		A
	CUG		CCG		CAG		CGG		G
A	AUU	(Ile/I) Isoleucine †	ACU	(Thr/T) Threonine †	AAU	(Asn/N) Asparagine †	AGU	(Ser/S) Serine †	U
	AUC		ACC		AAC		AGC		C
	AUA		ACA		AAA	(Lys/K) Lysine ‡	AGA	(Arg/R) Arginine ‡	A
	AUG	(Met/M) Methionine †	ACG		AAG		AGG		G
G	GUU	(Val/V) Valine †	GCU	(Ala/A) Alanine †	GAU	(Asp/D) Aspartic acid ↓	GGU	(Gly/G) Glycine †	U
	GUC		GCC		GAC		GGC		C
	GUA		GCA		GAA	(Glu/E) Glutamic acid	GGA		A
	GUG		GCG		GAG		GGG		G

Fig. 1.3 (https://en.wikipedia.org/wiki/DNA_and_RNA_codon_tables)

Standard DNA codon table

Standard genetic code

1st base	2nd base								3rd base
	T		C		A		G		
T	TTT	(Phe/F) Phenylalanine †	TCT	(Ser/S) Serine †	TAT	(Tyr/Y) Tyrosine †	TGT	(Cys/C) Cysteine †	T
	TTC		TCC		TAC		TGC		C
	TTA	(Leu/L) Leucine †	TCA		TAA	Stop (*Ochre*)[note 2]	TGA	Stop (*Opal*)[note 2]	A
	TTG		TCG		TAG	Stop (*Amber*)[note 2]	TGG	(Trp/W) Tryptophan †	G
C	CTT		CCT	(Pro/P) Proline †	CAT	(His/H) Histidine ‡	CGT	(Arg/R) Arginine ‡	T
	CTC		CCC		CAC		CGC		C
	CTA		CCA		CAA	(Gln/Q) Glutamine †	CGA		A
	CTG		CCG		CAG		CGG		G
A	ATT	(Ile/I) Isoleucine †	ACT	(Thr/T) Threonine †	AAT	(Asn/N) Asparagine †	AGT	(Ser/S) Serine †	T
	ATC		ACC		AAC		AGC		C
	ATA		ACA		AAA	(Lys/K) Lysine ‡	AGA	(Arg/R) Arginine ‡	A
	ATG[note 4]	(Met/M) Methionine †	ACG		AAG		AGG		G
G	GTT	(Val/V) Valine †	GCT	(Ala/A) Alanine †	GAT	(Asp/D) Aspartic acid ↓	GGT	(Gly/G) Glycine †	T
	GTC		GCC		GAC		GGC		C
	GTA		GCA		GAA	(Glu/E) Glutamic acid ↓	GGA		A
	GTG		GCG		GAG		GGG		G

Fig. 1.4 (https://en.wikipedia.org/wiki/DNA_and_RNA_codon_tables)

CHAPTER II

BASIC PROCEDURE FOR CARRYING OUT A BLAST RESEARCH AND ITS MAIN GOALS

GOALS OF THE CHAPTER II

The goal of this chapter is to teach motivated beginners the basic procedure for searching the **BLAST** database (**B**asic **L**ocal **A**lignment **S**earch **T**ool (1)) and the easiest way to use the results obtained from this research. Information on the characteristics of the sequences is acquired directly from **NCBI** [*National Center for Biotechnology Information* (2)] website.

By using **BLAST,** one can compare the *nucleotide sequences* or the *protein sequences* of DNA or RNA molecules. A **BLAST** re*search* allows the researcher to compare a *sequence of interest* with a database of *already known sequences* (nucleotide or protein) and to identify, among the latter, those that have **similarities** with the *sequence of interest*. For example, following the ongoing infection with Covid-19, after having **sequenced** part of the genome of this new Coronavirus - that is, having determined the order of the **nitrogenous bases** [Guanine (**G**), Uracil or Thymine (**U** or **T**), Adenine (**A**) and Cytosine (**C**)] in the case of a nucleic acid (DNA or RNA) or amino acids in the case of **proteins** - scientists have been and are still interested in doing a BLAST research in the genome (of all "organisms") stored in the database to see if it contains **genes with similar sequences**.

After finding, **genes with similar sequences**, the next step is to experimentally investigate the function of the identified genes and thus obtain clues about the possible functions of the genes of the new Coronavirus. The **BLAST** research assigns **scores** to the *similarly identified sequences* and *sorts* these sequences according to those scores. The *similar sequences* considered by the **BLAST** research will ONLY be those whose **score** will exceed a certain *critical threshold* and will be defined as **HPS** (**H**igh-scoring **S**egment **P**air).

In conclusion, an **HSP** is a subsegment of a *pair of sequences*, *nucleotides* or *amino acids*, which share a **high level of similarity**. The **level of similarity between the sequences** depends on the *settings* of the local or global alignment *algorithm* that generated them. Depending on the sequences, the alignment algorithm and the settings used, it is possible to find numerous HSPs within a given **pair of sequences entered**, i.e. the **known sequence** (present in the database) and the **sequence of interest** (the one being studied) (https://genomevolution.org/wiki/index.php/High-scoring_segment_pair).

(1) Altschul S. F., Madden T. L., Schaffer A. A., Zhang J., Zhang Z., Miller W. and D. J. Lipman. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res., 1997, 25 (17) :3389-3402. PMID: 9254694. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC146917/

(2) National Center for Biotechnology Information (**NCBI**)[Internet]. Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information; [1988]. Available from https://www.ncbi.nlm.nih.gov/

2.1 HOW TO PERFORM A BASIC BLAST RESEARCH

To carry out a **BLAST** research, write the word "BLAST" in the search engine and click on "BLAST: Basic Local Alignment Search Tool" (red arrows in Fig. 2.1).

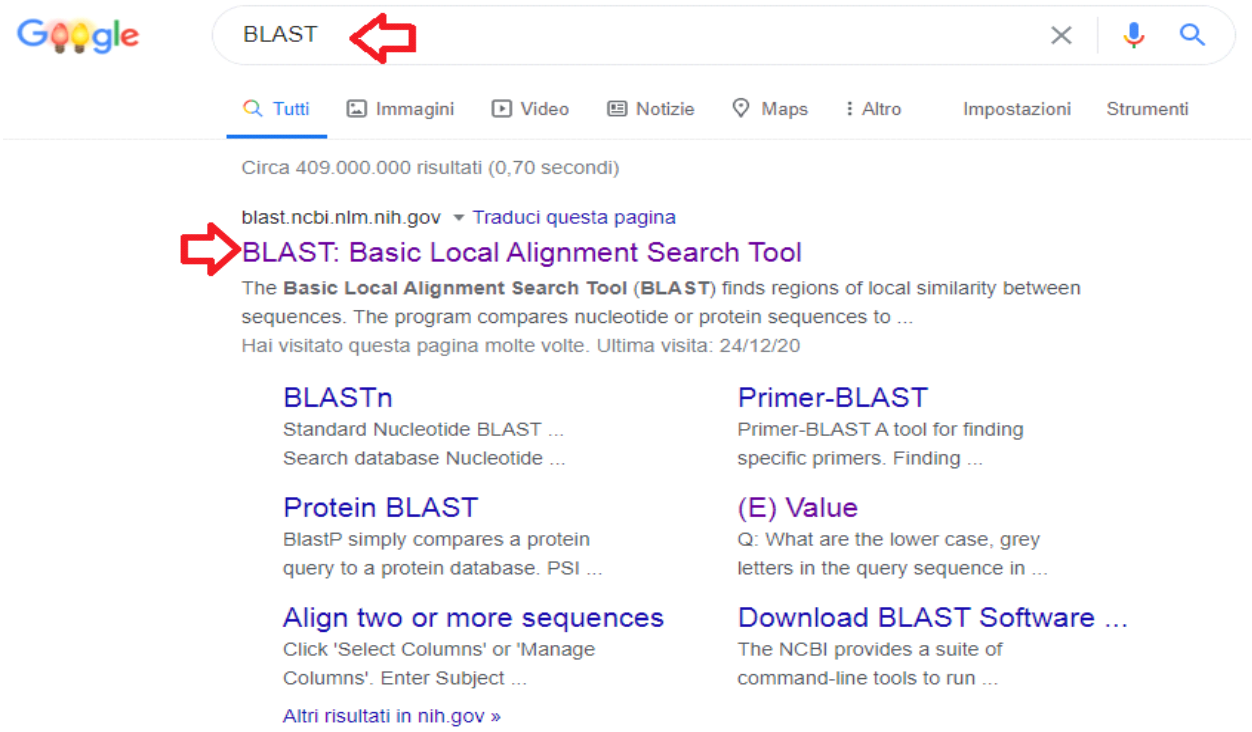


Fig. 2.1

Continue by clicking on the button provided to choose the type of search you wish to carry out. There are **several possibilities of choice** (see Fig.2.2 on the next page), but for the purposes of this study and teaching, we will consider only **two**:

1) **Nucleotide Blast** (red arrow in Fig.2.2 on the next page)
With this type of search, we have the possibility to insert a nucleotide sequence (a sequence of T, C, G and A) of *interest* (such as, for example, the *ATTGCGCGGATTATGTCC* sequence) and identify those highly similar to that of *interest*, among the known nucleotide sequences;

2) **Protein Blast** (green arrow in Fig.2.2 on the next page)
With this type of search, we have the possibility to insert a sequence of amino acids [a sequence of letters corresponding to *triplets* (or codons) of T bases, C, G and A] of *interest* (such as the NRPNH sequence) and identify those highly similar to that of *interest*, among the known amino acid sequences. The amino acids are the 20 basic constituents of proteins (*20 letters*). Proteins are formed by sequences of letters and each letter refers to a codon.

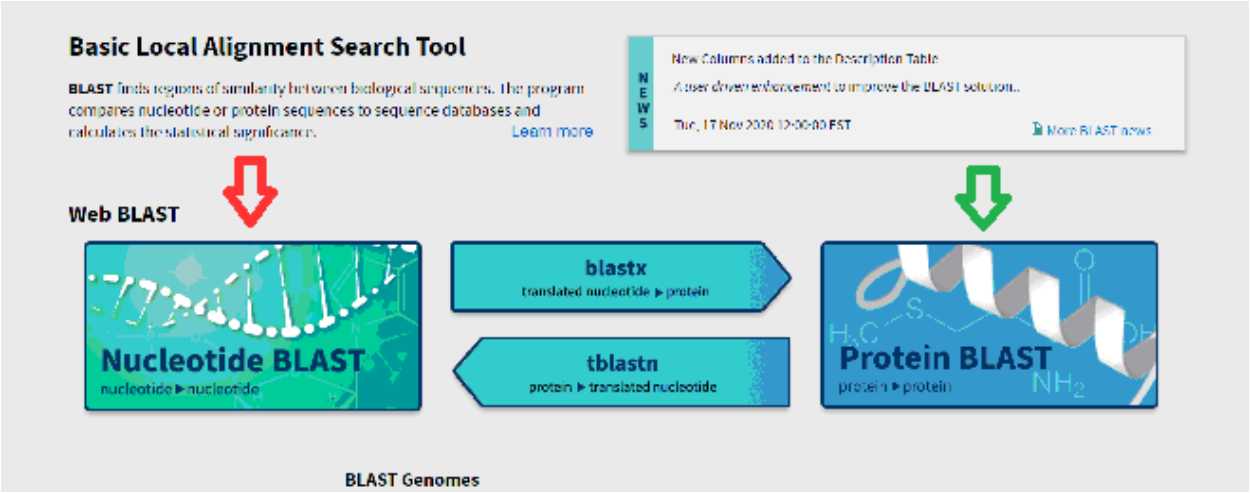


Fig. 2.2

For educational purposes, we perform a **simple basic** type **1 search** (**Nucleotide Blast**, red arrow in Fig. 2.2).

- we create a **sequence of interest**, a **RANDOM NUCLEOTIDE SEQUENCE of 18 BASES**:

ATTGCGCGGATTATGTCC

- **we copy and paste** the sequence in the space provided in BLAST (red arrow in Fig. 2.3).

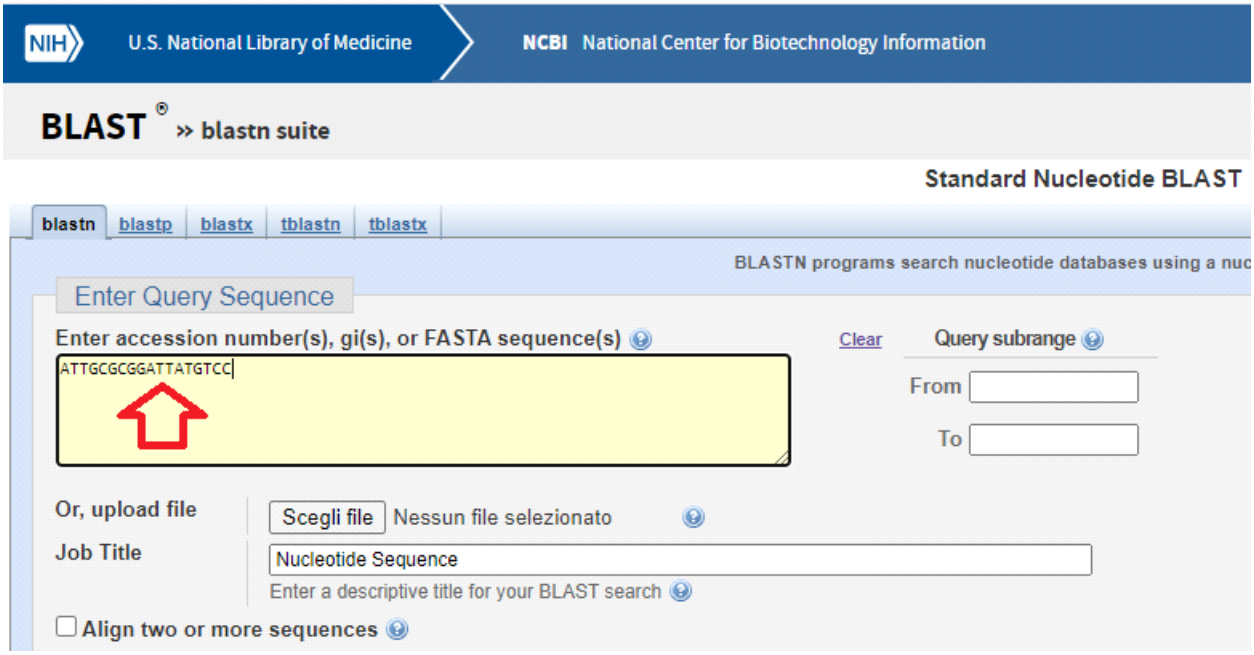


Fig. 2.3

- **click** at the bottom **on Algorithm parameters** (red arrow in Fig. 2.4).

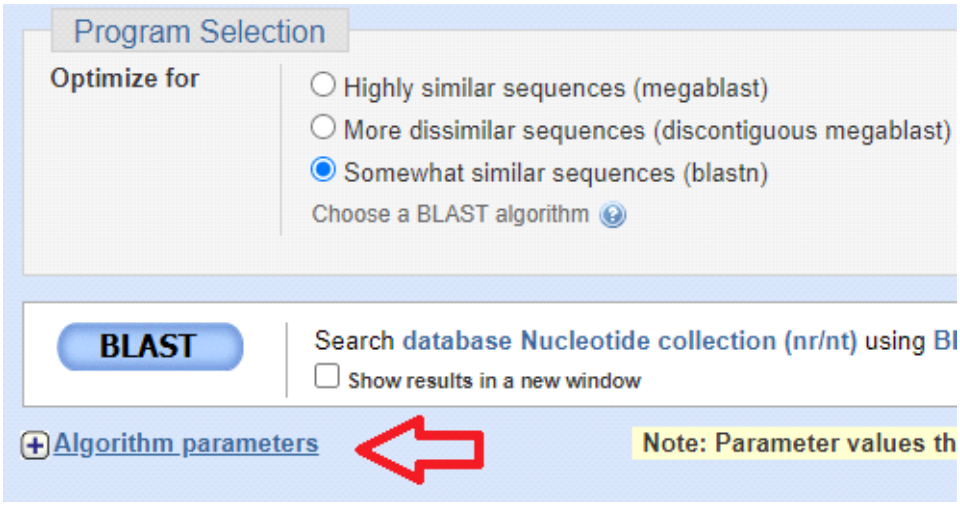


Fig. 2.4

- **- we establish a maximum number of sequences** we wish to obtain from the **BLAST** research (*Max target sequences*, red arrow in Fig. 2.5. In the example shown below, the number **10** has been set as default) and then **click** on the **BLAST** key (green arrow in Fig. 2.5).

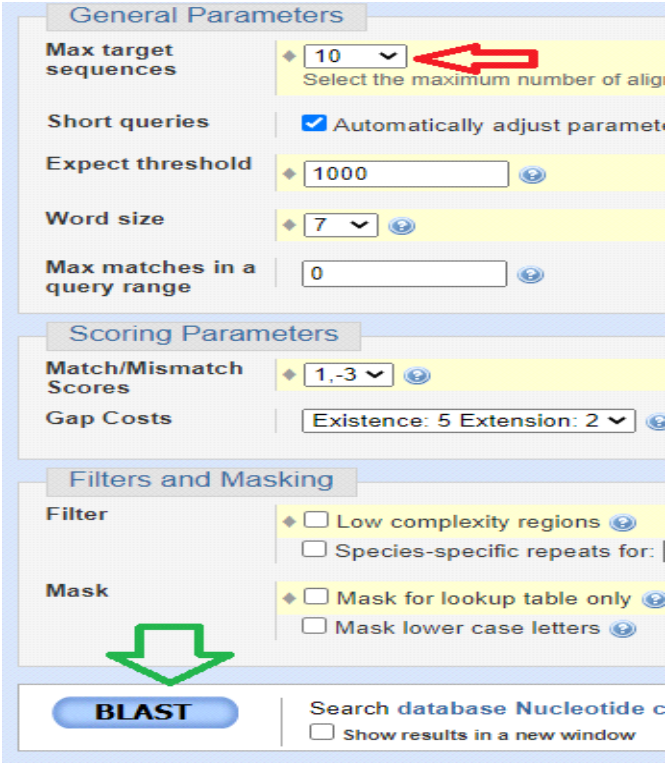


Fig. 2.5

- after waiting the time necessary for carrying out the **BLAST** research, we can **save** the search (*Save Search*, red arrow in Fig. 2.6), get the the **summary** of the various search parameters (*Search Summary*, green arrow in Fig. 2.6) and view the search in the recent results (Recent Results blue arrow in Fig.2.6).

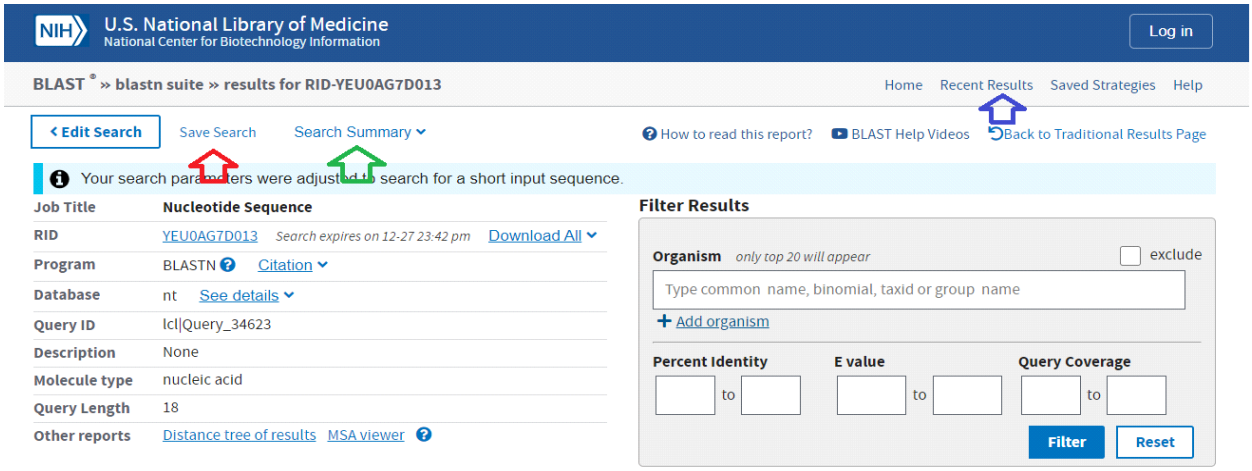
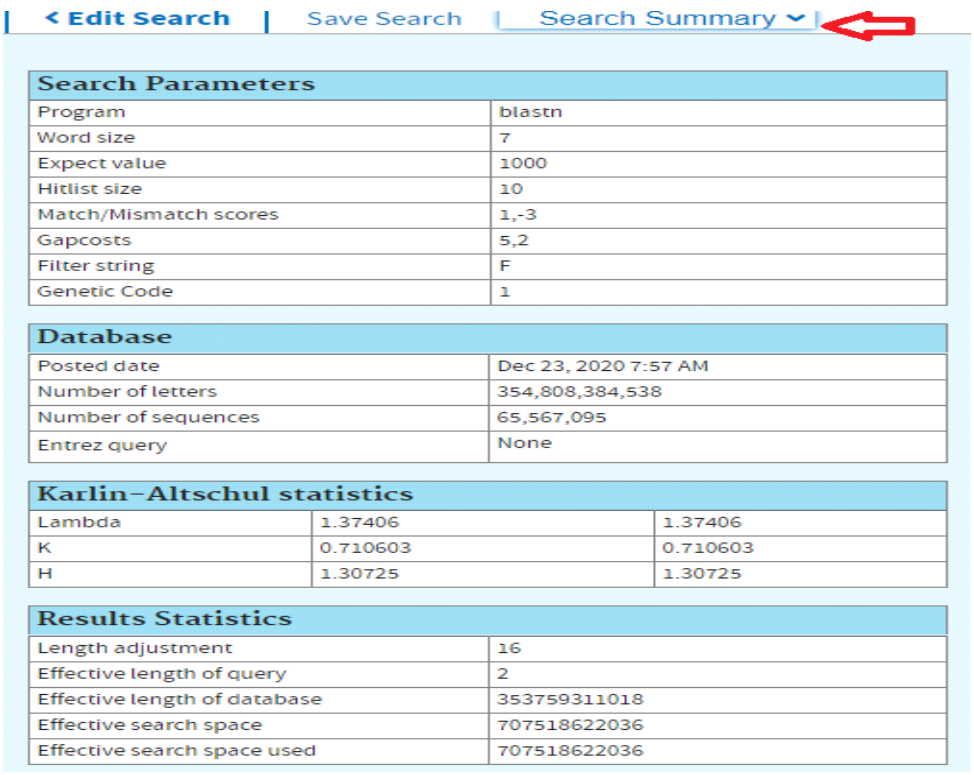


Fig. 2.6

- **by clicking** on the **summary** of the various search parameters [(*Search Summary*, red arrow in Fig. 2.7 (a)] the general data of the search are shown.



Search Parameters	
Program	blastn
Word size	7
Expect value	1000
Hitlist size	10
Match/Mismatch scores	1,-3
Gapcosts	5,2
Filter string	F
Genetic Code	1

Database	
Posted date	Dec 23, 2020 7:57 AM
Number of letters	354,808,384,538
Number of sequences	65,567,095
Entrez query	None

Karlin-Altschul statistics		
Lambda	1.37406	1.37406
K	0.710603	0.710603
H	1.30725	1.30725

Results Statistics	
Length adjustment	16
Effective length of query	2
Effective length of database	353759311018
Effective search space	707518622036
Effective search space used	707518622036

Fig. 2.7 (a)

- **scrolling** downwards, in **Descriptions,** the sequences producing **significant alignments** with the **sequence we have entered** (the sequence of interest) are shown. **By clicking in the small space "select all"** [red arrow on the left of the screen: see Fig. 2.7 (b)] we select all the sequences that have obtained significant alignments.

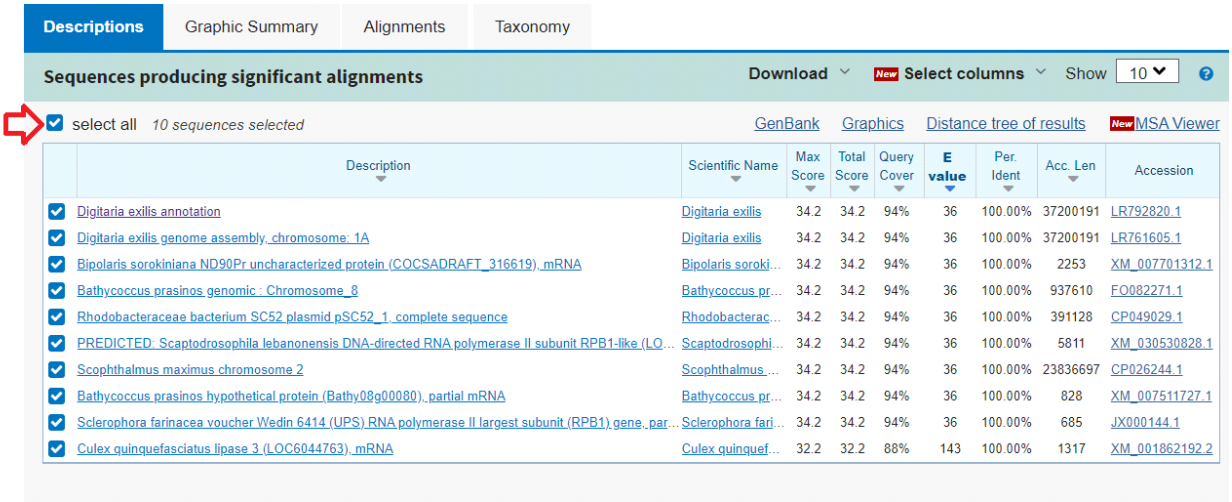


Fig. 2.7 (b)

- **By clicking on Alignments** [red arrow at the top of the screen: see Fig. 2.8 (a)] and then selecting, in *Alignments view*, *Flat query-anchored with letters in identities* [red arrow slightly to the right and further down the screen: see Fig. 2.8 (a)], it is possible to see the **bases of the sequence we have entered** (the sequence of interest, defined **Query**), **the access codes** of the **sequences identified** by the search and the *alignment* with the **bases of all the identified sequences.**

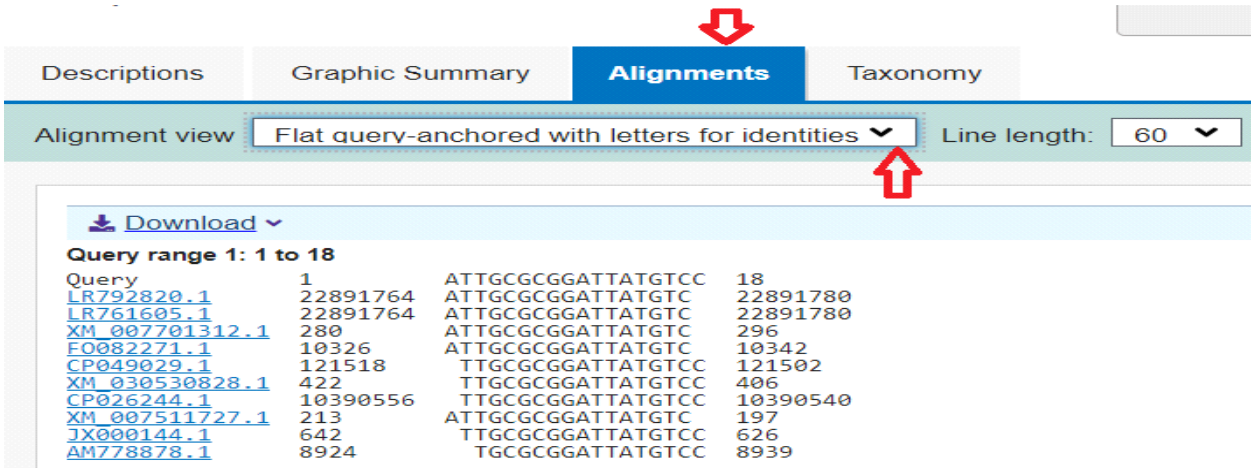


Fig. 2.8 (a)

- **By clicking in the small space** provided to the left of the names of the various identified sequences [red arrow on the left of the screen: see Fig. 2.8 (b)], it is possible to select a specific sequence; whilst by clicking on the **Access Code** link [*Accession*, red arrow on the right of the screen: see Fig. 2.8 (b)], it is possible to view the complete information sheet of the sequence.

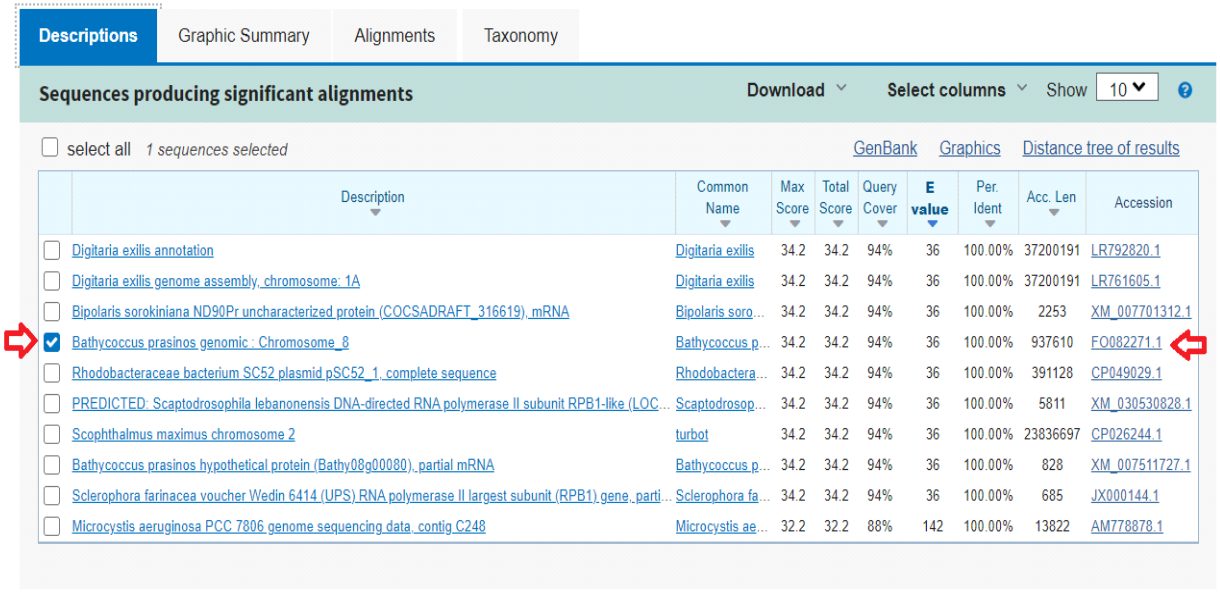


Fig. 2.8 (b)

Returning to Alignments, we select *Pairwise* (red arrow in the center, at the top and slightly to the right of the screen: see Fig.2.9)

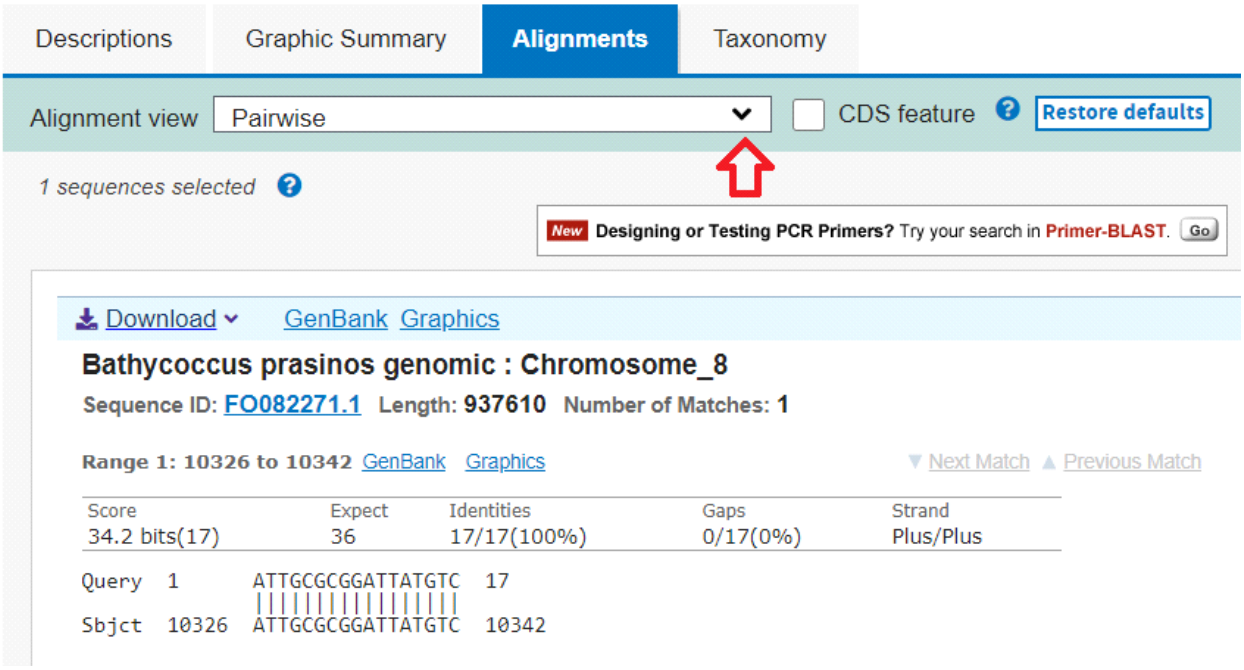


Fig. 2.9

- **By clicking** on *CDS features* or *Coding features* (red arrow in the center, lower than the previous arrow on the screen: see Fig.2.10) the most important references for the study of the **sequence we have entered** (the *sequence of interest*, **Query**) are shown. In this way, it is possible to see, in particular, the significant alignment with the sequence we had previously *selected* **by clicking in the small space** provided [red arrow on the left of the screen: see Fig. 2.8 (b)] in **Descriptions.**

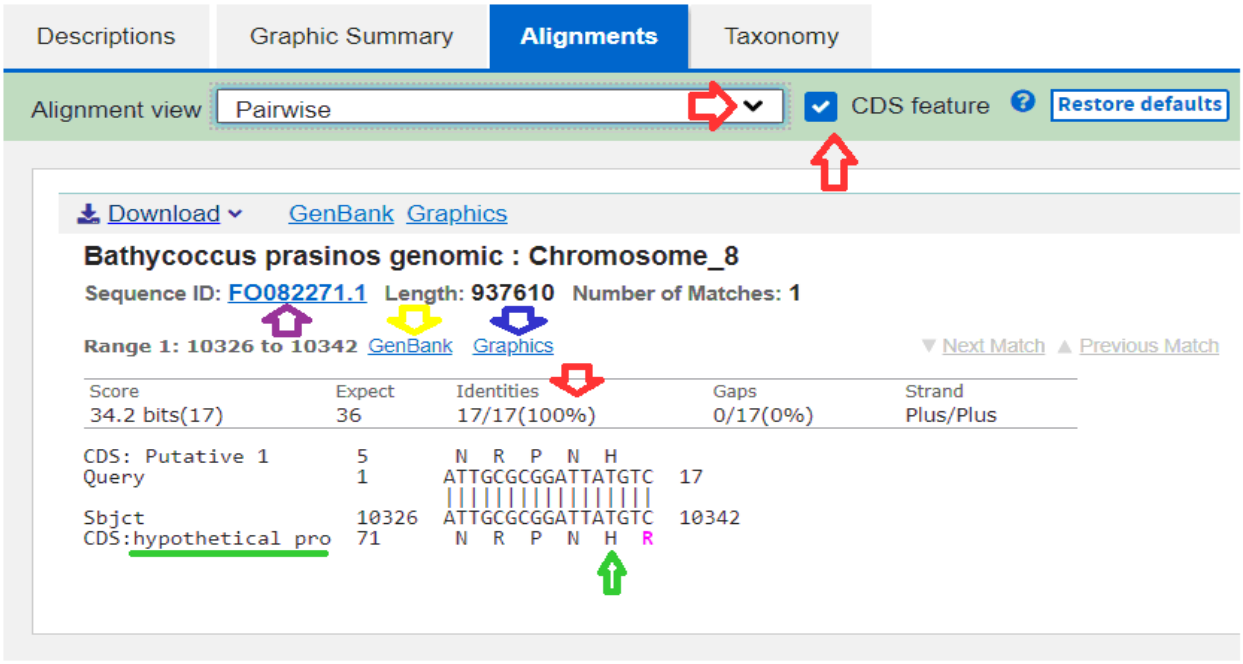


Fig. 2.10

In the example shown in Fig. 2.10, the sequence selected in **Descriptions** is the one that identifies "*Bathycoccus prasinos genomic: Chromosome_8*", whose **Access Code** link is FO082271.1 (purple arrow).

- **By clicking** on the *Access Code*, you can view the **complete information sheet** with the sequence selected in **Descriptions**, that is to say the sequence which was considered highly similar to the s**equence we have entered** (the *sequence of interest* or **Query**).

- **By clicking** on *GenBank* (yellow arrow in Fig. 2.10), it is possible to view the **Gene specific** information sheet of the **Query** sequence (see Fig. 2.11 shown on the next page). Among other important information, it is possible to view the **publication** dates (red arrow at the top right of the screen: see Fig. 2.11), the **presentation** for publication and **references** of who made the publication (red arrow further down and slightly to the left in the screen: see Fig.2.11).

GenBank

Bathycoccus prasinos genomic : Chromosome_8

GenBank: FO082271.1

FASTA Graphics

Go to:

```
LOCUS       FO082271                17 bp    DNA     linear   PLN 09-NOV-2012
DEFINITION  Bathycoccus prasinos genomic : Chromosome_8.
ACCESSION   FO082271 REGION: 10326..10342
VERSION     FO082271.1
KEYWORDS    green alga, Mamiellales, picoeukaryote, phytoplankton.
SOURCE      Bathycoccus prasinos
  ORGANISM  Bathycoccus prasinos
            Eukaryota; Viridiplantae; Chlorophyta; Mamiellophyceae;
            Mamiellales; Bathycoccaceae; Bathycoccus.
REFERENCE   1  (bases 1 to 17)
  AUTHORS   Genoscope -,C.E.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-OCT-2011) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
FEATURES             Location/Qualifiers
     source          1..17
                     /organism="Bathycoccus prasinos"
                     /mol_type="genomic DNA"
                     /strain="RCC 1105"
                     /db_xref="taxon:41875"
     gene            complement(<1..>17)
                     /locus_tag="Bathy08g00080"
     mRNA            complement(<1..>17)
                     /locus_tag="Bathy08g00080"
                     /product="hypothetical protein"
     CDS             complement(<1..>17)
                     /locus_tag="Bathy08g00080"
                     /codon_start=3
                     /product="hypothetical protein"
                     /protein_id="CCO17910.1"
                     /translation="MAIVDVSPFQKCIGPASHHQTHAKVNTPKTTTRLKNTTSQEEKD
                     DDDTAINKTLFRQSLPLSLQPRHNPRNTLFREFYDAGEIPIAVAHGGGSNAVTWRVDP
                     ASLDYHRYLPLFFDGIREIDHPYRLLSVQGAKQLLAVRPEKVADVLPQVILPLQNALK
                     TRDAPIVAVAMDILRHLATCEKSIGQMLVPYYRQLLPILNVFRNCDCLKVTLNVLSDG
                     SKRDGKELIGKGRSAKSLRVLVKETLRVLERVGGEDALINIKYMVPTYESCVAIAER"
ORIGIN
        1 attgcgcgga ttatgtc
//
```

Fig. 2.11

The **yellow** arrow in Fig. 2.11 indicates a part or the entire **sequence we have entered** (the *sequence of interest* or **Query**) and that was found identical (*significant alignment*) to the sequence selected in **Descriptions**, the one that identifies "*Bathycoccus prasinos genomic: Chromosome_8* ".
The **blue** arrow in Fig. 2.11 (shown on the previous page) indicates another very important information, that is the **Product** of the sequence selected in **Descriptions**. In the example shown, the **Product** is a **hypothetical protein** whose **amino acid** sequence (which in the tab is described after *translation* =) is indicated by the **green** arrow. Here, I remind the reader that an **amino acid sequence**, like that of *interest*, the *NRPNH* sequence, is described by a sequence of letters and each letter refers to a codon (or triplet of bases T, C, G and A).

In Fig. 2.12 below, underlined in **green,** the **Product** (*hypothetical protein*) of the Sequence selected in **Descriptions**, the one that identifies "*Bathycoccus prasinos genomic: Chromosome_8* " is presented. In the same image, it is possible to observe **Query**, the **sequence that we have entered** (the *sequence of interest*), to which **Sbjct** (the sequence resulting from the significant alignment with *Query*) is aligned. The **Identity** (in Identities) between *Query* and *Sbjct* is 100% (17 letters out of 17, red arrow at the bottom in Fig. 2.12). Here, it is of absolute importance to underline that, the **Identity** is 100% because **the sequence of identical bases is uninterrupted** (i.e., there are no missing or different letters); however, the initial bases entered for the search were **18** (one more base), so *in relation to the* total number of bases, the percentage that reflects the aforementioned **Identity is lower**.

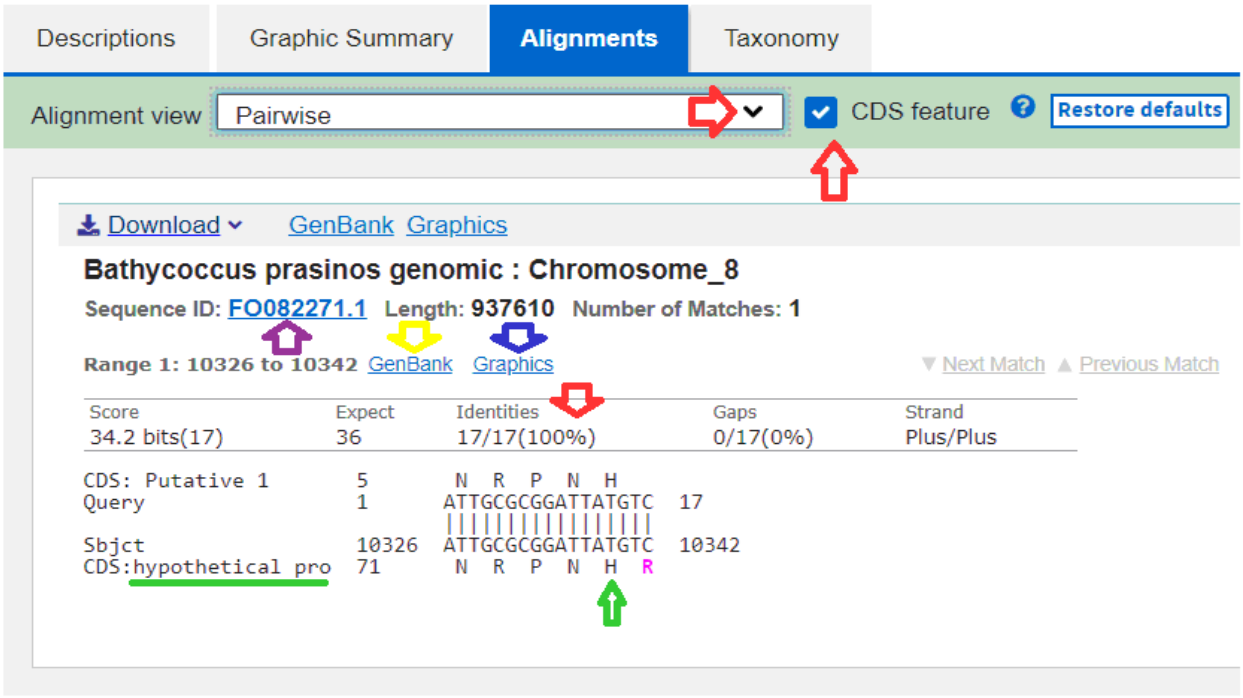


Fig. 2.12

The **green** arrow in Fig. 2.12 instead indicates the sequence of **amino acids** (the *NRPNH* sequence) and of it, in particular, indicates the letter **H**. The letter **H** has been chosen only as an example to point out that, right above it, there are THREE letters (**ATG**) of the **Sbjct sequence** and that **H** is placed under the middle letter (the letter **T**).
This way of presenting the letters indicates that the triplet (or codon) **ATG** encodes the **amino acid H**, *Histidine* (**His / H**). However, if we check the table of amino acids (see Fig. 1.4 on page 23) we discover that the letter **H** does not correspond to the **ATG** codon, but it corresponds to the **CAT** and **CAC** codons. On page 37, I will explain the apparent inconsistency.

- **By clicking** on the *Graphics* link (blue arrow in Fig. 2.12 shown on the previous page) a window opens in which two sequences, running in the opposite direction, are shown (see Fig. 2.13 presented on the next page), in full and with **white** letters highlighted in **gray**. *In a double helix, the direction of one strand is* ***opposite*** *to that of the complementary strand*. The asymmetric ends of a DNA strand are defined as *5′* end and *3′* end and, *in vivo*, direction, which is why the DNA and RNA sequences are written in the sense 5′ → 3′ (https://en.wikipedia.org/wiki/DNA).

In Fig. 2.13, in the two mirror chains, the direction 5 ′ → 3 ′ is indicated by the **red** and **black** arrows. To understand what, in Fig. 2.13, is the correct reading sense (the sense 5 '→ 3') it is necessary to observe the small **number 5** ' written in **blue** at the beginning of the two sequences (see the enlargement of these details in Fig. 2.14 shown, at the bottom of the next page). In Fig. 2.13, reading from the left to the right of the screen, the **sequence placed at the top** (*Sequence*) follows the direction 5 ′ → 3 ′, while the **one at the bottom** (*Genes*) flows in the opposite or antisense direction (3 ′ → 5 ′).

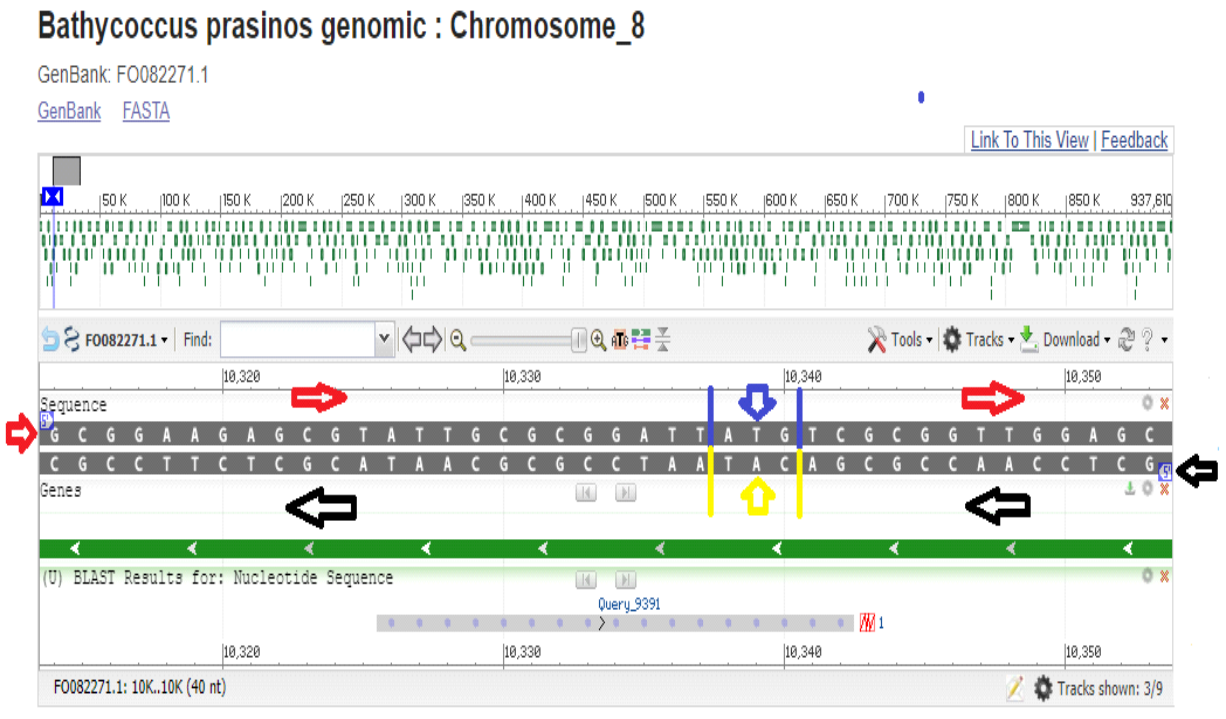


Fig. 2.13

Fig. 2.14

In the alignment established in BLAST, the **sequence placed at the top** corresponds to the **Query sequence**, the **sequence we entered** (the *sequence of interest*) (https://en.wikipedia.org/wiki/Molecular_biology), to which **Sbjct** (the sequence resulting from meaningful alignment with *Query*) is aligned. In our example, as it has already been said, the **Identity** (in Identities) between *Query* and *Sbjct* is 100%.
As you can see, the **blue** arrow (shown in Fig. 2.13) indicates the **ATG** triplet and corresponds the **TAC** mirror triplet, the one indicated below ATG by the **yellow** arrow.
As it has been mentioned in Chapter II, the letter **A** placed on one chain is associated with the mirror letter **T** placed on the other. The same goes for the letters **G** and **C**. So, simplifying, **ATG** and **TAC** are a different way to refer to the same DNA sequence, because they correspond to the letters placed on the two mirror chains. However, **TAC** is not the codon that encodes the **H amino acid**, which, as it has already been anticipated on page 36, is encoded by the codons **CAT** and **CAC**. However, if we read the **sequence placed at the bottom** (*Genes*), the *Sbjct* sequence in Fig. 2.13, in the 5 ′ → 3 ′ direction (the direction indicated by the black arrows), i.e. from the right of the screen to its left, the **TAC** codon will be read **CAT** and matches with one of the two codons that encodes the **H amino acid** (see Fig. 2.15 below on this page). **Attention:** to decide the reading direction of the codon it is necessary to refer to the reading direction established by the direction indicated by the **Purple Arrow** shown in Fig. 2.13.

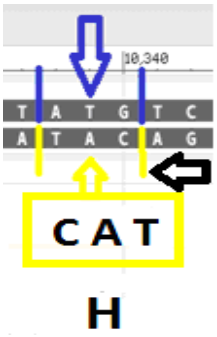


Fig. 2.15

If the reading direction established by the direction indicated by the **Purple Arrow** runs in the opposite direction to that shown in Fig. 2.13, that is to say, in a 5 ′ → 3 ′ direction parallel to the **sequence at the top** [see Fig. 2.16 shown on the next page. The figure refers to the graph of the **RNA** sequence (**GTGAAATGGTCATGTGTGGCGG**) which is the object of this study on **SARS-COV-2**, (that is) the **sequence placed at the top**], then the

coded amino acid will not be Histidine (**H**), but that indicated by the letter **M** (red arrow at the bottom), that is Methionine (**Met / M**) and whose codon is constituted by the **ATG** triplet corresponding directly with the **sequence placed at the top** indicated by the black arrow directed downwards.

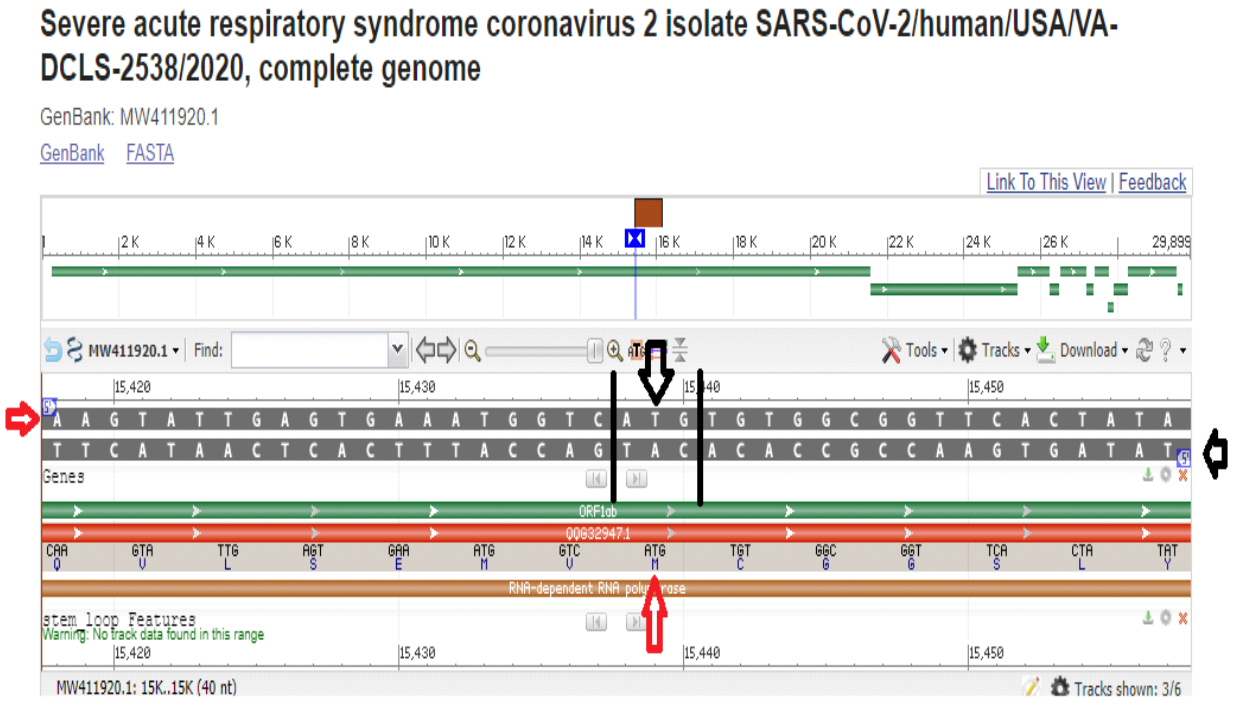


Fig. 2.16

Compare Fig. 2.16, which has already discussed in the previous pages, with Fig. 2.17 shown below, an identical copy of Fig. 2.13.

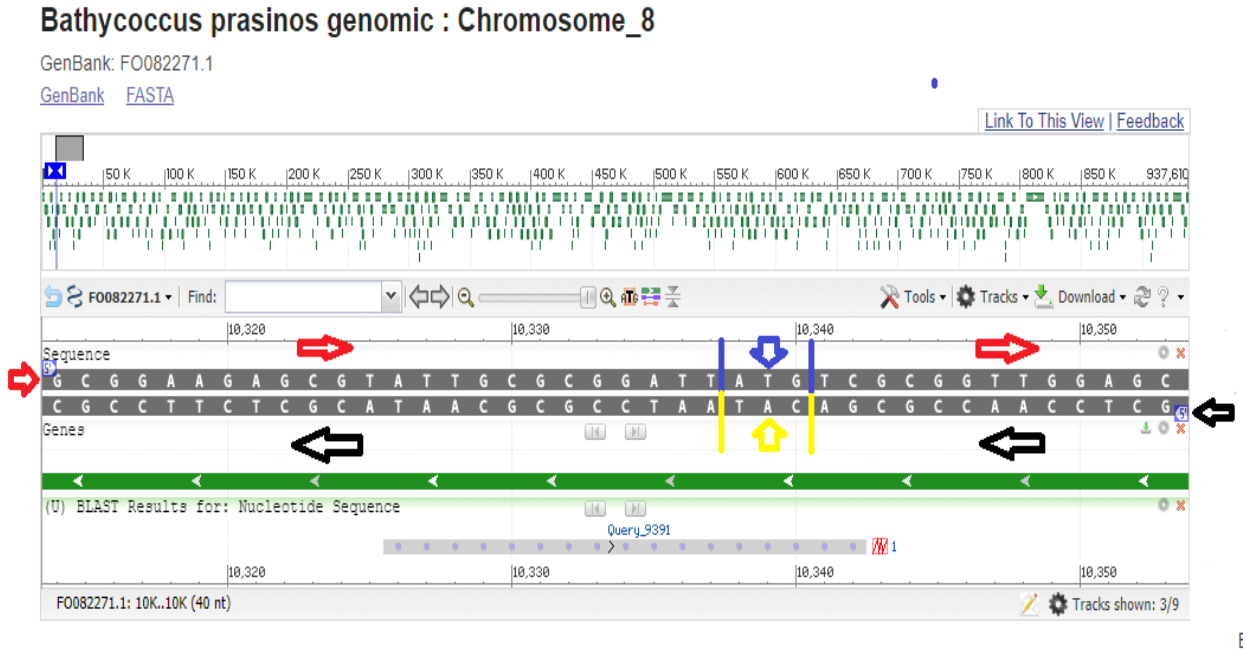


Fig. 2.17

One more detail to clarify before concluding the paragraph. In Fig. 2.12 on page 35 and in the table with active links on the next page, the letter **R** is shown in **purple**. This color of the letter indicates to the reader that in the ***Sbjct* sequence** (the sequence resulting from the significant alignment with **Query**) the **amino acid** that follows **H** (Histidin) is **Arginine** (Arginine, **Arg / R**), but that the last base of the codon (the third base) of Arginine (base that is necessary to encode **R**) is not present in the **Query sequence**, because the **Query sequence** has a *different base* from the **Sbjct Sequence** or simply because it *is shorter* and, therefore, missing.

In the case of our example, the **Query sequence** has a *different base* than the ***Sbjct* Sequence**. While in the **Query sequence** the base "**TC**" follows the base "**C**", in the ***Sbjct* Sequence** the base "**G**" follows (as you can see in the **chain placed at the top** in Fig. 2.17 after the ATG codon), the necessary base to encode amino acid **Arginine** (DNA codon "TCG" = Codon RNA "CGA" = **R**).

In the table with **active links** shown below, you can gain some experience on the contents covered in this paragraph. For this purpose, I also recommend creating random sequences, putting them in BLAST and practicing (both with DNA and RNA sequences) to identify the coding codons following the instructions provided.

Bathycoccus prasinos genomic : Chromosome_8
Sequence ID: FO082271.1 **Length: 937610 Number of Matches: 1**
Related Information
Genome Data Viewer-aligned genomic context
Range 1: 10326 to 10342 GenBank Graphics Next Match Previous Match

Alignment statistics for match #1

Score	Expect	Identities	Gaps	Strand
34.2 bits(17)	36	17/17(100%)	0/17(0%)	Plus/Plus

```
CDS: Putative 1        5        N  R  P  N  H
Query                  1        ATTGCGCGGATTATGTC  17
                                |||||||||||||||||
Sbjct                  10326    ATTGCGCGGATTATGTC  10342
CDS:hypothetical pro   71       N  R  P  N  H  R
```

2.2 HOW TO USE THE RESULTS OF A BASIC BLAST RESEARCH

BLAST research constitutes the **FOURTH PHASE** of the **complete procedure** that has been proposed in *Introduction to the Trichromatic Theory of Systems Equilibrium applied to the Analysis and Modification of DNA or RNA Sequences* (http://www.ttesystems.eu/Cms/files/Pdf/Introduzione%20DNA%20ENG.pdf).
The **Fourth Phase** is preceded by the phases of the **Analysis of the Original Sequence** (the sequence we intend to modify), the **Modification of the Original Sequence** and the **Comparison of the Graphic Representations** of the **Original Sequence** with those of the **New Sequences** obtained from its modifications.

As mentioned in the **Preamble**, unfortunately the pandemic forces me to publish a partially incomplete writing with a reduced amount of insights and internal links. Indeed, this paper lacks, among other things, the *comparison of the four main graphic representations* of the **original sequence** [the **RNA** sequence (**GTGAAATGGTCATGTGTGGCGG**) which is the object of this study on **SARS-COV-2**] with those of the **new sequences** (new sequences that I will present in subsequent chapters after teaching the procedure for *modifying the original sequence*), as well as the analyses of the *similarities and differences* of the *four main graphic representations* are missing.
Only after the carrying out of the **Three Phases** described above, it would have been necessary to carry out the **BLAST searches**, respectively on the **original sequence** and on all **new sequences** [see DNA or RNA Sequence Analysis and Modification through T_T_E_S_ (Chapter I° - First Part); DNA or RNA Sequence Analysis and Modification through T_T_E_S_ (Chapter I° - Second Part)].

As explained in the previous paragraph, using **BLAST** it is possible to compare the *nucleotide sequences* or the *protein sequences* of the DNA or RNA molecules. A **BLAST** re*search* allows the researcher to compare a *sequence of interest* with a database of *already known sequences* (nucleotide or protein) and to identify, among the latter, those that have **similarities** with the *sequence of interest*.

In this work, the main purpose of the **BLAST research** is therefore to identify, for the **original sequence** (the **SARS-COV-2 RNA** sequence) and for the **new sequences** (the **new sequences** obtained from the modifications of the *original sequence*), all *significant alignments* with all of the sequences and *organisms*, which are stored in the BLAST database.

As we saw in the previous paragraph, the information that can be obtained from *significant alignments* are many and all very important.
Below I summarize the main information, which is also the object of further investigation in the subsequent **Phases** of the complete procedure proposed in the *Introduction to the Trichromatic Theory of Systems Equilibrium applied to the Analysis and Modification of DNA or RNA Sequences*:

- possibility of identifying sequences producing **significant alignments** with the **sequence we have entered** (the s*equence of interest* or *Query*), as shown in Fig. 2.7 (b) of the previous paragraph.
- possibility to visualize the **alignment of all the bases of the identified sequences**, as shown in Fig. 2.8 (a) of the previous paragraph.
- possibility of selecting a **specific sequence** and **displaying the complete information sheet** with the selected sequence, as shown in Fig. 2.8 (b) of the previous paragraph;
- possibility to view the information sheet which is **specific** to the **Gene** of the **Query** sequence (see Figures 2.10 and 2.11 shown in the previous paragraph). Among other important information, we note the dates of **publication**, **submission** for publication and the **references** of who published. In addition, the following information are also shown: **1)** a part or the **entire sequence that we have entered** (**Query**); **2)** the **Product** (for example, a *hypothetical protein*)) of the **specific sequence** selected among the sequences identified by the *significant alignments*; **3)** the **complete sequence** of **amino acids** that encodes the *product* obtained.
- possibility to view the sequence of **amino acids** corresponding to a part or to the entire **sequence that we have entered** (for example, the *NRPNH* sequence) and the **Identities** (in Identities) between *Query* and *Sbjct* (the sequence resulting from the significant alignment with Query), as shown in Fig. 2.12 in the previous paragraph.
- possibility to view the **dynamic Graph** (as shown in Fig. 2.13 in the previous paragraph) in which *two sequences running in the opposite direction* are shown, **the sequence placed at the top** (*Sequence*, which coincides with **Query**) that follows the direction $5' \rightarrow 3'$ and the one **at the bottom** (*Genes*, which coincides with a **specific sequence** identified by the *significant alignments*) flowing in the opposite or antisense direction ($3' \rightarrow 5'$).
- possibility of easily identifying, in at the bottom of the graph of Fig. 2.15 and 2.16, the codons and amino acids corresponding to the **Query** sequence.

The BLAST RESEARCH PHASE is preparatory to the subsequent Phases, first of all the **FIFTH PHASE**, that is, the phase of **DISCOVERY** and **HIGHLIGHTING** of the "**organisms**" that are *in common* between the **original sequence** (in our case, the **SARS-COV-2 RNA** sequence) and the **new sequences** (the *new sequences* obtained from the modifications of the *original sequence*).

Let's make an example concerning the BLAST RESEARCH of the **SARS-COV-2 RNA** sequence, which is the object of this study.
As we will see in chapter IV, the "**organism**" (among all the "**organisms**") highlighted by the *BLAST* RESEARCH of the **SARS-COV-2 RNA** sequence (**GTGAAATGGTCATGTGTGGCGG**) is only **Homo Sapiens** (the Human Being), this is because the sequence under study has been found only in humans.

"**Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2**", diagnosed by the presence of the "**GTGAAATGGTCATGTGTGGCGG**" sequence, does not concern organisms other than **Homo Sapiens**, at least until now.
In conclusion, the only "**organisms**" (among all the "**organisms**") that can be **discovered** and **highlighted** as the one *in common* between the **original sequence** (the **SARS-COV-2 RNA** sequence) and the **new sequences** (those obtained from the modifications of the *original sequence*) will be, if they are identified, only **Homo Sapiens**.

On the contrary, if (as we will see in chapter V) we consider **two sequences** almost totally identical to the sequence "**GTGAAATGGTCATGTGTGGCGG**", that is, those that differ ONLY for one basis, such as the sequences "**GTGAAATGGTTATGTGTGGCGG**" and "**GTGAGATGGTCATGTGTG**", things radically change.
The BLAST research made on the first sequence ("**GTGAAATGGTTATGTGTGGCGG**") **does not highlight** the presence of Homo Sapiens among the "**organisms**" identified, except when the search is extended up to the mutation of the base highlighted in red, respectively the **T** with the **C**.
The "organisms" identified by the BLAST research on this sequence are: the ***Bat***, the ***Malayan*** and ***Chinese Pangolin*** (species of anteaters) and the ***Little Egret*** (a species of bird).
Instead, the "organisms" identified by the BLAST research on the "**GTGAGATGGTCATGTGTGGCGG**" sequence are: ***Homo Sapiens***, ***Rhinolophus Sinicus*** (called *Horseshoe*) and ***Rhinolophus Pusillus*** and ***Rhinolophus Monoceros*** (which are variants of the species of the ***Bat***), ***Altererythrobacter namhicola*** (a Gram-negative bacterium), ***Cebus Imitator*** (Panamanian white-faced Capuchin, a species of monkey) and ***Timema bartmani*** (a species of stick insect in the Timematidae family and its variant) e ***Timema Tahoe*.**
As you can see, the **Bat** is the only "organism" that the **two sequences** have in common, this is because this mammalian species can have one or the other sequence in its genome ("**GTGAAATGGTTATGTGTGGCGG**" or "**GTGAGATGGTCATGTGTGGCGG**").

This result of the BLAST research indicates that the change, more or less significant (in our example, the change was minimal because it involved ONLY a base for both the **new sequences** generated), of an **original sequence** (specifically **GTGAAATGGTCATGTGTGGCGG**) can generate **new sequences** whose significant alignments identified by the search identify ***common organisms*** not among the **original sequence** and the **new sequences**, but ONLY among the **new sequences**.

This implies the need to make *comparative analyses* among the alignments of the **new sequences** generated and compare both the identified "***common organisms***" and the **PRODUCTS** (for example proteins, but not only. For the concept of "Product" see the previous paragraph) related to their sequences.

This aspect concerns the SIXTH PHASE of the **complete procedure**, that is that of the **Product Specification of the Significant Alignment**, in which it is necessary to **SPECIFY**, for all the "organisms" that are *in common* with the **original sequence** (or that could be in common only among the **new sequences**), the "**Product**" of the *significant alignment* and which *DNA (or Rna) modified bases* they have in common.

As already mentioned in the *Introduction*, the results of this phase are very important because (along with the results of the SUBSEQUENT PHASES, **Bibliographic Research Targeted on a single Organism** or **on two or more Organisms** and **Concluding Considerations and Pragmatic Choices**) on them will depend the eventual choice to intervene concretely in the modification of the **biological material** of the analyzed sequences for the realization of the different and possible practical purposes.

Before concluding this chapter, it is essential to underline that over time, as we have been told by the BLAST DATABASE managers themselves, it is possible to find **"inconsistencies and subsequent variations in *significant alignments* ..." and they ".... must be attributed to regular updates and / or changes due to the moving parts of this complex DATABASE".**

CHAPTER III

FROM BATS TO HUMAN BEINGS THROUGH PANGOLINS: THE RdRp GENE AND THE RNA SEQUENCE "GTGAAATGGTCATGTGTGGCGG"

GOALS OF CHAPTER III

In paragraph 3.1 of this chapter, it is presented a short, but very recent, **bibliographic research** that investigates the hypothesis that there may be a transmission relationship of *SARS-CoV* from *bats* to *pangolins* and from the latter to *humans*.
In paragraph 3.2, it is described the importance of the **RdRp gene** or ***R****NA-****d****ependent* ***R****NA* ***p****olymerase* of *SARS-CoV-2*; clarified the reason for which the sequence of **22 bases of RNA** "**GTGAAATGGTCATGTGTGGCGG**"; the reasons that justify its multiple modifications and broadly explained the development plan of the next chapters.

3.1 THE MOST ACCREDITED HYPOTHESIS: FROM BATS TO HUMAN BEINGS THROUGH PANGOLINS

The abstract of "**Genomic characterization and infectivity of a novel SARS-like coronavirus in Chinese bats"** [(JOURNAL Emerg Microbes Infect 7 (1), 154 (**2018**) PMCID: PMC6135831; DOI: 10.1038/s41426-018-0155-5], says: "***SARS coronavirus (SARS-CoV)****, the causative agent of the large SARS outbreak in 2003, originated in* ***bats****. Many SARS-like coronaviruses (SL-CoVs) have been detected in bats, particularly those that reside in China, Europe, and Africa. To further understand the evolutionary relationship between SARS-CoV and its reservoirs, 334 bats were collected from Zhoushan city, Zhejiang province, China, between 2015 and 2017. PCR amplification of the conserved* ***coronaviral protein RdRp*** *detected coronaviruses in 26.65% of bats belonging to this region, and this number was influenced by seasonal changes..… Importantly, using pathogenic tests, we found that the virus can reproduce and cause disease in suckling rats, and further studies showed that the virus-like particles can be observed in the brains of suckling rats by electron microscopy. Thus, this study increased our understanding of the genetic diversity of the SL-CoVs carried by bats and also provided a new perspective to study the possibility of cross-species transmission of SL-CoVs using suckling rats as an animal model."*.

Here, it has been advanced the hypothesis that there may be a transmission relationship of *SARS-CoV* from *bats* to *pangolins* and from the latter to *humans*.

In "**Molecular evolution and phylogenetic analysis of SARS-CoV-2 and hosts ACE2 protein suggest Malayan pangolin as intermediary host**", Braz J Microbiol. 2020 Dec; 51(4):1593-1599. Published: **26 June 2020**. PMCID: PMC7319214 DOI: 10.1007/s42770-020-00321-1, the authors concluded that*"...pangolin has become an opportune host to intermediates bat-to-human SARS-CoV-2 jump and entry."*.

In "**An update on the origin of SARS-CoV-2: Despite closest identity, bat (RaTG13) and pangolin derived coronaviruses varied in the critical binding site and O-linked glycan residues**", J Med Virol. **2020 Jul 7**;10.1002/jmv.26261. doi: 10.1002/jmv.26261. PMCID: PMC7361880 DOI: 10.1002/jmv.26261, the authors raise doubts and claim that *"...the initial source's transmission routes of SARS‐CoV‐2 remain obscure and controversial.....Only* ***Serine*** *(Ser) in pangolin and both* ***Threonine*** *(Thr) and* ***Serine*** *(Ser) O‐linked glycans were seen in RaTG13, suggesting that a detailed study needed in pangolin (Manis javanica) and bat (Rhinolophus affinis) related CoV."*.

On the contrary, in "**Conservation analysis of SARS-CoV-2 spike suggests complicated viral adaptation history from bat to human**", Evol Med Public Health. **2020 Nov 5**; 2020(1):290-303. PMCID: PMC7665476 DOI: 10.1093/emph/eoaa041 it is concluded *"…that the spike sequence of SARS-CoV-2 is the result of multiple recombination events during its transmission from bat to human, and we propose a framework of evolutionary history that resolve the relationship of BCoV-RaTG13 and pangolin coronaviruses with SARS-CoV-2."*.

For the authors of "**Computational Analysis of SARS-CoV-2 and SARS-Like Coronavirus Diversity in Human, Bat and Pangolin Populations**" **2020 Dec 30**;13(1):E49. PMID: 33396801 DOI: 10.3390/v13010049*,"…The evolutionary origins of the virus remain elusive and understanding its complex mutational signatures could guide vaccine design and development…..we have characterised several high impact variants (in-frame insertion/deletion or stop gain) in bat-CoV and pangolin-CoV populations, some of which are found in the same amino acid position and may be highlighting loci of potential functional relevance."*.

Finally, in "**Novel coronavirus disease (COVID-19) pandemic: A recent mini review**", Comput Struct Biotechnol J. **2020 Dec 31.** PMCID: PMC7773542 DOI: 10.1016/j.csbj.2020.12.033, according to the authors*"…Phylogenetic analysis of the genomes suggests that the possible occurrence of recombination between SARS-like-CoVs from pangolin and bat might have led to the origin of SARS-CoV-2 and the COVID-19 outbreak…"* and end by saying that *"…SARS-CoV-2 infection, with a heavy viral load in the body, destroys the human lungs through cytokine storm, especially in elderly persons and people with immunosuppressed disorders. A number of drugs have been repurposed and employed, but still, no specific antiviral medicine has been approved by the FDA to treat this disease. This review provides a current status of the COVID-19, epidemiology, an overview of phylogeny, mode of action, diagnosis, and possible treatment methods and vaccines."*.

In conclusion, **at the present moment, there seems to be no certainty on the veracity of the of the most accredited hypothesis of a transmission relationship of *SARS-CoV* from *bats* to *pangolins* and from the latter to *human beings.***

3.2 THE RdRp GENE AND THE RNA SEQUENCE "GTGAAATGGTCATGTGTGGCGG" OF SARS-COV-2

The **RdRp gene** (or ***R**NA-**d**ependent **R**NA **p**olymerase* of *SARS-CoV-2*), is part of **ORF1ab** (**O**pen **R**eading **F**rame, reading phase that allows to encode a whole *protein*, without encountering premature stop codons and thus forming a truncated protein - https://en.wikipedia.org/wiki/Open_reading_frame) **is specific to SARS-CoV-2** (https://en.wikipedia.org/wiki/Severe_acute_respiratory_syndrome_coronavirus_2)], while the **gene "E"** (SARS-CoV-2 envelope protein) belongs to all **beta coronaviruses** (https://en.wikipedia.org/wiki/COVID-19_testing).

Figures 3.1, 3.2 and 3.3 (presented on the following pages) show **six sequences obtained from 2019-nCoV** which are used as a positive control of the nasopharyngeal swab (https://www.eurosurveillance.org/content/table/10.2807/1560-7917.ES.2020.25.3.2000045.t1?fmt=ahah&fullscreen=true).

The goal of each **virus** is to replicate itself as much as possible. Viruses achieve this goal through the action of **RNA polymerase** (or enzyme that forms *long RNA chains*). In the SARS-CoV-2 infection, during the process called **transcription**, the virus genome is replicated (copied) through **RNA-dependent RNA polymerase** (**RdRp**).
For the above reasons, of the **RdRp Gene** (*pSARS-F*), it has been selected the sequence of **22 bases of RNA "GTGAAATGGTCATGTGTGGCGG"** (indicated by the green arrow in Fig. 3.1), currently used, together with other sequences, as a positive control (in the nasopharyngeal swab) for the presence of *SARS-CoV-2* infection.

All the in-depth work presented in the following chapters concerns ONLY this sequence and its many modifications.

In the initial phase, in order to identify its various *significant alignments* and "*products*", the sequence was the object of a **complete BLAST Research** (see Chapter IV).
Subsequently **two complete BLAST Researches** were performed, respectively on the sequences "**GTGAAATGGTTATGTGTGGCGG**" " (RNA sequence of the RdRP Gene of the "*Bat coronavirus BM48-31 / BGR / 2008, complete genome*", see Chapter V) and "**GTGAGATGGTCATGTGTGGCGG**" (sequence of RNA of the RdRP Gene of the *"Bat SARS-like coronavirus isolate bat-SL-CoVZC45, complete genome"*, see Chapter V), indicated by the two red arrows and the blue one in Fig. 3.1. These two sequences differ from the **"GTGAAATGGTCATGTGTGGCGG"** sequence ONLY for one basis (respectively, the **T** instead of **C** and the **G** instead of **A**). As mentioned in paragraph 2.2 of Chapter II and as we will see in Chapter V, these sequences are important because the simple variation of these bases (indicated in red in the two sequences) allows to highlight *significant alignments* with specific "organisms" and to make hypotheses about their relationships, in particular between the ***Bat***, the ***Pangolin*** and the ***Human Being***.

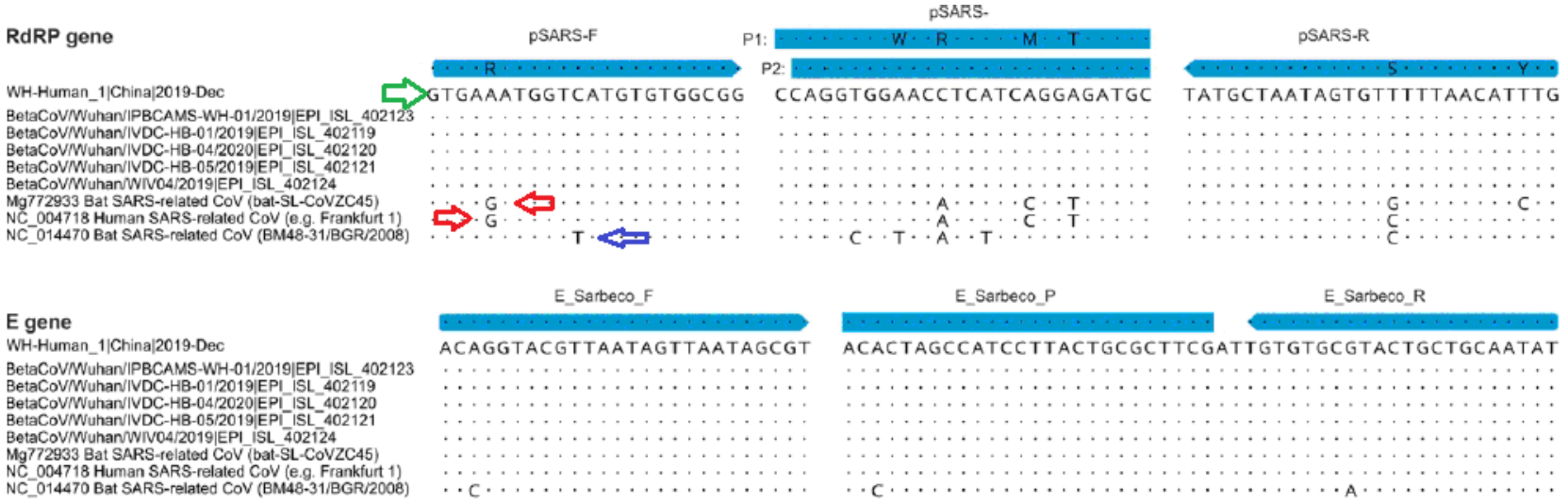


Figure 2 Partial alignments of oligonucleotide binding regions. Panels show six available sequences of 2019-nCoV, aligned to the corresponding partial sequences of SARS-CoV strain Frankfurt 1, which can be used as a positive control for all three RT-PCR assays. The alignment also contains the most closely-related bat virus (Bat SARS-related CoV isolate bat-SL-CoVZC45, GenBank Acc.No. MG772933.1) as well as the most distant member within the SARS-related bat CoV clade, detected in Bulgaria (GenBank Acc. No. NC_014470). Dots represent identical nucleotides compared to sequence Wuhan-Hu 1. Substitutions are specified. More comprehensive alignments in the Appendix.

Fig. 3.1

https://www.who.int/docs/default-source/coronaviruse/protocol-v2-1.pdf

Annex:

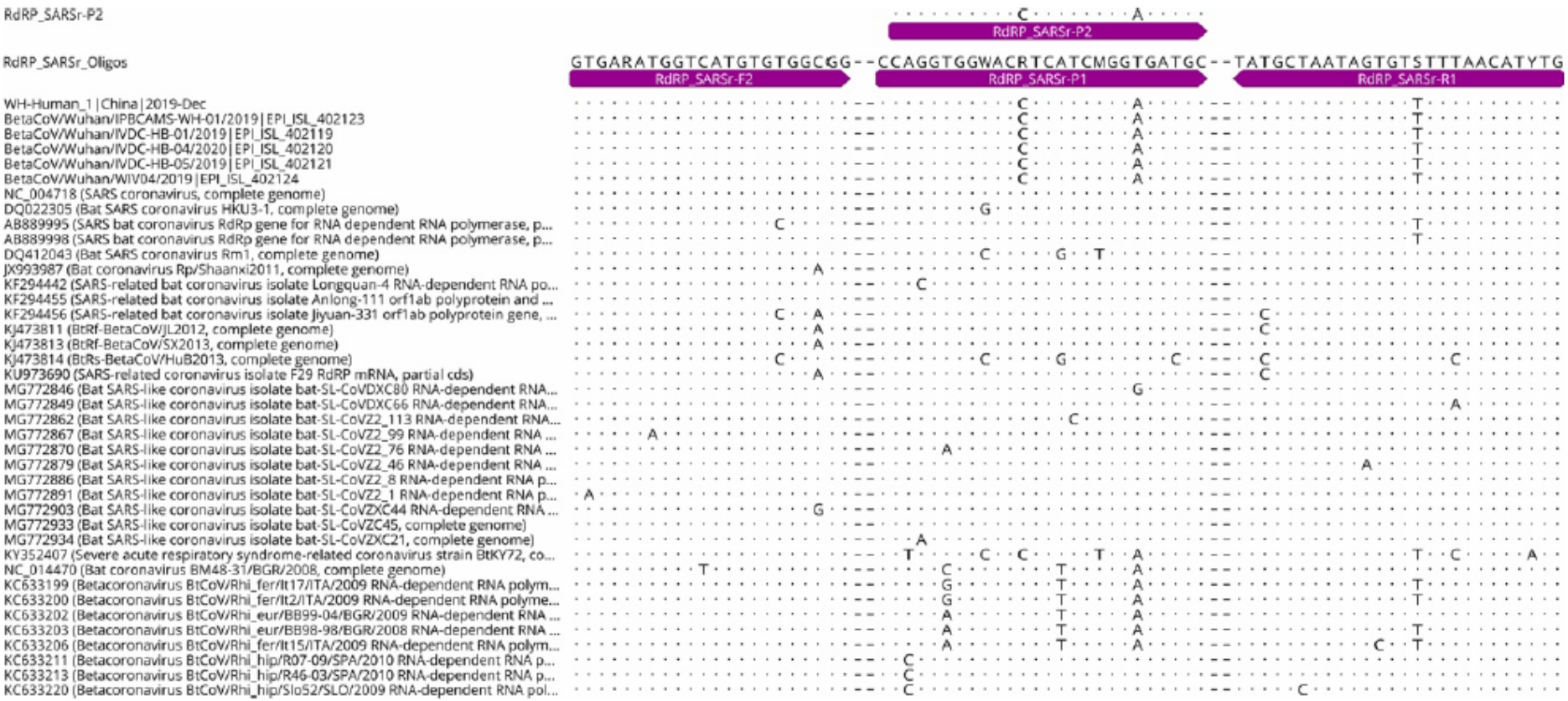
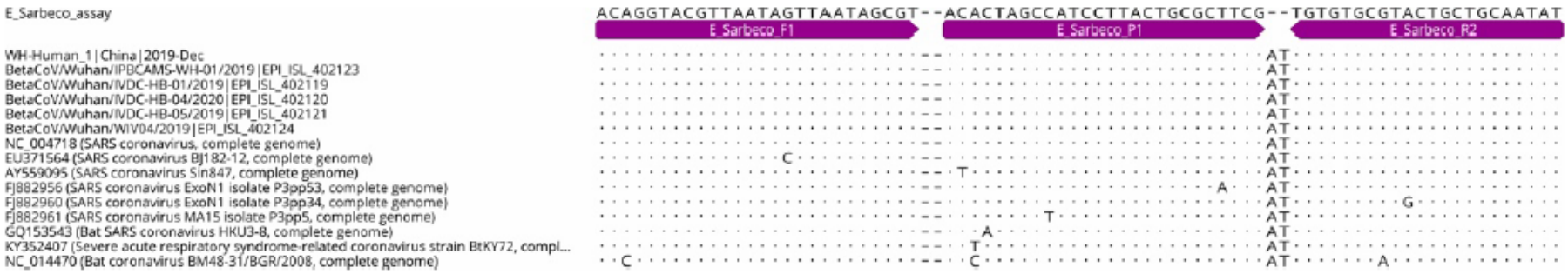


Fig. 3.2

https://www.who.int/docs/default-source/coronaviruse/protocol-v2-1.pdf



Annex figure. Non-redundant alignments of SARS-related CoVs focused on oligonucleotide binding sites of all assays (top to bottom: RdRp, E, N). Viruses not present in these alignments have been removed because their binding sites are 100% identical to one of the members of the alignment. ("--") means sequence gaps not covered by oligonucleotides. Note that these alignments contain only one sequence of 2019-nCoV while Figure 2 above contains all presently released sequences. We will fuse this into one figure.

Fig. 3.3

https://www.who.int/docs/default-source/coronaviruse/protocol-v2-1.pdf

In order not to neglect the impact of any other changes in nitrogenous bases in the same positions highlighted in blue in the sequence **“GTGAAATGGTCATGTGTGGCGG”** and identify any other “organisms” relevant for the purposes of this study, **all the possible base substitutions** were performed, obtaining as a result the following **4 sequences**:

“**GTGAAATGGTAATGTGTGGCGG**”, “**GTGAAATGGTGATGTGTGGCGG**”

“**GTGATATGGTCATGTGTGGCGG**”, “**GTGACATGGTCATGTGTGGCGG**”

These sequences have also been the object of **four *complete Blast searches*** (see Chapter VI).

Having carried out this laborious preliminary work of substitutions of bases to identify all the "organisms" and "products" dependent on these modifications of the sequence **“GTGAAATGGTCATGTGTGGCGG”**, it is possible to begin the modifications of this same sequence through the new methodology proposed in this paper, the one based on the principles of the **TRICHROMATIC THEORY OF EQUILIBRIUM OF SYSTEMS** (**T.T.E.S.**) and presented in Chapters VIII and IX.

This methodology provides for the **replacement of at least 70% of the bases** (but it can also reach 100%), thus obtaining **almost or totally new sequences**.
As we will see in Chapters VIII and IX, the substitution of the bases must necessarily be carried out following the "**Non- Obvious Trends**" of the **original sequence**, the **“GTGAAATGGTCATGTGTGGCGG”** sequence.
The theoretical model proposed here predicts that **all the new sequences** generated by the **original sequence** will have a *close "functional" relationship* with the original sequence because they are generated in strict compliance with its numerous "Non-Obvious Trends".
All the **new sequences generated** will then be subjected to *Blast Research* and used to try to verify the most accredited hypothesis of a transmission relationship of *SARS-CoV* from bats to pangolins and from the latter to humans.
This hypothesis could be more plausible if all or most of the **new sequences** generated by the **original sequence** presented *significant alignments* with the expected "**organisms**", namely the **Bat** and / or the **Pangolin** or, alternatively, with other "**intermediate organisms**", such as those identifiable by the modifications of **ONLY** one base of the sequence “**GTGAAATGGTCATGTGTGGCGG”**.
As already anticipated in the **Summary** and in the **Preamble**, this work of modifications of the **original sequence** reserved unexpected and sensational surprises, in particular the presence of *significant alignments* corresponding to a sequence with unknown characteristics of the ***Human Immunodeficiency Virus (HIV-1)*** and those corresponding to different sequences of the ***Anthrax Bacillus (Bacillus Anthracis)***, an infectious gram-positive bacterium, aetiological agent of ***anthrax*** in animals and ***anthrax*** in humans.

CHAPTER IV

BLAST Research on the RNA SEQUENCE "GTGAAATGGTCATGTGTGGCGG" of the RdRp Gene (RdRp_SARSr-F) (WH-Human_1|China|2019-Dec)

GOALS OF CHAPTER IV

This chapter presents the complete *Blast Research* performed on the **22 base** RNA sequence **"GTGAAATGGTCATGTGTGGCGG"**, a sequence taken from the **RdRp Gene** (*pSARS-F*) and currently used, along with other sequences, as a positive control (in the nasophraryngeal swab) for the presence of *SARS-CoV-2* infection.
As previously mentioned, the **RdRp Gene** ("RNA-dependent RNA polymerase") is part of the gene **"ORF1ab", an open reading frame** 1ab of SARS-CoV-2.
The **RdRp Gene** is important because through the action of **RNA dependent RNA polymerase (RdRp)** the virus (*SARS-CoV-2*) responsible for **Covid-19** carries out its replication (copy of itself).
RNA-dependent RNA polymerase (RdRp) is therefore the driving force behind the **transcription** process of *SARS-CoV-2*.
Although there are differences, "***RNA polymerase*** *performs essentially the same function in all cells from bacteria to humans .... transcription begins when RNA polymerase attaches to the* ***promoter,*** *a special region, at the beginning of the gene ... until it reaches a* ***terminator*** *.... The immediate* ***product*** *of* ***transcription*** *is called* ***primary transcript*** *and consists of an RNA that extends from the promoter to the terminator but is very unstable ...* (https://www.tesionline.it/appunti/Biologia-molecolare/Le-RNA-polimerasi-e-il-ciclo-della-trascrizione/255/75).
One of the possible ways to prevent the virus from replicating is to prevent the **Rna-dependent RNA polymerase (RdRp)** from adequately performing its function. Technically this is possible by the introduction of **modified Rna sequences**, so as to sabotage the viral **transcription** phase. However, *SARS-CoV-2* is able to correct **transcription errors** caused by the detection of **modified Rna** and thus to continue replication undisturbed.
Therefore, to achieve the goal of sabotaging the **RdRp**, a sophisticated masking process of the **modified RNA** is required or, alternatively, developing **potent chemical inhibitors** (drugs) of the **RdRp**.

4.1 BLAST Research on the RNA SEQUENCE "GTGAAATGGTCATGTGTGGCGG" of the RdRp Gene (RdRp_SARSr-F) – (Hitlist size: 100)

For the reasons explained in the previous page and in paragraph 3.2 of Chapter III, of the **RdRp Gene** *(pSARS-F)*, it has been selected the **22 base RNA sequence "GTGAAATGGTCATGTGTGGCGG"**. A **complete Blast Research** was performed on this sequence and, for further studies, the first sequence of the list of the first **100** *significant alignments* found was selected (the one highlighted in bold).

FOR THE BODY OF INFORMATION USEFUL TO INTERPRET THE DATA PRESENTED IN THE NEXT PAGES, PLEASE SEE CHAPTER 2.

Query: **GTGAAATGGTCATGTGTGGCGG**

Search Parameters	
Program	blastn
Word size	7
Expect value	1000
Hitlist size	100
Match/Mismatch scores	1,-3
Gapcosts	5,2
Filter string	F
Genetic Code	1

100 sequences selected:

Accession Seq	Description	Comomm Name	Max Score	Total Score	Query C.	E value	Per. Id.	Acc. Len
MW411920.1	**Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/VA-DCLS-2538/2020, complete genome**	**Severe acute respiratory syndrome coronavirus 2**	44.1	44.1	100%	0.037	100.00%	29899
MW411919.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/VA-DCLS-2537/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	44.1	44.1	100%	0.037	100.00%	29859
MW411899.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/VA-DCLS-2291/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	44.1	44.1	100%	0.037	100.00%	29881
MW411918.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/VA-DCLS-2199/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	44.1	44.1	100%	0.037	100.00%	29865
MW411898.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/VA-DCLS-2290/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	44.1	44.1	100%	0.037	100.00%	29881
MW411917.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/VA-DCLS-2198/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	44.1	44.1	100%	0.037	100.00%	29900
MW411897.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/VA-DCLS-2287/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	44.1	44.1	100%	0.037	100.00%	29901

MW411916.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/VA-DCLS-2197/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	44.1	44.1	100%	0.037	100.00%	29823
MW411896.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/VA-DCLS-2285/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	44.1	44.1	100%	0.037	100.00%	29867
MW411915.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/VA-DCLS-2196/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	44.1	44.1	100%	0.037	100.00%	29879
MW411895.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/VA-DCLS-2281/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	44.1	44.1	100%	0.037	100.00%	29901
MW411914.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/VA-DCLS-2195/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	44.1	44.1	100%	0.037	100.00%	29866
MW411894.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/VA-DCLS-2277/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	44.1	44.1	100%	0.037	100.00%	29901
MW411913.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/VA-DCLS-2194/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	44.1	44.1	100%	0.037	100.00%	29862
MW411893.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/VA-DCLS-2276/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	44.1	44.1	100%	0.037	100.00%	29898
MW411912.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/VA-DCLS-2193/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	44.1	44.1	100%	0.037	100.00%	29859
MW411911.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/VA-DCLS-2163/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	44.1	44.1	100%	0.037	100.00%	29888
MW411892.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/VA-DCLS-2273/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	44.1	44.1	100%	0.037	100.00%	29899
MW411910.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/VA-DCLS-2161/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	44.1	44.1	100%	0.037	100.00%	29880
MW411891.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/VA-DCLS-2272/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	44.1	44.1	100%	0.037	100.00%	29866
MW411909.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/VA-DCLS-2160/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	44.1	44.1	100%	0.037	100.00%	29895
MW411890.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/VA-DCLS-2271/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	44.1	44.1	100%	0.037	100.00%	29899
MW411908.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/VA-DCLS-2159/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	44.1	44.1	100%	0.037	100.00%	29873
MW411889.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/VA-DCLS-2269/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	44.1	44.1	100%	0.037	100.00%	29865
MW411907.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/VA-DCLS-2158/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	44.1	44.1	100%	0.037	100.00%	29866
MW411888.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/VA-DCLS-2267/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	44.1	44.1	100%	0.037	100.00%	29899
MW411906.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/VA-DCLS-2157/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	44.1	44.1	100%	0.037	100.00%	29873
MW411887.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-	Severe acute respiratory	44.1	44.1	100%	0.037	100.00%	29869

	2/human/USA/VA-DCLS-2266/2020, complete genome	syndrome coronavirus 2						
MW411905.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/VA-DCLS-2155/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	44.1	44.1	100%	0.037	100.00%	29892
MW411886.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/VA-DCLS-2265/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	44.1	44.1	100%	0.037	100.00%	29892
MW411904.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/VA-DCLS-2154/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	44.1	44.1	100%	0.037	100.00%	29899
MW411885.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/VA-DCLS-2264/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	44.1	44.1	100%	0.037	100.00%	29894
MW411903.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/VA-DCLS-2153/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	44.1	44.1	100%	0.037	100.00%	29855
MW411884.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/VA-DCLS-2263/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	44.1	44.1	100%	0.037	100.00%	29881
MW411902.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/VA-DCLS-2149/2020 ORF1ab polyprotein (ORF1ab), ORF1a polyprotein (ORF1ab), surface glycoprotein (S), ORF3a protein (ORF3a), envelope protein (E), membrane glycoprotein (M), ORF6 protein (ORF6), ORF7a protein (ORF7a), ORF7b (ORF7b), ORF8 protein (ORF8), nucleocapsid phosphoprotein (N), and ORF10 protein (ORF10) genes, complete cds	Severe acute respiratory syndrome coronavirus 2	44.1	44.1	100%	0.037	100.00%	29897
MW411883.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/VA-DCLS-2259/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	44.1	44.1	100%	0.037	100.00%	29868
MW411901.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/VA-DCLS-2147/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	44.1	44.1	100%	0.037	100.00%	29867
MW411882.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/VA-DCLS-2258/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	44.1	44.1	100%	0.037	100.00%	29867
MW411900.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/VA-DCLS-2145/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	44.1	44.1	100%	0.037	100.00%	29865
MW411881.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/VA-DCLS-2257/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	44.1	44.1	100%	0.037	100.00%	29901
MW411946.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/CA-LACPHL-AF00026/2020 ORF1ab polyprotein (ORF1ab), ORF1a polyprotein (ORF1ab), surface glycoprotein (S), ORF3a protein (ORF3a), envelope protein (E), and membrane glycoprotein (M) genes, complete cds; ORF6 protein (ORF6) and ORF7a protein (ORF7a) genes, partial cds; and ORF7b (ORF7b), ORF8 protein (ORF8), nucleocapsid phosphoprotein (N), and ORF10 protein (ORF10) genes, complete cds	Severe acute respiratory syndrome coronavirus 2	44.1	44.1	100%	0.037	100.00%	29847
MW411880.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/VA-DCLS-2256/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	44.1	44.1	100%	0.037	100.00%	29903
MW411945.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/CA-LACPHL-AF00025/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	44.1	44.1	100%	0.037	100.00%	29842

MW411944.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/CA-LACPHL-AF00024/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	44.1	44.1	100%	0.037	100.00%	29859
MW411943.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/CA-LACPHL-AF00023/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	44.1	44.1	100%	0.037	100.00%	29827
MW411942.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/CA-LACPHL-AF00022/2020 ORF1ab polyprotein (ORF1ab), ORF1a polyprotein (ORF1ab), surface glycoprotein (S), ORF3a protein (ORF3a), envelope protein (E), membrane glycoprotein (M), ORF6 protein (ORF6), ORF7a protein (ORF7a), ORF7b (ORF7b), ORF8 protein (ORF8), nucleocapsid phosphoprotein (N), and ORF10 protein (ORF10) genes, complete cds	Severe acute respiratory syndrome coronavirus 2	44.1	44.1	100%	0.037	100.00%	29866
MW411879.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/VA-DCLS-2255/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	44.1	44.1	100%	0.037	100.00%	29891
MW411941.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/CA-LACPHL-AF00021/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	44.1	44.1	100%	0.037	100.00%	29897
MW411940.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/CA-LACPHL-AF00020/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	44.1	44.1	100%	0.037	100.00%	29823
MW411939.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/CA-LACPHL-AF00019/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	44.1	44.1	100%	0.037	100.00%	29844
MW411938.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/CA-LACPHL-AF00018/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	44.1	44.1	100%	0.037	100.00%	29860
MW411878.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/VA-DCLS-2254/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	44.1	44.1	100%	0.037	100.00%	29865
MW411937.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/CA-LACPHL-AF00017/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	44.1	44.1	100%	0.037	100.00%	29826
MW411936.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/CA-LACPHL-AF00016/2020 ORF1ab polyprotein (ORF1ab), ORF1a polyprotein (ORF1ab), surface glycoprotein (S), ORF3a protein (ORF3a), envelope protein (E), membrane glycoprotein (M), ORF6 protein (ORF6), ORF7a protein (ORF7a), ORF7b (ORF7b), ORF8 protein (ORF8), nucleocapsid phosphoprotein (N), and ORF10 protein (ORF10) genes, complete cds	Severe acute respiratory syndrome coronavirus 2	44.1	44.1	100%	0.037	100.00%	29830
MW411935.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/CA-LACPHL-AF00015/2020 ORF1ab polyprotein (ORF1ab), ORF1a polyprotein (ORF1ab), surface glycoprotein (S), ORF3a protein (ORF3a), envelope protein (E), membrane glycoprotein (M), ORF6 protein (ORF6), ORF7a protein (ORF7a), ORF7b (ORF7b), ORF8 protein (ORF8), nucleocapsid phosphoprotein (N), and ORF10 protein (ORF10) genes, complete cds	Severe acute respiratory syndrome coronavirus 2	44.1	44.1	100%	0.037	100.00%	29824
MW411934.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/CA-LACPHL-AF00014/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	44.1	44.1	100%	0.037	100.00%	29858
MW411877.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/VA-DCLS-2253/2020,	Severe acute respiratory syndrome	44.1	44.1	100%	0.037	100.00%	29893

	complete genome	coronavirus 2						
MW411933.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/CA-LACPHL-AF00013/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	44.1	44.1	100%	0.037	100.00%	29858
MW411932.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/CA-LACPHL-AF00012/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	44.1	44.1	100%	0.037	100.00%	29851
MW411931.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/CA-LACPHL-AF00011/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	44.1	44.1	100%	0.037	100.00%	29892
MW411930.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/CA-LACPHL-AF00010/2020 ORF1ab polyprotein (ORF1ab), ORF1a polyprotein (ORF1ab), surface glycoprotein (S), ORF3a protein (ORF3a), envelope protein (E), membrane glycoprotein (M), ORF6 protein (ORF6), ORF7a protein (ORF7a), ORF7b (ORF7b), ORF8 protein (ORF8), nucleocapsid phosphoprotein (N), and ORF10 protein (ORF10) genes, complete cds	Severe acute respiratory syndrome coronavirus 2	44.1	44.1	100%	0.037	100.00%	29823
MW411876.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/VA-DCLS-2252/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	44.1	44.1	100%	0.037	100.00%	29900
MW411929.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/CA-LACPHL-AF00009/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	44.1	44.1	100%	0.037	100.00%	29851
MW411928.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/CA-LACPHL-AF00008/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	44.1	44.1	100%	0.037	100.00%	29857
MW411927.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/CA-LACPHL-AF00007/2020 ORF1ab polyprotein (ORF1ab), ORF1a polyprotein (ORF1ab), surface glycoprotein (S), ORF3a protein (ORF3a), envelope protein (E), membrane glycoprotein (M), ORF6 protein (ORF6), and ORF7a protein (ORF7a) genes, complete cds; ORF7b (ORF7b) and ORF8 protein (ORF8) genes, partial cds; and nucleocapsid phosphoprotein (N) and ORF10 protein (ORF10) genes, complete cds	Severe acute respiratory syndrome coronavirus 2	44.1	44.1	100%	0.037	100.00%	29831
MW411926.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/CA-LACPHL-AF00006/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	44.1	44.1	100%	0.037	100.00%	29852
MW411875.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/VA-DCLS-2251/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	44.1	44.1	100%	0.037	100.00%	29903
MW411925.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/CA-LACPHL-AF00005/2020 ORF1ab polyprotein (ORF1ab), ORF1a polyprotein (ORF1ab), surface glycoprotein (S), ORF3a protein (ORF3a), envelope protein (E), membrane glycoprotein (M), ORF6 protein (ORF6), ORF7a protein (ORF7a), ORF7b (ORF7b), ORF8 protein (ORF8), nucleocapsid phosphoprotein (N), and ORF10 protein (ORF10) genes, complete cds	Severe acute respiratory syndrome coronavirus 2	44.1	44.1	100%	0.037	100.00%	29861
MW411924.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/CA-LACPHL-AF00004/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	44.1	44.1	100%	0.037	100.00%	29831
MW411923.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/CA-LACPHL-AF00003/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	44.1	44.1	100%	0.037	100.00%	29846

MW411922.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/CA-LACPHL-AF00002/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	44.1	44.1	100%	0.037	100.00%	29845
MW411874.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/VA-DCLS-2250/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	44.1	44.1	100%	0.037	100.00%	29853
MW411921.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/CA-LACPHL-AF00001/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	44.1	44.1	100%	0.037	100.00%	29823
MW411873.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/VA-DCLS-2249/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	44.1	44.1	100%	0.037	100.00%	29873
MW411872.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/VA-DCLS-2192/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	44.1	44.1	100%	0.037	100.00%	29865
MW411871.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/VA-DCLS-2191/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	44.1	44.1	100%	0.037	100.00%	29899
MW411656.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/VA-DCLS-2512/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	44.1	44.1	100%	0.037	100.00%	29858
MW411655.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/VA-DCLS-2511/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	44.1	44.1	100%	0.037	100.00%	29893
MW411654.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/VA-DCLS-2510/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	44.1	44.1	100%	0.037	100.00%	29899
MW411653.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/VA-DCLS-2509/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	44.1	44.1	100%	0.037	100.00%	29893
MW411652.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/VA-DCLS-2508/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	44.1	44.1	100%	0.037	100.00%	29893
MW411651.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/VA-DCLS-2507/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	44.1	44.1	100%	0.037	100.00%	29869
MW411650.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/VA-DCLS-2505/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	44.1	44.1	100%	0.037	100.00%	29902
MW411649.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/VA-DCLS-2503/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	44.1	44.1	100%	0.037	100.00%	29866
MW411648.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/VA-DCLS-2333/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	44.1	44.1	100%	0.037	100.00%	29867
MW411647.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/VA-DCLS-2332/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	44.1	44.1	100%	0.037	100.00%	29867
MW411646.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/VA-DCLS-2330/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	44.1	44.1	100%	0.037	100.00%	29867
MW411645.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/VA-DCLS-2329/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	44.1	44.1	100%	0.037	100.00%	29899
MW411644.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/VA-DCLS-2328/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	44.1	44.1	100%	0.037	100.00%	29903
MW411643.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/VA-DCLS-2326/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	44.1	44.1	100%	0.037	100.00%	29901
MW411642.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-	Severe acute respiratory	44.1	44.1	100%	0.037	100.00%	29868

	2/human/USA/VA-DCLS-2325/2020, complete genome	syndrome coronavirus 2						
MW411641.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/VA-DCLS-2324/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	44.1	44.1	100%	0.037	100.00%	29870
MW411640.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/VA-DCLS-2323/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	44.1	44.1	100%	0.037	100.00%	29895
MW411639.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/VA-DCLS-2322/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	44.1	44.1	100%	0.037	100.00%	29884
MW411638.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/VA-DCLS-2319/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	44.1	44.1	100%	0.037	100.00%	29868
MW411637.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/VA-DCLS-2318/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	44.1	44.1	100%	0.037	100.00%	29865
MW411636.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/VA-DCLS-2317/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	44.1	44.1	100%	0.037	100.00%	29902
MW411635.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/VA-DCLS-2316/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	44.1	44.1	100%	0.037	100.00%	29899
MW411634.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/VA-DCLS-2311/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	44.1	44.1	100%	0.037	100.00%	29865
MW411633.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/VA-DCLS-2309/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	44.1	44.1	100%	0.037	100.00%	29865

ALIGNMENT VIEW FLAT QUERY-ANCHORED WITH LETTERS FOR IDENTITIES

Query range 1: 1 to 22

Query	1	GTGAAATGGTCATGTGTGGCGG	22
MW411920.1	15427	GTGAAATGGTCATGTGTGGCGG	15448
MW411919.1	15429	GTGAAATGGTCATGTGTGGCGG	15450
MW411899.1	15421	GTGAAATGGTCATGTGTGGCGG	15442
MW411918.1	15393	GTGAAATGGTCATGTGTGGCGG	15414
MW411898.1	15409	GTGAAATGGTCATGTGTGGCGG	15430
MW411917.1	15428	GTGAAATGGTCATGTGTGGCGG	15449
MW411897.1	15429	GTGAAATGGTCATGTGTGGCGG	15450
MW411916.1	15393	GTGAAATGGTCATGTGTGGCGG	15414
MW411896.1	15395	GTGAAATGGTCATGTGTGGCGG	15416
MW411915.1	15407	GTGAAATGGTCATGTGTGGCGG	15428
MW411895.1	15429	GTGAAATGGTCATGTGTGGCGG	15450
MW411914.1	15394	GTGAAATGGTCATGTGTGGCGG	15415
MW411894.1	15429	GTGAAATGGTCATGTGTGGCGG	15450
MW411913.1	15427	GTGAAATGGTCATGTGTGGCGG	15448
MW411893.1	15426	GTGAAATGGTCATGTGTGGCGG	15447
MW411912.1	15429	GTGAAATGGTCATGTGTGGCGG	15450
MW411911.1	15416	GTGAAATGGTCATGTGTGGCGG	15437
MW411892.1	15427	GTGAAATGGTCATGTGTGGCGG	15448
MW411910.1	15408	GTGAAATGGTCATGTGTGGCGG	15429
MW411891.1	15394	GTGAAATGGTCATGTGTGGCGG	15415
MW411909.1	15423	GTGAAATGGTCATGTGTGGCGG	15444
MW411890.1	15427	GTGAAATGGTCATGTGTGGCGG	15448
MW411908.1	15401	GTGAAATGGTCATGTGTGGCGG	15422
MW411889.1	15394	GTGAAATGGTCATGTGTGGCGG	15415
MW411907.1	15394	GTGAAATGGTCATGTGTGGCGG	15415
MW411888.1	15427	GTGAAATGGTCATGTGTGGCGG	15448
MW411906.1	15401	GTGAAATGGTCATGTGTGGCGG	15422
MW411887.1	15397	GTGAAATGGTCATGTGTGGCGG	15418
MW411905.1	15420	GTGAAATGGTCATGTGTGGCGG	15441
MW411886.1	15420	GTGAAATGGTCATGTGTGGCGG	15441
MW411904.1	15427	GTGAAATGGTCATGTGTGGCGG	15448
MW411885.1	15422	GTGAAATGGTCATGTGTGGCGG	15443
MW411903.1	15425	GTGAAATGGTCATGTGTGGCGG	15446
MW411884.1	15409	GTGAAATGGTCATGTGTGGCGG	15430
MW411902.1	15425	GTGAAATGGTCATGTGTGGCGG	15446
MW411883.1	15396	GTGAAATGGTCATGTGTGGCGG	15417
MW411901.1	15395	GTGAAATGGTCATGTGTGGCGG	15416
MW411882.1	15395	GTGAAATGGTCATGTGTGGCGG	15416
MW411900.1	15393	GTGAAATGGTCATGTGTGGCGG	15414
MW411881.1	15429	GTGAAATGGTCATGTGTGGCGG	15450
MW411946.1	15414	GTGAAATGGTCATGTGTGGCGG	15435
MW411880.1	15431	GTGAAATGGTCATGTGTGGCGG	15452
MW411945.1	15407	GTGAAATGGTCATGTGTGGCGG	15428
MW411944.1	15429	GTGAAATGGTCATGTGTGGCGG	15450
MW411943.1	15396	GTGAAATGGTCATGTGTGGCGG	15417
MW411942.1	15394	GTGAAATGGTCATGTGTGGCGG	15415
MW411879.1	15419	GTGAAATGGTCATGTGTGGCGG	15440
MW411941.1	15425	GTGAAATGGTCATGTGTGGCGG	15446
MW411940.1	15395	GTGAAATGGTCATGTGTGGCGG	15416
MW411939.1	15415	GTGAAATGGTCATGTGTGGCGG	15436
MW411938.1	15429	GTGAAATGGTCATGTGTGGCGG	15450
MW411878.1	15393	GTGAAATGGTCATGTGTGGCGG	15414
MW411937.1	15396	GTGAAATGGTCATGTGTGGCGG	15417

MW411936.1	15400	GTGAAATGGTCATGTGTGGCGG	15421
MW411935.1	15393	GTGAAATGGTCATGTGTGGCGG	15414
MW411934.1	15423	GTGAAATGGTCATGTGTGGCGG	15444
MW411877.1	15421	GTGAAATGGTCATGTGTGGCGG	15442
MW411933.1	15426	GTGAAATGGTCATGTGTGGCGG	15447
MW411932.1	15421	GTGAAATGGTCATGTGTGGCGG	15442
MW411931.1	15420	GTGAAATGGTCATGTGTGGCGG	15441
MW411930.1	15394	GTGAAATGGTCATGTGTGGCGG	15415
MW411876.1	15428	GTGAAATGGTCATGTGTGGCGG	15449
MW411929.1	15421	GTGAAATGGTCATGTGTGGCGG	15442
MW411928.1	15427	GTGAAATGGTCATGTGTGGCGG	15448
MW411927.1	15396	GTGAAATGGTCATGTGTGGCGG	15417
MW411926.1	15421	GTGAAATGGTCATGTGTGGCGG	15442
MW411875.1	15431	GTGAAATGGTCATGTGTGGCGG	15452
MW411925.1	15431	GTGAAATGGTCATGTGTGGCGG	15452
MW411924.1	15396	GTGAAATGGTCATGTGTGGCGG	15417
MW411923.1	15416	GTGAAATGGTCATGTGTGGCGG	15437
MW411922.1	15425	GTGAAATGGTCATGTGTGGCGG	15446
MW411874.1	15425	GTGAAATGGTCATGTGTGGCGG	15446
MW411921.1	15393	GTGAAATGGTCATGTGTGGCGG	15414
MW411873.1	15401	GTGAAATGGTCATGTGTGGCGG	15422
MW411872.1	15393	GTGAAATGGTCATGTGTGGCGG	15414
MW411871.1	15427	GTGAAATGGTCATGTGTGGCGG	15448
MW411656.1	15395	GTGAAATGGTCATGTGTGGCGG	15416
MW411655.1	15421	GTGAAATGGTCATGTGTGGCGG	15442
MW411654.1	15427	GTGAAATGGTCATGTGTGGCGG	15448
MW411653.1	15421	GTGAAATGGTCATGTGTGGCGG	15442
MW411652.1	15421	GTGAAATGGTCATGTGTGGCGG	15442
MW411651.1	15397	GTGAAATGGTCATGTGTGGCGG	15418
MW411650.1	15430	GTGAAATGGTCATGTGTGGCGG	15451
MW411649.1	15394	GTGAAATGGTCATGTGTGGCGG	15415
MW411648.1	15395	GTGAAATGGTCATGTGTGGCGG	15416
MW411647.1	15395	GTGAAATGGTCATGTGTGGCGG	15416
MW411646.1	15395	GTGAAATGGTCATGTGTGGCGG	15416
MW411645.1	15427	GTGAAATGGTCATGTGTGGCGG	15448
MW411644.1	15431	GTGAAATGGTCATGTGTGGCGG	15452
MW411643.1	15429	GTGAAATGGTCATGTGTGGCGG	15450
MW411642.1	15396	GTGAAATGGTCATGTGTGGCGG	15417
MW411641.1	15398	GTGAAATGGTCATGTGTGGCGG	15419
MW411640.1	15423	GTGAAATGGTCATGTGTGGCGG	15444
MW411639.1	15412	GTGAAATGGTCATGTGTGGCGG	15433
MW411638.1	15396	GTGAAATGGTCATGTGTGGCGG	15417
MW411637.1	15393	GTGAAATGGTCATGTGTGGCGG	15414
MW411636.1	15430	GTGAAATGGTCATGTGTGGCGG	15451
MW411635.1	15427	GTGAAATGGTCATGTGTGGCGG	15448
MW411634.1	15393	GTGAAATGGTCATGTGTGGCGG	15414
MW411633.1	15393	GTGAAATGGTCATGTGTGGCGG	15414

Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/VA-DCLS-2538/2020, complete genome

GenBank: MW411920.1

FASTA Graphics

Go to:

LOCUS MW411920 22 bp RNA linear VRL **28-DEC-2020**

DEFINITION Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/VA-DCLS-2538/2020, complete genome.

ACCESSION MW411920 REGION: 15427..15448

VERSION MW411920.1

KEYWORDS .

SOURCE **Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)**

ORGANISM Severe acute respiratory syndrome coronavirus 2

Viruses; Riboviria; Orthornavirae; Pisuviricota; Pisoniviricetes; Nidovirales; Cornidovirineae; Coronaviridae; Orthocoronavirinae; Betacoronavirus; Sarbecovirus.

REFERENCE 1 (bases 1 to 22)

AUTHORS Fink,L.

TITLE Direct Submission

JOURNAL **Submitted (28-DEC-2020) Department of General Services Division of Consolidated Laboratory Services, Virginia Division of Consolidated Laboratory Services Sequencing Submission Group, 500 North 5th Street, Richmond, VA 23219, USA**

COMMENT ##Assembly-Data-START##

Assembly Method :: minimap2 v. 2.17; ivar v. 1.1

Sequencing Technology :: Illumina

##Assembly-Data-END##

FEATURES Location/Qualifiers

source 1..22

/organism="Severe acute respiratory syndrome coronavirus 2"

/mol_type="genomic RNA"

/isolate="**SARS-CoV-2/human/USA/VA-DCLS-2538/2020**"

/host="**Homo sapiens**"

/db_xref="taxon:2697049"

/country="**USA: Virginia**"

/collection_date="2020-11"

gene <1..>22

/gene="ORF1ab"

CDS <1..>22

/gene="**ORF1ab**"

/ribosomal slippage

/codon_start=3

/product="ORF1ab polyprotein"

/protein_id="QQG32947.1"

/translation="MESLVPGFNEKTHVQLSLPVLQVRDVLVRGFGDSVEEVLSEARQ
HLKDGTCGLVEVEKGVLPQLEQPYVFIKRSDARTAPHGHVMVELVAELEGIQYGRSGE
TLGVLVPHVGEIPVAYRKVLLRKNGNKGAGGHSYGADLKSFDLGDELGTDPYEDFQEN
WNTKHSSGVTRELMRELNGGAYTRYVDNNFCGPDGYPLECIKDLLARAGKASCTLSEQ
LDFIDTKRGVYCCREHEHEIAWYTERSEKSYELQTPFEIKLAKKFDIFNGECPNFVFP
LNSIIKTIQPRVEKKKLDGFMGRIRSVYPVASPNECNQMCLSTLMKCDHCGETSWQTG
DFVKATCEFCGTENLTKEGATTCGYLPQNAVVKIYCPACHNSEVGPEHSLAEYHNESG
LKTILRKGGRTIAFGGCVFSYVGCHNKCAYWVPRASANIGCNHTGVVGEGSEGLNDNL
LEILQKEKVNINIVGDFKLNEEIAIILASFSASTSAFVETVKGLDYKAFKQIVESCGN
FKVTKGKAKKGAWNIGEQKSILSPLYAFASEAARVVRSIFSRTLETAQNSVRVLQKAA
ITILDGISQYSLRLIDAMMFTSDLATNNLVVMAYITGGVVQLTSQWLTNIFGTVYEKL
KPVLDWLEEKFKEGVEFLRDGWEIVKFISTCACEIVGGQIVTCAKEIKESVQTFFKLV
NKFLALCADSIIIGGAKLKALNLGETFVTHSKGLYRKCVKSREETGLLMPLKAPKEII
FLEGETLPTEVLTEEVVLKTGDLQPLEQPTSEAVEAPLVGTPVCINGLMLLEIKDTEK
YCALAPNMMVTNNTFTLKGGAPTKVTFGDDTVIEVQGYKSVNITFELDERIDKVLNEK
CSAYTVELGTEVNEFACVVADAVIKTLQPVSELLTPLGIDLDEWSMATYYLFDESGEF
KLASHMYCSFYPPDEDEEEGDCEEEEFEPSTQYEYGTEDDYQGKPLEFGATSAALQPE
EEQEEDWLDDDSQQTVGQQDGSEDNQTTTIQTIVEVQPQLEMELTPVVQTIEVNSFSG
YLKLTDNVYIKNADIVEEAKKVKPTVVVNAANVYLKHGGGVAGALNKATNNAMQVESD
DYIATNGPLKVGGSCVLSGHNLAKHCLHVVGPNVNKGEDIQLLKSAYENFNQHEVLLA
PLLSAGIFGADPIHSLRVCVDTVRTNVYLAVFDKNLYDKLVSSFLEMKSEKQVEQKIA
EIPKEEVKPFITESKPSVEQRKQDDKKIKACVEEVTTTLEETKFLTENLLLYIDINGN
LHPDSATLVSDIDITFLKKDAPYIVGDVVQEGVLTAVVIPTKKAGGTTEMLAKALRKV
PTDNYITTYPGQGLNGYTVEEAKTVLKKCKSAFYILPSIISNEKQEILGTVSWNLREM
LAHAEETRKLMPVCVETKAIVSTIQRKYKGIKIQEGVVDYGARFYFYTSKTTVASLIN
TLNDLNETLVTMPLGYVTHGLNLEEAARYMRSLKVPATVSVSSPDAVTAYNGYLTSSS
KTPEEHFIETISLAGSYKDWSYSGQSTQLGIEFLKRGDKSVYYTSNPTTFHLDGEVIT
FDNLKTLLSLREVRTIKVFTTVDNINLHTQVVDMSMTYGQQFGPTYLDGADVTKIKPH
NSHEGKTFYVLPNDDTLRVEAFEYYHTTDPSFLGRYMSALNHTKKWKYPQVNGLTSIK
WADNNCYLATALLTLQQIELKFNPPALQDAYYRARAGEAANFCALILAYCNKTVGELG
DVRETMSYLFQHANLDSCKRVLNVVCKTCGQQQTTLKGVEAVMYMGTLSYEQFKKGVQ
IPCTCGKQATKYLVQQESPFVMMSAPPAQYELKHGTFTCASEYTGNYQCGHYKHITSK
ETLYCIDGALLTKSSEYKGPITDVFYKENSYTTTIKPVTYKLDGVVCTEIDPKLDNYY
KKDNSYFTEQPIDLVPNQPYPNASFDNFKFVCDNIKFADDLNQLTGYKKPASRELKVT
FFPDLNGDVVAIDYKHYTPSFKKGAKLLHKPIVWHVNNATNKATYKPNTWCIRCLWST
KPVETSNSFDVLKSEDAQGMDNLACEDLKPVSEEVVENPTIQKDVLECNVKTTEVVGD
IILKPANNSLKITEEVGHTDLMAAYVDNSSLTIKKPNELSRVLGLKTLATHGLAAVNS
VPWDTIANYAKPFLNKVVSTTTNIVTRCLNRVCTNYMPYFFTLLLQLCTFTRSTNSRI
KASMPTTIAKNTVKSVGKFCLEASFNYLKSPNFSKLINIIIWFLLLSVCLGSLIYSTA
ALGVLMSNLGMPSYCTGYREGYLNSTNVTIATYCTGSIPCSVCLSGLDSLDTYPSLET
IQITISSFKWDLTAFGLVAEWFLAYILFTRFFYVLGLAAIMQLFFSYFAVHFISNSWL
MWLIINLVQMAPISAMVRMYIFFASFYYVWKSYVHVVDGCNSSTCMMCYKRNRATRVE

CTTIVNGVRRSFYVYANGGKGFCKLHNWNCVNCDTFCAGSTFISDEVARDLSLQFKRP
INPTDQSSYIVDSVTVKNGSIHLYFDKAGQKTYERHSLSHFVNLDNLRANNTKGSLPI
NVIVFDGKSKCEESSAKSASVYYSQLMCQPILLLDQALVSDVGDSAEVAVKMFDAYVN
TFSSTFNVPIEKLKTLVATAEAELAKNVSLDNVLSTFISAARQGFVDSDVETKDVVEC
LKLSHQSDIEVTGDSCNNYMLTYNKVENMTPRDLGACIDCSARHINAQVAKSHNIALI
WNVKDFMSLSEQLRKQIRSAAKKNNLPFKLTCATTRQVVNVVTTKIALKGGKIVNNWL
KQLIKVTLVFLFVAAIFYLITPVHVMSKHTDFSSEIIGYKAIDGGVTRDIASTDTCFA
NKHADFDTWFSQRGGSYTNDKACPLIAAVITREVGFVVPGLPGTILRTTNGDFLHFLP
RVFSAVGNICYTPSKLIEYTDFATSACVLAAECTIFKDASGKPVPYCYDTNVLEGSVA
YESLRPDTRYVLMDGSIIQFPNTYLEGSVRVVTTFDSEYCRHGTCERSEAGVCVSTSG
RWVLNNDYYRSLPGVFCGVDAVNLLTNMFTPLIQPIGALDISASIVAGGIVAIVVTCL
AYYFMRFRRAFGEYSHVVAFNTLLFLMSFTVLCLTPVYSFLPGVYSVIYLYLTFYLTN
DVSFLAHIQWMVMFTPLVPFWITIAYIICISTKHFYWFFSNYLKRRVVFNGVSFSTFE
EAALCTFLLNKEMYLKLRSDVLLPLTQYNRYLALYNKYKYFSGAMDTTSYREAACCHL
AKALNDFSNSGSDVLYQPPQTSITSAVLQSGFRKMAFPSGKVEGCMVQVTCGTTTLNG
LWLDDVVYCPRHVICTSEDMLNPNYEDLLIRKSNHNFLVQAGNVQLRVIGHSMQNCVL
KFKVDTANPKTPKYKFVRIQPGQTFSVLACYNGSPSGVYQCAMRPNFTIKGSFLNGSC
GSVGFNIDYDCVSFCYMHHMELPTGVHAGTDLEGNFYGPFVDRQTAQAAGTDTTITVN
VLAWLYAAVINGDRWFLNRFTTTLNDFNLVAMKYNYEPLTQDHVDILGPLSAQTGIAV
LDMCASLKELLQNGMNGRTILGSALLEDEFTPFDVVRQCSGVTFQSAVKRTIKGTHHW
LLLTILTSLLVLVQSTQWSLFFFLYENAFLPFAMGIIAMSAFAMMFVKHKHAFLCLFL
LPSLATVAYFNMVYMPASWVMRIMTWLDMVDTSLSGFKLKDCVMYASAVVLLILMTAR
TVYDDGARRVWTLMNVLTLVYKVYYGNALDQAISMWALIISVTSNYSGVVTTVMFLAR
GIVFMCVEYCPIFFITGNTLQCIMLVYCFLGYFCTCYFGLFCLLNRYFRLTLGVYDYL
VSTQEFRYMNSQGLLPPKNSIDAFKLNIKLLGVGGKPCIKVATVQSKMSDVKCTSVVL
LSVLQQLRVESSSKLWAQCVQLHNDILLAKDTTEAFEKMVSLLSVLLSMQGAVDINKL
CEEMLDNRATLQAIASEFSSLPSYAAFATAQEAYEQAVANGDSEVVLKKLKKSLNVAK
SEFDRDAAMQRKLEKMADQAMTQMYKQARSEDKRAKVTSAMQTMLFTMLRKLDNDALN
NIINNARDGCVPLNIIPLTTAAKLMVVIPDYNTYKNTCDGTTFTYASALWEIQQVVDA
DSKIVQLSEISMDNSPNLAWPLIVTALRANSAVKLQNNELSPVALRQMSCAAGTTQTA
CTDDNALAYYNTTKGGRFVLALLSDLQDLKWARFPKSDGTGTIYTELEPPCRFVTDTP
KGPKVKYLYFIKGLNNLNRGMVLGSLAATVRLQAGNATEVPANSTVLSFCAFAVDAAK
AYKDYLASGGQPITNCVKMLCTHTGTGQAITVTPEANMDQESFGGASCCLYCRCHIDH
PNPKGFCDLKGKYVQIPTTCANDPVGFTLKNTVCTVCGMWKGYGCSCDQLREPMLQSA
DAQSFLNRVCGVSAARLTPCGTGTSTDVVYRAFDIYNDKVAGFAKFLKTNCCRFQEKD
EDDNLIDSYFVVKRHTFSNYQHEETIYNLLKDCPAVAKHDFFKFRIDGDMVPHISRQR
LTKYTMADLVYALRHFDEGNCDTLKEILVTYNCCDDDYFNKKDWYDFVENPDILRVYA
NLGERVRQALLKTVQFCDAMRNAGIVGVLTLDNQDLNGNWYDFGDFIQTTPGSGVPVV
DSYYSLLMPILTLTRALTAESHVDTDLTKPYIKWDLLKYDFTEERLKLFDRYFKYWDQ
TYHPNCVNCLDDRCILHCANFNVLFSTVFPLTSFGPLVRKIFVDGVPFVVSTGYHFRE
LGVVHNQDVNLHSSRLSFKELLVYAADPAMHAASGNLLLDKRTTCFSVAALTNNVAFQ
TVKPGNFNKDFYDFAVSKGFFKEGSSVELKHFFFAQDGNAAISDYDYYRYNLPTMCDI
RQLLFVVEVVDKYFDCYDGGCINANQVIVNNLDKSAGFPFNKWGKARLYYDSMSYEDQ
DA
LFAYTKRNVIPTITQMNLKYAISAKNRARTVAGVSICSTMTNRQFHQKLLKSIAAT

RGATVVIGTSKFYGGWHNMLKTVYSDVENPHLMGWDYPKCDRAMPNMLRIMASLVLAR
KHTTCCSLSHRFYRLANECAQVLSEMVMCGGSLYVKPGGTSSGDATTAYANSVFNICQ
AVTANVNALLSTDGNKIADKYVRNLQHRLYECLYRNRDVDTDFVNEFYAYLRKHFSMM
ILSDDAVVCFNSTYASQGLVASIKNFKSVLYYQNNVFMS
EAKCWTETDLTKGPHEFCS
QHTMLVKQGDDYVYLPYPDPSRILGAGCFVDDIVKTDGTLMIERFVSLAIDAYPLTKH
PNQEYADVFHLYLQYIRKLHDELTGHMLDMYSVMLTNDNTSRYWEPEFYEAMYTPHTV
LQAVGACVLCNSQTSLRCGACIRRPFLCCKCCYDHVISTSHKLVLSVNPYVCNAPGCD
VTDVTQLYLGGMSYYCKSHKPPISFPLCANGQVFGLYKNTCVGSDNVTDFNAIATCDW
TNAGDYILANTCTERLKLFAAETLKATEETFKLSYGIATVREVLSDRELHLSWEVGKP
RPPLNRNYVFTGYRVTKNSKVQIGEYTFEKGDYGDAVVYRGTTTYKLNVGDYFVLTSH
TVMPLSAPTLVPQEHYVRITGLYPTLNISDEFSSNVANYQKVGMQKYSTLQGPPGTGK
SHFAIGLALYYPSARIVYTACSHAAVDALCEKALKYLPIDKCSRIIPARARVECFDKF
KVNSTLEQYVFCTVNALPETTADIVVFDEISMATNYDLSVVNARLRAKHYVYIGDPAQ
LPAPRTLLTKGTLEPEYFNSVCRLMKTIGPDMFLGTCRRCPAEIVDTVSALVYDNKLK
AHKDKSAQCFKMFYKGVITHDVSSAINRPQIGVVREFLTRNPAWRKAVFISPYNSQNA
VASKILGLPTQTVDSSQGSEYDYVIFTQTTETAHSCNVNRFNVAITRAKVGILCIMSD
RDLYDKLQFTSLEIPRRNVATLQAENVTGLFKDCSKVITGLHPTQAPTHLSVDTKFKT
EGLCVDIPGIPKDMTYRRLISMMGFKMNYQVNGYPNMFITREEAIRHVRAWIGFDVEG
CHATREAVGTNLPLQLGFSTGVNLVAVPTGYVDTPDNTDFSRVSAKPPPGDQFKHLIP
LMYKGLPWNVVRIKIVQMLSDTLKNLSDGVVFVLWAHGFELTSMKYFVKIGPERTCCL
CDRRATCFSTASDTYACWHHSIGFDYVYNPFMIDVQQWGFTGNLQSNHDLYCQVHGNA
HVASCDAIMTRCLAVHECFVKRVDWTIEYPIIGDELKINAACRKVQHMVVKAALLADK
FPVLHDIGNPKAIKCVPQADVEWKFYDAQPCSDKAYKIEELFYSYATHSDKFTDGVCL
FWNCNVDRYPANSIVCRFDTRVLSNLNLPGCDGGSLYVNKHAFHTPAFDKSAFVNLKQ
LPFFYYSDSPCESHGKQVVSDIDYVPLKSATCITRCNLGGAVCRHHANEYRLYLDAYN
MMISAGFSLWVYKQFDTYNLWNTFTRLQSLENVAFNVVNKGHFDGQQGEVPVSIINNT
VYTKVDGVDVELFENKTTLPVNVAFELWAKRNIKPVPEVKILNNLGVDIAANTVIWDY
KRDAPAHISTIGVCSMTDIAKKPTETICAPLTVFFDGRVDGQVDLFRNARNGVLITEG
SVKGLQPSVGPKQASLNGVTLIGEAVKTQFNYYKKVDGVVQQLPETYFTQSRNLQEFK
PRSQMEIDFLELAMDEFIERYKLEGYAFEHIVYGDFSHSQLGGLHLLIGLAKRFKESP
FELEDFIPMDSTVKNYFITDAQTGSSKCVCSVIDLLLDDFVEIIKSQDLSVVSKVVKV
TIDYTEISFMLWCKDGHVETFYPKLQSSQAWQPGVAMPNFYKMQRMLLEKCDLQNYGD
SATLPKGIMMNVAKYTQLCQYLNTLTLAVPYNMRVIHFGAGSDKGVAPGTAVLRQWLP
TGTLLVDSDLNDFVSDADSTLIGDCATVHTANKWDLIISDMYDPKTKNVTKENDSKEG
FFTYICGFIQQKLALGGSVAIKITEHSWNADLYKLMGHFAWWTAFVTNVNASSSEAFL
IGCNYLGKPCEQIDGYVMHANYIFWRNTNPIQLSSYSLFDMSKFPLKLRGTAVMSLKE
GQINDMILSLLSKGRLIIRENNRVVISSDVLVNN"

mat_peptide <1..>22

/gene="ORF1ab"

/product="RNA-dependent RNA polymerase"

ORIGIN

1 gtgaaatggt catgtgtggc gg

//

ALIGNMENT VIEW PAIRWISE (CDS FEATURE)

Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/VA-DCLS-2538/2020, complete genome.
Sequence ID: MW411920.1 Length: 29899 Number of Matches: 1
Range 1: 15427 to 15448 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
44.1 bits(22)	0.037	22/22(100%)	0/22(0%)	Plus/Plus

```
CDS: Putative 1        1          E  M  V  M  C  G  G
Query                  1      GTGAAATGGTCATGTGTGGCGG  22
                              ||||||||||||||||||||||
Sbjct                  15427  GTGAAATGGTCATGTGTGGCGG  15448
CDS:ORF1ab polyprote   5056   S  E  M  V  M  C  G  G
```

NOTE 1: Pay particular attention to the amino acids M and V (see also Fig. 4.1)
As they will be object of further studies in the next chapters.

NOTA 2: "ORF1ab", open reading frame 1ab of SARS-CoV-2, includes the RdRp Gene ("RNA-dependent RNA polymerase").
The sequence "GTGAAATGGTCATGTGTGGCGG" (object of the present study) is part of /product="RNA-dependent RNA polymerase" and it is used, along with other sequences of the same Gene, in the positive control (in the nasophraryngeal swab) of SARS-CoV-2.

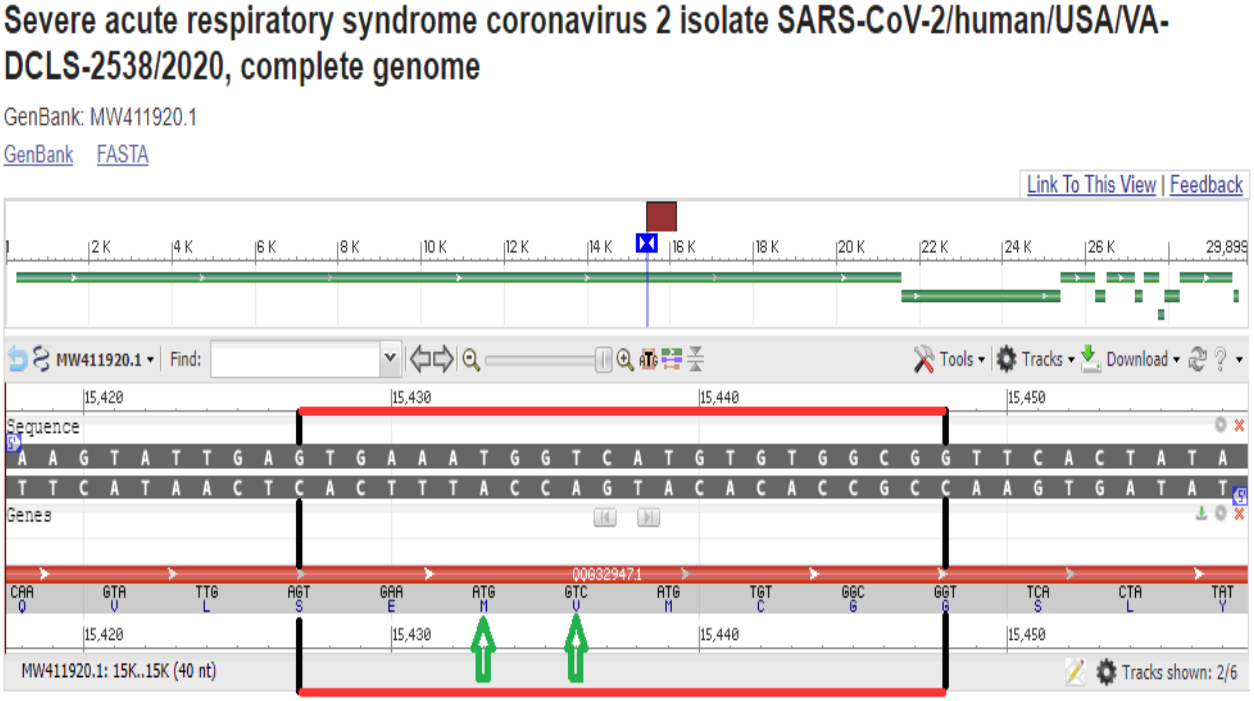


Fig. 4.1

Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/BEL/GHB-03021/2020 orf1ab polyprotein, RdRp region, (orf1ab) gene, partial cds

GenBank: MT072668.1

FASTA Graphics

Go to:

```
LOCUS       MT072668                522 bp    RNA     linear   VRL 06-APR-2020
DEFINITION  Severe acute respiratory syndrome coronavirus 2 isolate
            SARS-CoV-2/human/BEL/GHB-03021/2020 orf1ab polyprotein, RdRp
            region, (orf1ab) gene, partial cds.
ACCESSION   MT072668 REGION: 170..691
VERSION     MT072668.1
KEYWORDS    .
SOURCE      Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)
  ORGANISM  Severe acute respiratory syndrome coronavirus 2
            Viruses; Riboviria; Orthornavirae; Pisuviricota; Pisoniviricetes;
            Nidovirales; Cornidovirineae; Coronaviridae; Orthocoronavirinae;
            Betacoronavirus; Sarbecovirus.
REFERENCE   1  (bases 1 to 522)
  AUTHORS   Wollants,E., Vanmechelen,B., Wawina,T., Van Ranst,M. and Maes,P.
  TITLE     First detection of COVID-19 in Belgium
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 522)
  AUTHORS   Wollants,E., Vanmechelen,B., Wawina,T., Van Ranst,M. and Maes,P.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-FEB-2020) Microbiology, Immunolgy and
            Transplantation, KU Leuven Rega Institute, Herestraat 49 BOX1040,
            Leuven 3000, Belgium
COMMENT     ##Assembly-Data-START##
            Assembly Method       :: KANU v. 1.8; MEDACA v. 0.8
            Coverage              :: 10.000
            Sequencing Technology :: Oxford Nanopore technologies MinION
            ##Assembly-Data-END##
FEATURES             Location/Qualifiers
     source          1..522
                     /organism="Severe acute respiratory syndrome coronavirus
                     2"
                     /mol_type="genomic RNA"
                     /isolate="SARS-CoV-2/human/BEL/GHB-03021/2020"
                     /host="Homo sapiens"
                     /db_xref="taxon:2697049"
                     /country="Belgium"
                     /collection_date="03-Feb-2020"
     gene            <1..>522
                     /gene="orf1ab"
     CDS             <1..>522
```

/gene="orf1ab"
/codon_start=1
/product="**orf1ab polyprotein**"
/protein_id="QIB84671.1"
/translation="LFAYTKRNVIPTITQMNLKYAISAKNRARTVAGVSICSTMTNRQ
FHQKLLKSIAATRGATVVIGTSKFYGGWHNMLKTVYSDVENPHLMGWDYPKCDRAMPN
MLRIMASLVLARKHTTCCSLSHRFYRLANECAQVL**SEMVMCGG**SLYVKPGGTSSGDAT
TAYANSVFNICQAVTANVNALLSTDGNKIADKYVRNLQHRLYECLYRNRDVDTDFVNE
FYAYLRKHFSMMILSDDAVVCFNSTYASQGLVASIKNFKSVLYYQNNVFMS"

mat_peptide <1..>522
/gene="orf1ab"
/product="RdRp"

ORIGIN

```
      1 agaggagcta ctgtagtaat tggaacaagc aaattctatg gtggttggca caacatgtta
     61 aaaactgttt atagtgatgt agaaaaccct caccttatgg gttgggatta tcctaaatgt
    121 gatagagcca tgcctaacat gcttagaatt atggcctcac ttgttcttgc tcgcaaacat
    181 acaacgtgtt gtagcttgtc acaccgtttc tatagattag ctaatgagtg tgctcaagta
    241 ttgagtgaaa tggtcatgtg tggcggttca ctatatgtta aaccaggtgg aacctcatca
    301 ggagatgcca caactgctta tgctaatagt gtttttaaca tttgtcaagc tgtcacggcc
    361 aatgttaatg cacttttatc tactgatggt aacaaaattg ccgataagta tgtccgcaat
    421 ttacaacaca gactttatga gtgtctctat agaaatagag atgttgacac agactttgtg
    481 aatgagtttt acgcatattt gcgtaaacat ttctcaatga tg
```

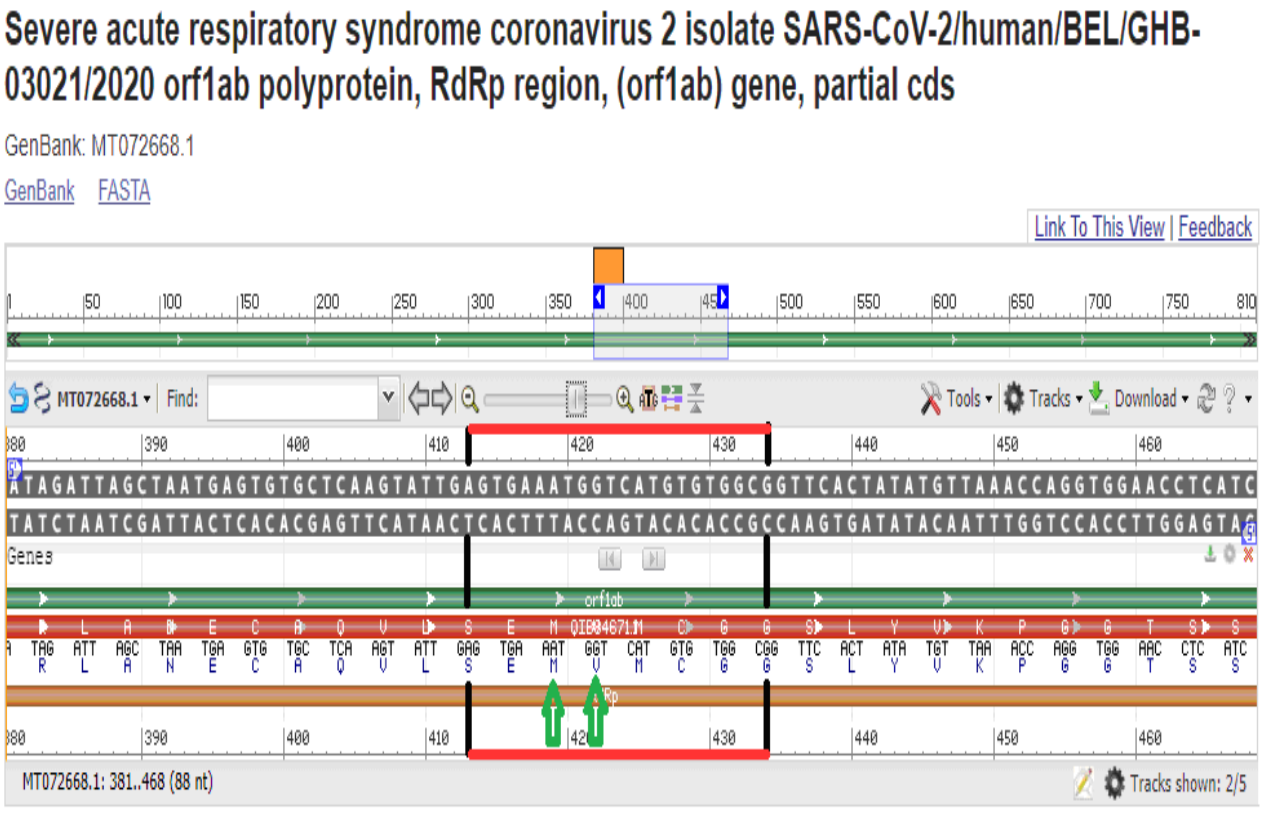


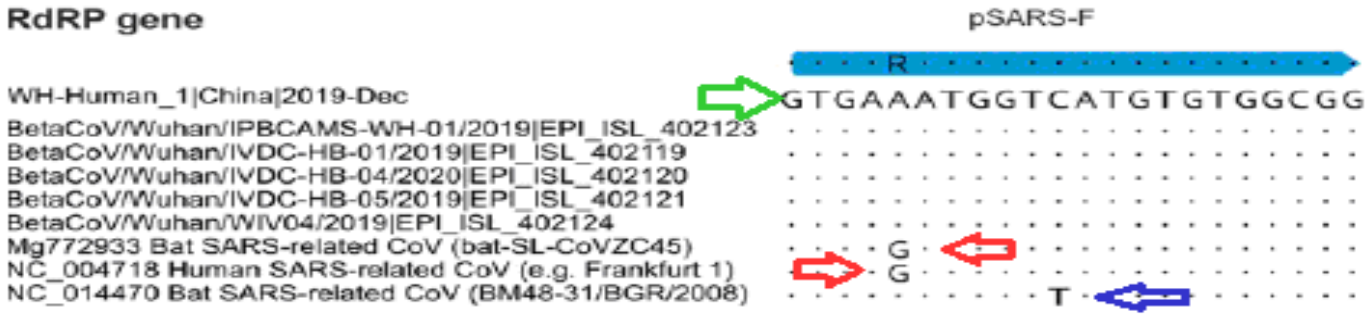
Fig. 4.2 **(Attention: Although the bases are different from those highlighted in Fig.4.1, the amino acids are the same, i. e. M and V)**

CHAPTER V

BLAST RESEARCH ON TWO RNA SPECIFIC SEQUENCES OF THE RdRp GENE

GOALS OF CHAPTER V

In this chapter BLAST searches were carried out on the sequences **two sequences** almost totally identical to the sequence "**GTGAAATGGTCATGTGTGGCGG**" (the sequence considered in the previous paragraph and the subject of study in this document), i. e. those which differ ONLY by a basis, respectively the sequences "**GTGAAATGGTTATGTGTGGCGG**" and "**GTGAGATGGTCATGTGTGGCGG**". As already seen in Chapter III in Fig. 3.1 (of which I shall give the segment concerned below), these two sequences are important because they are the two sequences of the **RdRp gene** (*RNA dependent RNA polymerase*, the engine of the *SARS-CoV-2* **transcription** process) whose significant alignments of BLAST research identify the ***Bat*** and ***Pangolin***.



More precisely, the "organisms" identified by the BLAST research on the sequence "**GTGAAATGGTTATGTGTGGCGG**" are: the ***Bat***, the ***Malay Pangolin*** and the ***Chinese Pangolin*** (species of anteaters) and the ***Litte Egret*** (a species of bird). BLAST research in this sequence **does not show** the presence of **Homo sapiens** among the identified "organisms", except if the research extends to the mutation of the base (the **T** with **C**) highlighted in red.
Instead, the "organisms" identified by the BLAST research on the "**GTGAGATGGTCATGTGTGGCGG**" sequence are: ***Homo Sapiens***, **Rhinolophus Sinicus** (called *Horseshoe*) and ***Rhinolophus Pusillus*** and ***Rhinolophus Monoceros*** (which are variants of the species of the ***Bat***), ***Altererythrobacter namhicola*** (a Gram-negative bacterium), ***Cebus Imitator*** (Panamanian white-faced Capuchin, a species of monkey) and ***Timema bartmani*** (a species of stick insect in the Timematidae family and its variant) e ***Timema Tahoe.***
As you can see, the **Bat** is the only "organism" that the **two sequences** have in common, this is because this mammalian species can have one or the other sequence in its genome ("**GTGAAATGGTTATGTGTGGCGG**" or "**GTGAGATGGTCATGTGTGGCGG**").

This result of the BLAST research indicates that the change, more or less significant, of an **original sequence** ("**GTGAAATGGTCATGTGTGGCGG"**) can generate **new sequences** whose significant alignments identified by the search identify ***common organisms*** not among the **original sequence** and the **new sequences**, but ONLY among the **new sequences** (see also paragraph 2.2 of Chapter II for further information).

5.1 BLAST Research on the RNA SEQUENCE "GTGAAATGGTTATGTGTGGCGG" of the RdRp Gene of the "Bat coronavirus BM48-31/BGR/2008, complete genome" – (Hitlist size: 100)

Query: **GTGAAATGGTTATGTGTGGCGG**

Search Parameters	
Program	blastn
Word size	7
Expect value	1000
Hitlist size	100
Match/Mismatch scores	1,-3
Gapcosts	5,2
Filter string	F
Genetic Code	1

100 sequences selected:

Accession Seq	Description	Comomm Name	Max Score	Total Score	Query C.	E value	Per. Id.	Acc. Len
MN312673.1	**Bat betacoronavirus isolate 7952 RNA-dependent RNA polymerase (RdRp) gene, partial cds**	Bat betacoronavirus	44.1	44.1	100%	0.036	100.00%	402
MN312672.1	**Bat betacoronavirus isolate 7931 RNA-dependent RNA polymerase (RdRp) gene, partial cds**	Bat betacoronavirus	44.1	44.1	100%	0.036	100.00%	402
MN312671.1	**Bat betacoronavirus isolate 7896 RNA-dependent RNA polymerase (RdRp) gene, partial cds**	Bat betacoronavirus	44.1	44.1	100%	0.036	100.00%	591
MN312636.1	**Bat betacoronavirus isolate 7924 RNA-dependent RNA polymerase (RdRp) gene, partial cds**	Bat betacoronavirus	44.1	44.1	100%	0.036	100.00%	589
MN312635.1	**Bat betacoronavirus isolate 7921 RNA-dependent RNA polymerase (RdRp) gene, partial cds**	Bat betacoronavirus	44.1	44.1	100%	0.036	100.00%	588
MN312634.1	**Bat betacoronavirus isolate 7909 RNA-dependent RNA polymerase (RdRp) gene, partial cds**	Bat betacoronavirus	44.1	44.1	100%	0.036	100.00%	592
MN312633.1	**Bat betacoronavirus isolate 7907 RNA-dependent RNA polymerase (RdRp) gene, partial cds**	Bat betacoronavirus	44.1	44.1	100%	0.036	100.00%	597
MN312632.1	**Bat betacoronavirus isolate 7905 RNA-dependent RNA polymerase (RdRp) gene, partial cds**	Bat betacoronavirus	44.1	44.1	100%	0.036	100.00%	566
GU190217.1	**Bat coronavirus BB98-16/BGR/2008 RNA-dependent RNA polymerase gene, partial cds**	Bat coronavirus BB98-16/BGR/2008	44.1	44.1	100%	0.036	100.00%	817
GU190215.1	**Bat coronavirus BM48-31/BGR/2008, complete genome**	Bat coronavirus BM48-31/BGR/2008	44.1	44.1	100%	0.036	100.00%	29276
XM_037013622.1	**PREDICTED: Manis javanica consortin, connexin sorting protein (CNST), transcript variant X6, mRNA**	Malayan pangolin	38.2	38.2	86%	2.2	100.00%	4505
XM_037013621.1	**PREDICTED: Manis javanica consortin, connexin sorting protein (CNST), transcript variant X5, mRNA**	Malayan pangolin	38.2	38.2	86%	2.2	100.00%	4346
XM_037013617.1	**PREDICTED: Manis javanica consortin, connexin sorting protein (CNST), transcript variant X1, mRNA**	Malayan pangolin	38.2	38.2	86%	2.2	100.00%	4929

XM_036918394.1	**PREDICTED: Manis pentadactyla consortin, connexin sorting protein (CNST), transcript variant X3, mRNA**	Chinese pangolin	38.2	38.2	86%	2.2	100.00%	4496
XM_035898451.1	**PREDICTED: Egretta garzetta zinc finger protein 639 (ZNF639), transcript variant X3, mRNA**	little egret	38.2	38.2	86%	2.2	100.00%	4342
XM_035898450.1	**PREDICTED: Egretta garzetta zinc finger protein 639 (ZNF639), transcript variant X2, mRNA**	little egret	38.2	38.2	86%	2.2	100.00%	4199
XM_035898448.1	**PREDICTED: Egretta garzetta zinc finger protein 639 (ZNF639), transcript variant X1, mRNA**	little egret	38.2	38.2	86%	2.2	100.00%	4394
MW295643.1	**Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/TUN/TUN-202058412/2020, complete genome**	Severe acute respiratory syndrome coronavirus 2	36.2	36.2	100%	8.9	95.45%	29856
MW295640.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/TUN/TUN-202055262/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	36.2	36.2	100%	8.9	95.45%	29897
MW295556.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/TUN/TUN-202053873/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	36.2	36.2	100%	8.9	95.45%	29897
MW294011.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/ECU/Z&Z_SARS_4/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	36.2	36.2	100%	8.9	95.45%	29903
MW294007.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/ECU/Z&Z_SARS_3/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	36.2	36.2	100%	8.9	95.45%	29903
MW292665.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/MD-MDH-0406/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	36.2	36.2	100%	8.9	95.45%	29864
MW292664.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/MD-MDH-0405/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	36.2	36.2	100%	8.9	95.45%	29832
MW292663.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/MD-MDH-0404/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	36.2	36.2	100%	8.9	95.45%	29817
MW292662.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/MD-MDH-0403/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	36.2	36.2	100%	8.9	95.45%	29827
MW292661.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/MD-MDH-0402/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	36.2	36.2	100%	8.9	95.45%	29821
MW292660.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/MD-MDH-0401/2020 ORF1ab polyprotein (ORF1ab), ORF1a polyprotein (ORF1ab), surface glycoprotein (S), ORF3a protein (ORF3a), envelope protein (E), membrane glycoprotein (M), ORF6 protein (ORF6), ORF7a protein (ORF7a), ORF7b (ORF7b), ORF8 protein (ORF8), nucleocapsid phosphoprotein (N), and ORF10 protein (ORF10) genes, complete cds	Severe acute respiratory syndrome coronavirus 2	36.2	36.2	100%	8.9	95.45%	29824
MW292659.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/MD-MDH-0400/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	36.2	36.2	100%	8.9	95.45%	29831

MW292658.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/MD-MDH-0399/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	36.2	36.2	100%	8.9	95.45%	29821
MW292657.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/MD-MDH-0398/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	36.2	36.2	100%	8.9	95.45%	29820
MW292656.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/MD-MDH-0397/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	36.2	36.2	100%	8.9	95.45%	29821
MW292655.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/MD-MDH-0396/2020 ORF1ab polyprotein (ORF1ab), ORF1a polyprotein (ORF1ab), surface glycoprotein (S), ORF3a protein (ORF3a), envelope protein (E), membrane glycoprotein (M), ORF6 protein (ORF6), ORF7a protein (ORF7a), ORF7b (ORF7b), ORF8 protein (ORF8), nucleocapsid phosphoprotein (N), and ORF10 protein (ORF10) genes, complete cds	Severe acute respiratory syndrome coronavirus 2	36.2	36.2	100%	8.9	95.45%	29825
MW292654.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/MD-MDH-0395/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	36.2	36.2	100%	8.9	95.45%	29824
MW292653.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/MD-MDH-0394/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	36.2	36.2	100%	8.9	95.45%	29820
MW292652.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/MD-MDH-0393/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	36.2	36.2	100%	8.9	95.45%	29837
MW292651.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/MD-MDH-0392/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	36.2	36.2	100%	8.9	95.45%	29818
MW292650.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/MD-MDH-0391/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	36.2	36.2	100%	8.9	95.45%	29822
MW292649.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/MD-MDH-0390/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	36.2	36.2	100%	8.9	95.45%	29806
MW292648.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/MD-MDH-0389/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	36.2	36.2	100%	8.9	95.45%	29851
MW292647.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/MD-MDH-0388/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	36.2	36.2	100%	8.9	95.45%	29823
MW292646.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/MD-MDH-0387/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	36.2	36.2	100%	8.9	95.45%	29815
MW292645.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/MD-MDH-0386/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	36.2	36.2	100%	8.9	95.45%	29849
MW292644.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/MD-MDH-0385/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	36.2	36.2	100%	8.9	95.45%	29867
MW292643.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/MD-MDH-0384/2020 ORF1ab polyprotein (ORF1ab), ORF1a polyprotein (ORF1ab), surface glycoprotein (S), ORF3a protein (ORF3a), envelope protein (E), membrane glycoprotein (M), ORF6 protein (ORF6), ORF7a protein (ORF7a), ORF7b (ORF7b), ORF8 protein (ORF8), nucleocapsid phosphoprotein (N), and ORF10 protein (ORF10) genes, complete cds	Severe acute respiratory syndrome coronavirus 2	36.2	36.2	100%	8.9	95.45%	29821

MW292642.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/MD-MDH-0383/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	36.2	36.2	100%	8.9	95.45%	29818
MW292641.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/MD-MDH-0382/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	36.2	36.2	100%	8.9	95.45%	29821
MW292640.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/MD-MDH-0381/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	36.2	36.2	100%	8.9	95.45%	29866
MW292639.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/MD-MDH-0380/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	36.2	36.2	100%	8.9	95.45%	29816
MW292638.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/MD-MDH-0379/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	36.2	36.2	100%	8.9	95.45%	29821
MW292637.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/MD-MDH-0378/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	36.2	36.2	100%	8.9	95.45%	29814
MW292636.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/MD-MDH-0377/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	36.2	36.2	100%	8.9	95.45%	29819
MW292635.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/MD-MDH-0376/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	36.2	36.2	100%	8.9	95.45%	29879
MW292634.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/MD-MDH-0375/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	36.2	36.2	100%	8.9	95.45%	29829
MW292633.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/MD-MDH-0374/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	36.2	36.2	100%	8.9	95.45%	29848
MW292632.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/MD-MDH-0373/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	36.2	36.2	100%	8.9	95.45%	29857
MW292631.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/MD-MDH-0372/2020 ORF1ab polyprotein (ORF1ab), ORF1a polyprotein (ORF1ab), surface glycoprotein (S), ORF3a protein (ORF3a), envelope protein (E), membrane glycoprotein (M), ORF6 protein (ORF6), ORF7a protein (ORF7a), ORF7b (ORF7b), ORF8 protein (ORF8), nucleocapsid phosphoprotein (N), and ORF10 protein (ORF10) genes, complete cds	Severe acute respiratory syndrome coronavirus 2	36.2	36.2	100%	8.9	95.45%	29820
MW292630.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/MD-MDH-0371/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	36.2	36.2	100%	8.9	95.45%	29855
MW292629.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/MD-MDH-0370/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	36.2	36.2	100%	8.9	95.45%	29818
MW290973.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/IRQ/ICGEB-5T/2020 ORF1ab polyprotein (ORF1ab), ORF1a polyprotein (ORF1ab), surface glycoprotein (S), ORF3a protein (ORF3a), envelope protein (E), membrane glycoprotein (M), ORF6 protein (ORF6), ORF7a protein (ORF7a), ORF7b (ORF7b), ORF8 protein (ORF8), nucleocapsid phosphoprotein (N), and ORF10 protein (ORF10) genes, complete cds	Severe acute respiratory syndrome coronavirus 2	36.2	36.2	100%	8.9	95.45%	29720
MW290972.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/IRQ/ICGEB-2T/2020 ORF1ab polyprotein (ORF1ab),	Severe acute respiratory syndrome coronavirus 2	36.2	36.2	100%	8.9	95.45%	29576

MW290971.1	ORF1a polyprotein (ORF1ab), surface glycoprotein (S), ORF3a protein (ORF3a), envelope protein (E), membrane glycoprotein (M), ORF6 protein (ORF6), ORF7a protein (ORF7a), ORF7b (ORF7b), ORF8 protein (ORF8), nucleocapsid phosphoprotein (N), and ORF10 protein (ORF10) genes, complete cds	Severe acute respiratory syndrome coronavirus 2	36.2	36.2	100%	8.9	95.45%	29865
MW290970.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/UT-UPHL-201131465/2020 ORF1ab polyprotein (ORF1ab), ORF1a polyprotein (ORF1ab), surface glycoprotein (S), ORF3a protein (ORF3a), envelope protein (E), membrane glycoprotein (M), ORF6 protein (ORF6), ORF7a protein (ORF7a), ORF7b (ORF7b), ORF8 protein (ORF8), nucleocapsid phosphoprotein (N), and ORF10 protein (ORF10) genes, complete cds	Severe acute respiratory syndrome coronavirus 2	36.2	36.2	100%	8.9	95.45%	29865
MW290969.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/UT-UPHL-201131133/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	36.2	36.2	100%	8.9	95.45%	29865
MW290968.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/UT-UPHL-201129708/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	36.2	36.2	100%	8.9	95.45%	29878
MW290967.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/UT-UPHL-201128910/2020 ORF1ab polyprotein (ORF1ab) gene, complete cds; ORF1a polyprotein (ORF1ab) gene, partial cds; surface glycoprotein (S), ORF3a protein (ORF3a), envelope protein (E), membrane glycoprotein (M), and ORF6 protein (ORF6) genes, complete cds; ORF7a protein (ORF7a) and ORF7b (ORF7b) genes, partial cds; and ORF8 protein (ORF8), nucleocapsid phosphoprotein (N), and ORF10 protein (ORF10) genes, complete cds	Severe acute respiratory syndrome coronavirus 2	36.2	36.2	100%	8.9	95.45%	29864
MW290966.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/UT-UPHL-201128144/2020 ORF1ab polyprotein (ORF1ab) gene, complete cds; ORF1a polyprotein (ORF1ab) gene, partial cds; surface glycoprotein (S), ORF3a protein (ORF3a), envelope protein (E), membrane glycoprotein (M), and ORF6 protein (ORF6) genes, complete cds; ORF7a protein (ORF7a) and ORF7b (ORF7b) genes, partial cds; and ORF8 protein (ORF8), nucleocapsid phosphoprotein (N), and ORF10 protein (ORF10) genes, complete cds	Severe acute respiratory syndrome coronavirus 2	36.2	36.2	100%	8.9	95.45%	29816
MW290965.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/UT-UPHL-201128010/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	36.2	36.2	100%	8.9	95.45%	29871
MW290964.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/UT-UPHL-201127728/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	36.2	36.2	100%	8.9	95.45%	29887
MW290963.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/UT-UPHL-201126219/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	36.2	36.2	100%	8.9	95.45%	29872
	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/UT-UPHL-201124901/2020 ORF1ab polyprotein (ORF1ab) gene, complete cds; ORF1a polyprotein (ORF1ab) gene, partial cds; and surface glycoprotein (S), ORF3a protein (ORF3a), envelope	Severe acute respiratory syndrome coronavirus 2	36.2	36.2	100%	8.9	95.45%	29790

	protein (E), membrane glycoprotein (M), ORF6 protein (ORF6), ORF7a protein (ORF7a), ORF7b (ORF7b), ORF8 protein (ORF8), nucleocapsid phosphoprotein (N), and ORF10 protein (ORF10) genes, complete cds							
MW290962.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/UT-UPHL-201124395/2020 ORF1ab polyprotein (ORF1ab) gene, complete cds; ORF1a polyprotein (ORF1ab) gene, partial cds; and surface glycoprotein (S), ORF3a protein (ORF3a), envelope protein (E), membrane glycoprotein (M), ORF6 protein (ORF6), ORF7a protein (ORF7a), ORF7b (ORF7b), ORF8 protein (ORF8), nucleocapsid phosphoprotein (N), and ORF10 protein (ORF10) genes, complete cds	Severe acute respiratory syndrome coronavirus 2	36.2	36.2	100%	8.9	95.45%	29865
MW290961.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/UT-UPHL-201122426/2020 ORF1ab polyprotein (ORF1ab) gene, complete cds; ORF1a polyprotein (ORF1ab) gene, partial cds; and surface glycoprotein (S), ORF3a protein (ORF3a), envelope protein (E), membrane glycoprotein (M), ORF6 protein (ORF6), ORF7a protein (ORF7a), ORF7b (ORF7b), ORF8 protein (ORF8), nucleocapsid phosphoprotein (N), and ORF10 protein (ORF10) genes, complete cds	Severe acute respiratory syndrome coronavirus 2	36.2	36.2	100%	8.9	95.45%	29865
MW290960.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/UT-UPHL-201121351/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	36.2	36.2	100%	8.9	95.45%	29865
MW290959.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/UT-UPHL-201120724/2020 ORF1ab polyprotein (ORF1ab) gene, complete cds; ORF1a polyprotein (ORF1ab) gene, partial cds; and surface glycoprotein (S), ORF3a protein (ORF3a), envelope protein (E), membrane glycoprotein (M), ORF6 protein (ORF6), ORF7a protein (ORF7a), ORF7b (ORF7b), ORF8 protein (ORF8), nucleocapsid phosphoprotein (N), and ORF10 protein (ORF10) genes, complete cds	Severe acute respiratory syndrome coronavirus 2	36.2	36.2	100%	8.9	95.45%	29864
MW290958.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/UT-UPHL-201120091/2020 ORF1ab polyprotein (ORF1ab) gene, complete cds; ORF1a polyprotein (ORF1ab) gene, partial cds; and surface glycoprotein (S), ORF3a protein (ORF3a), envelope protein (E), membrane glycoprotein (M), ORF6 protein (ORF6), ORF7a protein (ORF7a), ORF7b (ORF7b), ORF8 protein (ORF8), nucleocapsid phosphoprotein (N), and ORF10 protein (ORF10) genes, complete cds	Severe acute respiratory syndrome coronavirus 2	36.2	36.2	100%	8.9	95.45%	29865
MW290957.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/UT-UPHL-201119669/2020 ORF1ab polyprotein (ORF1ab) gene, complete cds; ORF1a polyprotein (ORF1ab) gene, partial cds; surface glycoprotein (S), ORF3a protein (ORF3a), envelope protein (E), membrane glycoprotein (M), and ORF6 protein (ORF6) genes, complete cds; ORF7a protein (ORF7a) and ORF7b (ORF7b) genes, partial cds; and ORF8 protein (ORF8), nucleocapsid phosphoprotein (N), and ORF10 protein (ORF10) genes, complete cds	Severe acute respiratory syndrome coronavirus 2	36.2	36.2	100%	8.9	95.45%	29777

MW290956.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/UT-UPHL-201118755/2020 ORF1ab polyprotein (ORF1ab), ORF1a polyprotein (ORF1ab), and surface glycoprotein (S) genes, partial cds; ORF3a protein (ORF3a), envelope protein (E), membrane glycoprotein (M), and ORF6 protein (ORF6) genes, complete cds; ORF7a protein (ORF7a) and ORF7b (ORF7b) genes, partial cds; and ORF8 protein (ORF8), nucleocapsid phosphoprotein (N), and ORF10 protein (ORF10) genes, complete cds	Severe acute respiratory syndrome coronavirus 2	36.2	36.2	100%	8.9	95.45%	29860
MW290955.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/UT-UPHL-201117542/2020 ORF1ab polyprotein (ORF1ab), ORF1a polyprotein (ORF1ab), and surface glycoprotein (S) genes, partial cds; ORF3a protein (ORF3a), envelope protein (E), membrane glycoprotein (M), and ORF6 protein (ORF6) genes, complete cds; ORF7a protein (ORF7a) and ORF7b (ORF7b) genes, partial cds; and ORF8 protein (ORF8), nucleocapsid phosphoprotein (N), and ORF10 protein (ORF10) genes, complete cds	Severe acute respiratory syndrome coronavirus 2	36.2	36.2	100%	8.9	95.45%	29854
MW290954.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/UT-UPHL-201117199/2020 ORF1ab polyprotein (ORF1ab), ORF1a polyprotein (ORF1ab), surface glycoprotein (S), ORF3a protein (ORF3a), envelope protein (E), membrane glycoprotein (M), ORF6 protein (ORF6), ORF7a protein (ORF7a), ORF7b (ORF7b), ORF8 protein (ORF8), nucleocapsid phosphoprotein (N), and ORF10 protein (ORF10) genes, complete cds	Severe acute respiratory syndrome coronavirus 2	36.2	36.2	100%	8.9	95.45%	29865
MW290953.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/UT-UPHL-201117069/2020 ORF1ab polyprotein (ORF1ab), ORF1a polyprotein (ORF1ab), surface glycoprotein (S), ORF3a protein (ORF3a), envelope protein (E), membrane glycoprotein (M), ORF6 protein (ORF6), ORF7a protein (ORF7a), ORF7b (ORF7b), ORF8 protein (ORF8), nucleocapsid phosphoprotein (N), and ORF10 protein (ORF10) genes, complete cds	Severe acute respiratory syndrome coronavirus 2	36.2	36.2	100%	8.9	95.45%	29865
MW290952.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/UT-UPHL-201116393/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	36.2	36.2	100%	8.9	95.45%	29896
MW290951.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/UT-UPHL-201116223/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	36.2	36.2	100%	8.9	95.45%	29824
MW290950.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/UT-UPHL-201115394/2020 ORF1ab polyprotein (ORF1ab) gene, complete cds; ORF1a polyprotein (ORF1ab) gene, partial cds; and surface glycoprotein (S), ORF3a protein (ORF3a), envelope protein (E), membrane glycoprotein (M), ORF6 protein (ORF6), ORF7a protein (ORF7a), ORF7b (ORF7b), ORF8 protein (ORF8), nucleocapsid phosphoprotein (N), and ORF10 protein (ORF10) genes, complete cds	Severe acute respiratory syndrome coronavirus 2	36.2	36.2	100%	8.9	95.45%	29864
MW290949.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-	Severe acute respiratory	36.2	36.2	100%	8.9	95.45%	29865

	2/human/USA/UT-UPHL-201114835/2020 ORF1ab polyprotein (ORF1ab) gene, complete cds; ORF1a polyprotein (ORF1ab) gene, partial cds; surface glycoprotein (S), ORF3a protein (ORF3a), envelope protein (E), membrane glycoprotein (M), and ORF6 protein (ORF6) genes, complete cds; ORF7a protein (ORF7a) and ORF7b (ORF7b) genes, partial cds; and ORF8 protein (ORF8), nucleocapsid phosphoprotein (N), and ORF10 protein (ORF10) genes, complete cds	syndrome coronavirus 2						
MW290948.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/UT-UPHL-201113704/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	36.2	36.2	100%	8.9	95.45%	29865
MW290947.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/UT-UPHL-201113629/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	36.2	36.2	100%	8.9	95.45%	29876
MW290946.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/UT-UPHL-201112041/2020 ORF1ab polyprotein (ORF1ab), ORF1a polyprotein (ORF1ab), surface glycoprotein (S), ORF3a protein (ORF3a), envelope protein (E), membrane glycoprotein (M), ORF6 protein (ORF6), ORF7a protein (ORF7a), ORF7b (ORF7b), ORF8 protein (ORF8), nucleocapsid phosphoprotein (N), and ORF10 protein (ORF10) genes, complete cds	Severe acute respiratory syndrome coronavirus 2	36.2	36.2	100%	8.9	95.45%	29865
MW290945.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/UT-UPHL-201111636/2020 ORF1ab polyprotein (ORF1ab) gene, complete cds; ORF1a polyprotein (ORF1ab) gene, partial cds; surface glycoprotein (S), ORF3a protein (ORF3a), envelope protein (E), membrane glycoprotein (M), and ORF6 protein (ORF6) genes, complete cds; ORF7a protein (ORF7a) and ORF7b (ORF7b) genes, partial cds; and ORF8 protein (ORF8), nucleocapsid phosphoprotein (N), and ORF10 protein (ORF10) genes, complete cds	Severe acute respiratory syndrome coronavirus 2	36.2	36.2	100%	8.9	95.45%	29887
MW290944.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/UT-UPHL-201109153/2020 ORF1ab polyprotein (ORF1ab) gene, complete cds; ORF1a polyprotein (ORF1ab) gene, partial cds; surface glycoprotein (S), ORF3a protein (ORF3a), envelope protein (E), membrane glycoprotein (M), and ORF6 protein (ORF6) genes, complete cds; ORF7a protein (ORF7a) and ORF7b (ORF7b) genes, partial cds; and ORF8 protein (ORF8), nucleocapsid phosphoprotein (N), and ORF10 protein (ORF10) genes, complete cds	Severe acute respiratory syndrome coronavirus 2	36.2	36.2	100%	8.9	95.45%	29854
MW290943.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/UT-UPHL-201108802/2020 ORF1ab polyprotein (ORF1ab) gene, complete cds; ORF1a polyprotein (ORF1ab) gene, partial cds; surface glycoprotein (S), ORF3a protein (ORF3a), envelope protein (E), membrane glycoprotein (M), and ORF6 protein (ORF6) genes, complete cds; ORF7a protein (ORF7a) and ORF7b (ORF7b) genes, partial cds; and ORF8 protein (ORF8), nucleocapsid phosphoprotein (N), and ORF10 protein (ORF10) genes,	Severe acute respiratory syndrome coronavirus 2	36.2	36.2	100%	8.9	95.45%	29813

	complete cds							
MW290942.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/UT-UPHL-201108622/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	36.2	36.2	100%	8.9	95.45%	29900
MW290941.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/UT-UPHL-201108102/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	36.2	36.2	100%	8.9	95.45%	29865
MW290940.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/UT-UPHL-201107239/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	36.2	36.2	100%	8.9	95.45%	29899
MW290939.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/UT-UPHL-201105052/2020 ORF1ab polyprotein (ORF1ab), ORF1a polyprotein (ORF1ab), and surface glycoprotein (S) genes, partial cds; ORF3a protein (ORF3a), envelope protein (E), membrane glycoprotein (M), and ORF6 protein (ORF6) genes, complete cds; ORF7a protein (ORF7a) and ORF7b (ORF7b) genes, partial cds; and ORF8 protein (ORF8), nucleocapsid phosphoprotein (N), and ORF10 protein (ORF10) genes, complete cds	Severe acute respiratory syndrome coronavirus 2	36.2	36.2	100%	8.9	95.45%	29864
MW290938.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/UT-UPHL-201104262/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	36.2	36.2	100%	8.9	95.45%	29865
MW290937.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/UT-UPHL-201103661/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	36.2	36.2	100%	8.9	95.45%	29897
MW290936.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/UT-UPHL-201102669/2020 ORF1ab polyprotein (ORF1ab) gene, complete cds; ORF1a polyprotein (ORF1ab) gene, partial cds; and surface glycoprotein (S), ORF3a protein (ORF3a), envelope protein (E), membrane glycoprotein (M), ORF6 protein (ORF6), ORF7a protein (ORF7a), ORF7b (ORF7b), ORF8 protein (ORF8), nucleocapsid phosphoprotein (N), and ORF10 protein (ORF10) genes, complete cds	Severe acute respiratory syndrome coronavirus 2	36.2	36.2	100%	8.9	95.45%	29815
MW290935.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/UT-UPHL-201102520/2020 ORF1ab polyprotein (ORF1ab) gene, complete cds; ORF1a polyprotein (ORF1ab) gene, partial cds; and surface glycoprotein (S), ORF3a protein (ORF3a), envelope protein (E), membrane glycoprotein (M), ORF6 protein (ORF6), ORF7a protein (ORF7a), ORF7b (ORF7b), ORF8 protein (ORF8), nucleocapsid phosphoprotein (N), and ORF10 protein (ORF10) genes, complete cds	Severe acute respiratory syndrome coronavirus 2	36.2	36.2	100%	8.9	95.45%	29899
MW288285.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/MN-MDH-2048/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	36.2	36.2	100%	8.9	95.45%	29669
MW288284.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/MN-MDH-2047/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	36.2	36.2	100%	8.9	95.45%	29819

ALIGNMENT VIEW FLAT QUERY-ANCHORED WITH LETTERS FOR IDENTITIES

Query range 1: 1 to 22

Query	**1**	**GTGAAATGGTTATGTGTGGCGG**	**22**
MN312673.1	122	GTGAAATGGTTATGTGTGGCGG	143
MN312672.1	122	GTGAAATGGTTATGTGTGGCGG	143
MN312671.1	122	GTGAAATGGTTATGTGTGGCGG	143
MN312636.1	122	GTGAAATGGTTATGTGTGGCGG	143
MN312635.1	119	GTGAAATGGTTATGTGTGGCGG	140
MN312634.1	122	GTGAAATGGTTATGTGTGGCGG	143
MN312633.1	128	GTGAAATGGTTATGTGTGGCGG	149
MN312632.1	96	GTGAAATGGTTATGTGTGGCGG	117
GU190217.1	581	GTGAAATGGTTATGTGTGGCGG	602
GU190215.1	15260	GTGAAATGGTTATGTGTGGCGG	15281
XM_037013622.1	2494	GTGAAATGGTTATGTGTGG	2512
XM_037013621.1	2335	GTGAAATGGTTATGTGTGG	2353
XM_037013617.1	2918	GTGAAATGGTTATGTGTGG	2936
XM_036918394.1	2496	GTGAAATGGTTATGTGTGG	2514
XM_035898451.1	1327	GTGAAATGGTTATGTGTGG	1345
XM_035898450.1	1327	GTGAAATGGTTATGTGTGG	1345
XM_035898448.1	1327	GTGAAATGGTTATGTGTGG	1345
MW295643.1	15425	GTGAAATGGTCATGTGTGGCGG	15446
MW295640.1	15425	GTGAAATGGTCATGTGTGGCGG	15446
MW295556.1	15425	GTGAAATGGTCATGTGTGGCGG	15446
MW294011.1	15431	GTGAAATGGTCATGTGTGGCGG	15452
MW294007.1	15431	GTGAAATGGTCATGTGTGGCGG	15452
MW292665.1	15393	GTGAAATGGTCATGTGTGGCGG	15414
MW292664.1	15395	GTGAAATGGTCATGTGTGGCGG	15416
MW292663.1	15395	GTGAAATGGTCATGTGTGGCGG	15416
MW292662.1	15396	GTGAAATGGTCATGTGTGGCGG	15417
MW292661.1	15395	GTGAAATGGTCATGTGTGGCGG	15416
MW292660.1	15397	GTGAAATGGTCATGTGTGGCGG	15418
MW292659.1	15406	GTGAAATGGTCATGTGTGGCGG	15427
MW292658.1	15395	GTGAAATGGTCATGTGTGGCGG	15416
MW292657.1	15396	GTGAAATGGTCATGTGTGGCGG	15417
MW292656.1	15394	GTGAAATGGTCATGTGTGGCGG	15415
MW292655.1	15396	GTGAAATGGTCATGTGTGGCGG	15417
MW292654.1	15395	GTGAAATGGTCATGTGTGGCGG	15416
MW292653.1	15395	GTGAAATGGTCATGTGTGGCGG	15416
MW292652.1	15406	GTGAAATGGTCATGTGTGGCGG	15427
MW292651.1	15395	GTGAAATGGTCATGTGTGGCGG	15416
MW292650.1	15395	GTGAAATGGTCATGTGTGGCGG	15416
MW292649.1	15395	GTGAAATGGTCATGTGTGGCGG	15416
MW292648.1	15425	GTGAAATGGTCATGTGTGGCGG	15446
MW292647.1	15395	GTGAAATGGTCATGTGTGGCGG	15416
MW292646.1	15393	GTGAAATGGTCATGTGTGGCGG	15414
MW292645.1	15423	GTGAAATGGTCATGTGTGGCGG	15444
MW292644.1	15395	GTGAAATGGTCATGTGTGGCGG	15416
MW292643.1	15393	GTGAAATGGTCATGTGTGGCGG	15414
MW292642.1	15397	GTGAAATGGTCATGTGTGGCGG	15418
MW292641.1	15393	GTGAAATGGTCATGTGTGGCGG	15414
MW292640.1	15394	GTGAAATGGTCATGTGTGGCGG	15415
MW292639.1	15393	GTGAAATGGTCATGTGTGGCGG	15414
MW292638.1	15396	GTGAAATGGTCATGTGTGGCGG	15417
MW292637.1	15393	GTGAAATGGTCATGTGTGGCGG	15414
MW292636.1	15395	GTGAAATGGTCATGTGTGGCGG	15416
MW292635.1	15406	GTGAAATGGTCATGTGTGGCGG	15427

MW292634.1	15396	GTGAAATGGTCATGTGTGGCGG	15417
MW292633.1	15419	GTGAAATGGTCATGTGTGGCGG	15440
MW292632.1	15430	GTGAAATGGTCATGTGTGGCGG	15451
MW292631.1	15396	GTGAAATGGTCATGTGTGGCGG	15417
MW292630.1	15431	GTGAAATGGTCATGTGTGGCGG	15452
MW292629.1	15394	GTGAAATGGTCATGTGTGGCGG	15415
MW290973.1	15391	GTGAAATGGTCATGTGTGGCGG	15412
MW290972.1	15248	GTGAAATGGTCATGTGTGGCGG	15269
MW290971.1	15393	GTGAAATGGTCATGTGTGGCGG	15414
MW290970.1	15393	GTGAAATGGTCATGTGTGGCGG	15414
MW290969.1	15406	GTGAAATGGTCATGTGTGGCGG	15427
MW290968.1	15392	GTGAAATGGTCATGTGTGGCGG	15413
MW290967.1	15344	GTGAAATGGTCATGTGTGGCGG	15365
MW290966.1	15399	GTGAAATGGTCATGTGTGGCGG	15420
MW290965.1	15415	GTGAAATGGTCATGTGTGGCGG	15436
MW290964.1	15400	GTGAAATGGTCATGTGTGGCGG	15421
MW290963.1	15318	GTGAAATGGTCATGTGTGGCGG	15339
MW290962.1	15393	GTGAAATGGTCATGTGTGGCGG	15414
MW290961.1	15393	GTGAAATGGTCATGTGTGGCGG	15414
MW290960.1	15393	GTGAAATGGTCATGTGTGGCGG	15414
MW290959.1	15392	GTGAAATGGTCATGTGTGGCGG	15413
MW290958.1	15393	GTGAAATGGTCATGTGTGGCGG	15414
MW290957.1	15354	GTGAAATGGTCATGTGTGGCGG	15375
MW290956.1	15388	GTGAAATGGTCATGTGTGGCGG	15409
MW290955.1	15382	GTGAAATGGTCATGTGTGGCGG	15403
MW290954.1	15393	GTGAAATGGTCATGTGTGGCGG	15414
MW290953.1	15393	GTGAAATGGTCATGTGTGGCGG	15414
MW290952.1	15424	GTGAAATGGTCATGTGTGGCGG	15445
MW290951.1	15393	GTGAAATGGTCATGTGTGGCGG	15414
MW290950.1	15392	GTGAAATGGTCATGTGTGGCGG	15413
MW290949.1	15393	GTGAAATGGTCATGTGTGGCGG	15414
MW290948.1	15393	GTGAAATGGTCATGTGTGGCGG	15414
MW290947.1	15404	GTGAAATGGTCATGTGTGGCGG	15425
MW290946.1	15393	GTGAAATGGTCATGTGTGGCGG	15414
MW290945.1	15415	GTGAAATGGTCATGTGTGGCGG	15436
MW290944.1	15382	GTGAAATGGTCATGTGTGGCGG	15403
MW290943.1	15393	GTGAAATGGTCATGTGTGGCGG	15414
MW290942.1	15428	GTGAAATGGTCATGTGTGGCGG	15449
MW290941.1	15393	GTGAAATGGTCATGTGTGGCGG	15414
MW290940.1	15427	GTGAAATGGTCATGTGTGGCGG	15448
MW290939.1	15392	GTGAAATGGTCATGTGTGGCGG	15413
MW290938.1	15393	GTGAAATGGTCATGTGTGGCGG	15414
MW290937.1	15425	GTGAAATGGTCATGTGTGGCGG	15446
MW290936.1	15392	GTGAAATGGTCATGTGTGGCGG	15413
MW290935.1	15427	GTGAAATGGTCATGTGTGGCGG	15448
MW288285.1	15414	GTGAAATGGTCATGTGTGGCGG	15435
MW288284.1	15395	GTGAAATGGTCATGTGTGGCGG	15416

Bat coronavirus BM48-31/BGR/2008, complete genome

GenBank: GU190215.1

FASTA Graphics

Go to:

LOCUS GU190215 22 bp RNA linear VRL **08-OCT-2010**

DEFINITION Bat coronavirus BM48-31/BGR/2008, complete genome.

ACCESSION GU190215 REGION: 15260..15281
VERSION GU190215.1
KEYWORDS .
SOURCE Bat coronavirus BM48-31/BGR/2008
ORGANISM Bat coronavirus BM48-31/BGR/2008
Viruses; Riboviria; Orthornavirae; Pisuviricota; Pisoniviricetes;
Nidovirales; Cornidovirineae; Coronaviridae; Orthocoronavirinae;
Betacoronavirus; Sarbecovirus; unclassified Sarbecovirus.
REFERENCE 1 (bases 1 to 22)
AUTHORS Drexler,J.F., Gloza-Rausch,F., Glende,J., Corman,V.M., Muth,D.,
Goettsche,M., Seebens,A., Niedrig,M., Pfefferle,S., Yordanov,S.,
Zhelyazkov,L., Hermanns,U., Vallo,P., Lukashev,A., Muller,M.A.,
Deng,H., Herrler,G. and Drosten,C.
TITLE Genomic Characterization of Severe Acute Respiratory
Syndrome-Related Coronavirus in European Bats and Classification of
Coronaviruses Based on Partial RNA-Dependent RNA Polymerase Gene
Sequences
JOURNAL J. Virol. 84 (21), 11336-11349 (2010)
PUBMED 20686038
REFERENCE 2 (bases 1 to 22)
AUTHORS Drexler,J.F., Corman,V.M. and Drosten,C.
TITLE Direct Submission
JOURNAL **Submitted (09-NOV-2009) Institute of Virology, University of Bonn
Medical Centre, Sigmund Freud-Str. 25, Bonn 53127, Germany**
FEATURES Location/Qualifiers
source 1..22
/organism="Bat coronavirus BM48-31/BGR/2008"
/mol_type="genomic RNA"
/strain="BtCoV/BM48-31/BGR/2008"
/host="**Rhinolophus blasii**"
/db_xref="taxon:864596"
/country="Bulgaria"
/collection_date="2008"
gene <1..>22
/gene="ORF1ab"
CDS <1..>22
/gene="ORF1ab"
/ribosomal_slippage
/codon_start=3
/product="ORF1ab"
/protein_id="ADK66840.1"
/translation="MESLVPGFNEKTHVQLSLPVLQVRDVLVRGFGDTVEEAVAEARQ
HLIEGTCGIVDLQKGVLPQLEQPYIFLKRCDARTAPHGHVMVELVAELDGVQYGRSGE
SLGVLVPHVGETPIGYRKVLVRKNGNKGAGGHLYGADLRFYDLGDELGTDPLDDFQQD
WNTKHGSGLRRDLFRELNGGVYTRYVDNNFCGPDGYPLECIKDLLARAGKSSAPLAEQ

LDFLESKRGVYCCREHEHEIAWYTERSDKSYELQTPFDITNAKKFDSFKGECPKFVFP
LNSTVKVLQPRVEKKKTEGFLGRIRTVYQVASPGECNSMHLSTYMNCNHCGEKSWQTC
DFLTATCEMCGNQNTVEEGPTTCGYVPSNAVVKMVCPACQNPEIGPDHSVADYHNNSK
IETRLRKGGRIKSFGGCVFSYVGCYNKRAFWVPRAAANIGSNHTGVVGEGVETMNEDL
LQILSRERVVINIVGEFCLNEEIAILLASLSASTSAFVETVKNLDFKTFKKIIESCGN
YKVTKGKFKPGVWNIGTSKSLLTPLHCFSSQAAGVVRSIFSRTLATANHSIVDLHRAA
MIIFSDISDQANRVLDAMVNTSDLVTESVVVMAYLTGGLVQQVSTWLSQLLNTSVDKF
SAVLRWLEQKLQGGIDFLRQAWGILKLLVTGAYVVIRGKIQVVNTSLIECVTSFVDVV
NKVFELCTDYITVAGARVRAINFGEVLIAQSRGLYRQCVRARDQLQLLMPLKSPKDVV
FLDGDAYDTLLTSEEVTVKNGTLEALDLELSDVVTGVAEGVPVCVNGLMLLELKEKEQ
YCALSPSLLATNNVFTLKGGAPTKGVTFGEDTVVEIQGYKSVKITFELDERVDKVLNE
KCASYTVETGTTAEELACVVAESVVKTLQPISELLTPMGIDLDEWSVAKFYLFDESGE
AVLSSHMYCSFYPPDEEEEEDLEESEDVEYGTEDDYTGAPLEFGASSTVEQDEVHDEE
EDWLAPQEESEVLYDQFTDYHKLTDNVFIKCADIVEESLKVNPTVVVNAANIHLKHGG
GVARALDKATGGSMQKESNDYISTNGPLRVGGSCLLSGHNLAKHCLHVVGPNKNAGED
IKLLDAAYENFNAYEVVLSPLLSAGIFGVSPIQSLETCKRVVRNTVYIVVNDSVVFDQ
LLAKTPGKTNERPVVESSEICEEVNQKPVVEFSETKELHEETNQKLKSSEEPVKTRIE
ELNTTVDEAKFLTTKLLLYADVNGNLSEDSKVLIGNDGASFKKGAPYIVGDIISEGEL
TCVVLPTKAVGGTTHMLTRALKNVPSDTYLTTYPGQGVSGYTLDEAKAALKKSRSVFY
ILPSANVNAKEEVLGTVAWNLREMLAHAEETRKVMPVCMDVRAIISTIQRKYKGIGIQ
EGLVDYKVRFYFYSSKTPIARVISNLNSLGEPLITMPLGYVTHGLNLEESARYMRSVK
VPVVVSVSSPDAVTSYNGYVTSASKSAEEHFIETVSLAGSYKDWSYSGQRTELGVEFL
KRGDKIVYHTVGNVIEFHMEGEVLPLEKLKTLLALREVKTIKVFTTVDNINLHTQVID
MSMTYGQQLGPTYMDGADLTKVKPHASHENKTFFVLPSDDTLRIEAFEYYHTVDESFF
GRYMSALNHTKRWKYPQVGGLTSIKWADNNCYLSSVLLSLQQIDIKFNAPALQDAYYR
ARAGDAANFCALVLAYSKKTVGELGDVRETMAHLLQHANLESAKRVLNVVCKHCGQKS
TTLSGVEAVMYMGTLSYDHLKRGVKIPCVCGREATQYLVKQESTFVMMSAPPAEYTLQ
TGEFLCANEYTGNYQCGHYTHITNRETIYKIDGALLTKITEYKGPVADVFYKETSYST
DIKPVSYKLDGVTYTEINPDLNGYYKKDNAYYTEQPIDLVPTQPLPNASFDNFRFVCA
NTKFADDLNQMTGFKKPPSRDLTITFFPDLNGDVVAIDYRHYTPTFKKGAKLVHKPIL
WHVNQTTTKSTFKPNMWCLRCLYSTKPVPTSNSFEVLSSDDAQGMDNLACESQQTVAE
EVVDNPTIQKDIIECDVKTTEVVGNVILKPSADGIKVTSELEHEDLMAAYVNETSITI
KKPNELSIMLGLKTIATHGAAAINSVPWIKICAYVKPFLGYVAEQSKNCIKRCFRRVF
NDYMPFLLTLLLQLCTFTKSTNFRIKAAMPIVIARNSVIGGVRFCLDALTMYVKSPKF
SGILTVVMWLLLLSVCLGCLVYAVASFGAILSGFGLMSYCDGVRAGYVNSSNVTIPDY
CAGSLPCGVCLGGLDSLDAYPALETIQVTISSYKLDLTFVGMMAEWFLAYMLFTKFFY
LLGLFALMQLFFGLFATHFVNNSWLMWLIINVVQMAPISAMVRMYVFFASFYYVWKAY
IHVINGCTSSTCIMCYKRNRATRVECTTIVNGMKKSFYVYANGGQGFCKLHNWNCLNC
DTFCSGSTFISDEVARDLSLQFKRPINPTDQSSYNVDSVTVKDGTLYLYFQKAGKLTY
ERHPLSYFVNLDNLRANNVKGTLPINVIVFDGKSKCEEAAAKSASVYYSQLMCQPILL
LDQALISDVGDSTEVAVKMFDAYVNAFSSTFNAPMEKLKTFIATAHAEIAKGVSLDSV
LSTFLSAARQGFVDSDVDTKDVMECLKLSHHSDLEITSDSCNNFMLTYNKVENMTPRD
LGACIDCSARHINAQVAKSHNVSLVWNVKDYMSLSEQLRKQIRSAAKKNNIPFKLTCA
TTRQVVNVITTKISLKGGKFVSNNWFRFLLKMTVLMVLVAFIFYFITPTHTLMGHDVF
SSEIIGYKAIHNGVTRDVLTTDDCFANKHTGFDHWFSQRGGSYRNDKTCPVIAAVITR

EVGFIVPGLPGTVRRASNGDFLHFLPRVFSAVGNICYTPAKLIEYTDFATSACVLAAE
CTIFKDAQGKPVPYCYDTNLLEGSISYSELRPDTRYVLMDGSIIQFPSTYLEGSVRVV
TTFDSEYCRHGTCERSDAGVCLSTNGRWVLNNDYYRSIPGVFCGADASDLLFNIFTPL
VRPVGTLDISASVVAGGLIAILVTCVAYYFMKFRRAFGEYNHVVFANALLFLLSFTIL
CLTPAYTFLPGIYSLLYLYLTFYFTNDVSFLAHLQWLAMFSPIVPFWITVTYVVCISI
KHCHWFFSNYLKKRVVFNGVTFSTFEEAALCTFLLNKEMYLKLRSETLLPLTQYNRYL
ALYNKYKYFSGALDTTSYREAACCHLAKALNDFSNSGADVLYQPPQTSITSAVLQSGF
RKMAFPSGKVEGCMVQVTCGTTTLNGLWLDDVVYCPRHVICTLEDMLNPNYDDLLIRK
SNHNFLVQASNVQLRVIGHTMQNCLLKLKVDIANPKTPKYKFVRIQPGQTFSVLACYN
GAPSGVYQCAMRSNHTIKGSFLNGSCGSVGFNIDYDCVSFCYMHHMELPTGVHAGTDL
EGNFYGPFVDRQTAQAAGTDTTITLNVLAWLYAAVINGERWFLNRFTTTLNDFNLVAM
KYNYEPLTQDQVDILGPLSAQTGVAVMDMCAALKELLQNGLNGRTILGSTILEDEFTP
FDVVRQCSGVTFQGKFKKVVKGTHHWLLLTLLTSLLILVQSTQWSLFFFVYEHAFLPF
TMGVVCFAACAMVLVKHKHAFLCLFLLPSLITVAYFNMIYMPASWVMRVMTWLDLVDT
SLSGYRLKDCVMYALAAFLLILMTARTVYDDAARRVWTVMNVITLVYKVYYGNSLDQA
LAMWALVISVTSNYSGVVTTIMFLARAIVFLCVEYYPILFITGNTLQCIMLVYCFLGY
CCCCYFGLFCLLNRYFRLTLGVYDYFVSTQEFRYMNSQGLLPPKTSLDAFKLNVKLLG
IGGKPCIKVATVQSKMSDIKCTSVVLLSVLQQLRIESSSKLWAQCVQLHNDILLAKDT
TEAFEKMVSLLSVLLSMQGAVDINKLCDEMLNNRATLQAIASEFSSLPSYAAYATAQE
AYEQAVANGDSEVVLKKLKKSLNVAKSEFDRDAAMQRKLEKMADQAMTQMYKQARSED
KRAKVTSAMQTMLFTMLRKLDNDALNNIINNARDGCVPLNIIPLTTAAKLMVVVPDYN
TYKNTCDGNTFTYASALWEIQQVVDADSKVVQLSEINMDNSQNLAWPLIVTALRSNSA
VKLQNNELSPVALRQMSCAAGTTQTACTDDNALAYYNTSKGGRFVLALLSDHQDLKWA
RFPKSDGTGTIYTELEPPCRFVTDTPKGPKVKYLYFIKGLNNLNRGMVLGSLAATVRL
QAGNATEVPANSTVLSFCAFAVDPAKAYKDYLASGGQPITNCVKMLCTHTGTGQAITV
IPEANMDQESFGGASCCLYCRCHIDHPNPKGFCDLKGKYVQIPTTCTNDPVGFILRNT
VCTVCGMWKGYGCSCDQLREPVMQAADAPAFLNRVCGVSAARLTPCGTGTSTDVVYRA
FDIYNEKVAGFAKFLKTNCCRFQEVDEEGNLLDSYFVVKRHTMSNYQHEETMYNLVKE
CPAVAVHDFFKFRVDGDMVPHISRQRLTKYTMADLVYALRHFDEGNCDTLKEILVTYN
CCDDAYFNKKDWYDFVENPDILRVYACLGERVRQALLKTVQFCDAMRDAGIVGVLTLD
NQDLNGNWYDFGDFVQVAPGAGIPIVDSYYSLLMPILTLTKALAAESHMDCDTTKPLI
KWDLLKYDFTEERLCLFNRYFKYWDQTYHPNCINCLDDRCILHCANFNVLFSTVFPPT
SFGPLVRKIFVDGVPFVVSTGYHFRELGVVHNQDVNLHSSRLSFKELLVYAADPAMHA
ASGNLLLDKRTTCFSVAALTNSVAFQTVKPGNFNKDFYDFAVSKGFFKEGSSVELKHF
FFAQDGNAAISDYDYYRYNLPTMCDIRQLLFVVEVVDKYFDCYDGGCINANQVIVNNL
DKSAGFPFNKWGKARLYYDSMSYEDQDALFAYTKRNVIPTITQMNLKYAISAKNRART
VAGVSICSTMTNRQFHQKLLKSIAATRGATVVIGTSKFYGGWHNMLKTVYSDVETPNL
MGWDYPKCDRAMPNMLRIMASLVLARKHSTCCNLSHRFYGLANECAQVL**S**EMVMCGGS
LYVKPGGTSSGDATTAYANSVFNICQAVTANVNALLSTDGNKIADKYVRNLQHRLYEC
LYRNRDVDHEFVEEFYAYLRKHFSMMILSDDAVVCYNSNYAAQGLVASIKNFKAVLYY
QNNVFMSEAKCWTETDLTKGPHEFCSQHTMLVKQGDDYVYLPYPDPSRILGAGCFVDD
IVKTDGTLMIERFVSLAIDAYPLTKHPNQEYADVFHLYLQYIRKLHDELTGHMLDMYS
VMLTNDNTSRYWEPEFYEAMYTPHTVLQAVGACVLCNSQTSLRCGSCIRRPFLCCKCC
YDHVISTSHKLVLSVNPYVCNAPGCDVTDVTQLYLGGMSYYCKSHKPPISFPLCANGQ
VFGLYKNTCVGSDNVTDFNAIATCDWTNAGDYILANTCTERLKLFAAETLKANEETFK

LSYGIATVREVLSDRELHLSWEIGKPRPPLNRNYVFTGYRVTKNSKVQIGEYTFEKGD
YGDAVVYRGTTTYKLNVGDYFVLTSHTVMPLTAPTLVPQEHYVRITGLYPTLNISDEF
SSNVANYQKVGMQKYSTLQGPPGTGKSHFAIGLALYYPSARIVYTACSHAAVDALCEK
ALKYLPIDKCSRIIPARARVECFDKFKVNSTLEQYVFCTVNALPETTADIVVFDEVSM
ATNYDLSVVNARLRAKHYVYIGDPAQLPAPRTLLTKGTLEPEYFNSVCRLMKTIGPDM
FLGTCRRCPAEIVDTVSALVYDNKLRAHKGKSSQCFKMFYKGVITHDVSSAINRPQIG
VVREFLTRNPAWRKAVFISPYNSQNAVASKILGLPTQTVDSSQGSEYDYVIFAQTTET
AHSCNVNRFNVAITRAKVGILCIMSDKDLYDKLQFTSLEVPRRSVAVLQSENVTGLFK
DCSKLITGLHPTQAPTYLSVDTKFKTEGLCVDIPGIPKDMTYRRLISMMGFKMNYQVN
GYPNMFITRDEAIKHVRAWIGFDVEGCHATRDAVGTNLPLQLGFSTGVNLVAVPTGYV
DTSAATEFSRVNAKPPPGDQFKHLIPLMYKGLPWNIVRVKIVQMLSDTLKDLSDRVVF
VLWAHGFELTSMKYFVKIGPERTCCLCDKRATCFSTSSDTYACWHHSVGFDYVYNPFM
IDVQQWGFTGNLQSNHDQHCQVHGNAHVASCDAIMTRCLAIHECFVKRVDWSVEYPII
GDELRINVACRKVQHMVVKSALLADKFPVLHDIGNPKAIKCVPQADVEWKFYDVQPCS
DKAYKIEELFYSYATHHDKFTDGVCLFWNCNVDRYPSNAIVCRFDTRVLSNLNLPGCD
GGSLYVNKHAFHTPAFDKGAFANLKQLPFFYYSDSPCESHGKQVVSDIDYVPLKSATC
ITRCNLGGAVCRHHASEYRQYLDAYNMMISAGFSLWIYKQFDTYNLWNTFTRLQSLEN
VAYNVVNKGHFDGQAGEKPVSIINNTVYTKVDGVDVEIFENKTTLPVNVAFELWAKRN
IKPVPEIKILNNLGVDIAANTVIWDYKRESPAYISTIGVCTMTDIAKKPTENACSSLT
VFFDGRVDGQVDSFRNARNGVLITEGSVKGLNPSKGPPQASLNGVTLIGESVKTQFNY
FKKVDGVVQQLPETYFTQSRSLDDFKPRSQMEVDFLQLAMDEFIERYKLEGYAFEHIV
YGDFSHGQLGGLHLMIGLAKRSLESLLKLEDFIPIDSTVKNYFVTDAQTGSSKCVCSV
IDLLLDDFVEIIKSQDLSVVSKVVTVTIDYAEISFMLWCKDGHVETFYPKLQANQTWQ
PGVAMPNLYKMQRMLLDKCDLHNYGENAVIPKGIMMNVAKYTQLCQYLNTLTIAVPYN
MRVIHFGAGSDKGVAPGSAVLKQWLPVGTLLVDSDINDFVSDADSTLIGDCSTVYTAN
KWDLIISDMYDPKTKHILKENDSKEGFFTYLCGFIKQKLALGGSVAIKITEHSWNADL
YKLMGYFSWWTAFVTNVNASSSEAFLIGVNYLGKQKESIDGYTMHANYIFWRNTNPIQ
LSSYSLFDMSKFPLKLRGTAVMSLKDNQINDMICSLLEKGRLIIRENNKVVFSSDVLV
NN"

ORIGIN

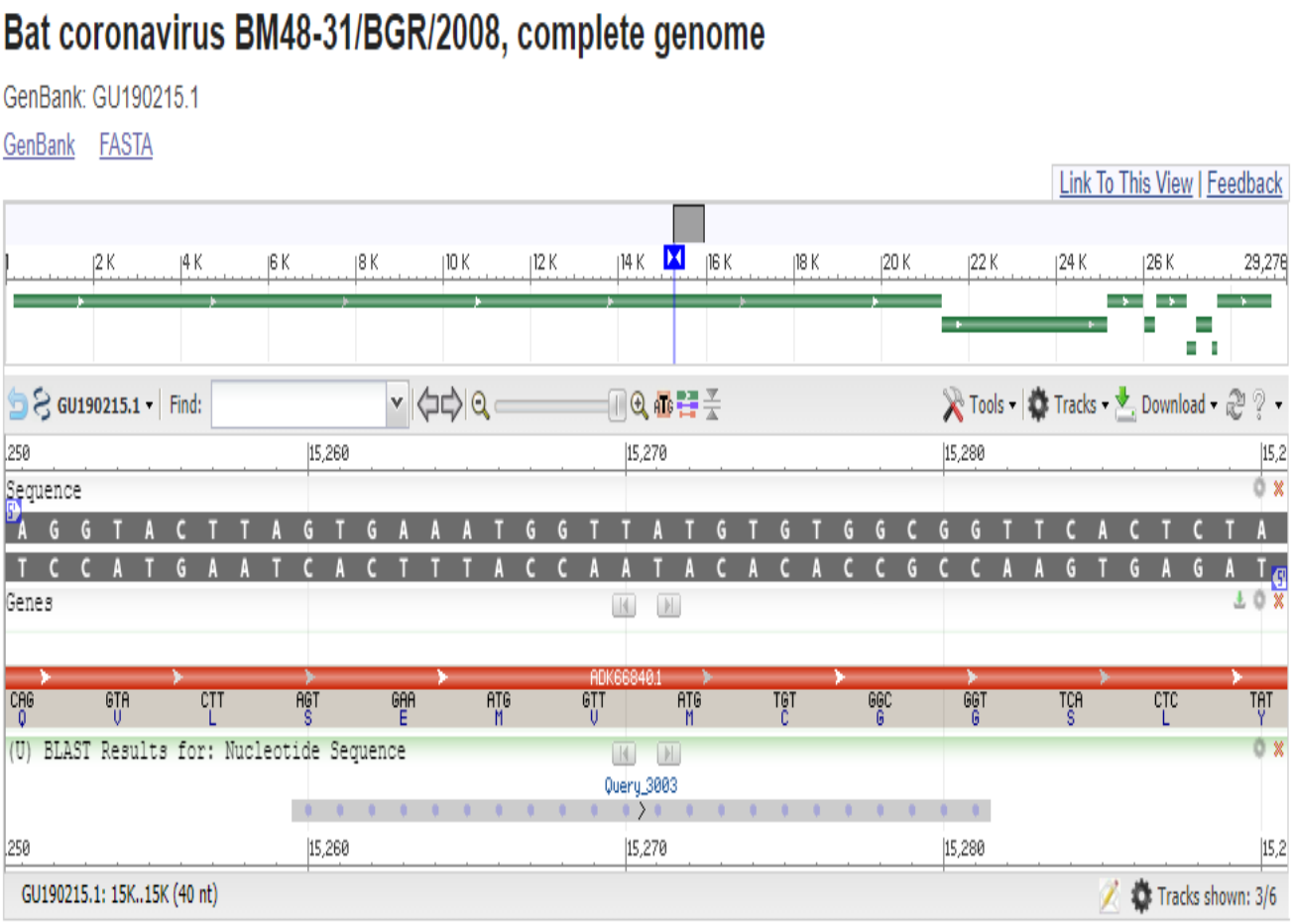
1 gtgaaatggt tatgtgtggc gg

//

Bat coronavirus BM48-31/BGR/2008, complete genome
Sequence ID: GU190215.1 **Length:** 29276 **Number of Matches:** 1
Range 1: 15260 to 15281 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
44.1 bits(22)	0.036	22/22(100%)	0/22(0%)	Plus/Plus

```
CDS: Putative 1        1          E   M   V   M   C   G   G
Query                  1      GTGAAATGGTTATGTGTGGCGG   22
                              ||||||||||||||||||||||
Sbjct                  15260  GTGAAATGGTTATGTGTGGCGG   15281
CDS:ORF1ab [Bat coro   5024   S  E   M   V   M   C   G   G
```



PREDICTED: Manis javanica consortin, connexin sorting protein (CNST), transcript variant X6, mRNA
Sequence ID: XM_037013622.1 Length: 4505 Number of Matches: 1
Range 1: 2494 to 2512 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
38.2 bits(19)	2.3	19/19(100%)	0/19(0%)	Plus/Plus

```
Query  1     GTGAAATGGTTATGTGTGG  19
             ||||||||||||||||||
Sbjct  2494  GTGAAATGGTTATGTGTGG  2512
```

PREDICTED: Manis javanica consortin, connexin sorting protein (CNST), transcript variant X5, mRNA
Sequence ID: XM_037013621.1 Length: 4346 Number of Matches: 1
Range 1: 2335 to 2353 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
38.2 bits(19)	2.3	19/19(100%)	0/19(0%)	Plus/Plus

```
Query  1     GTGAAATGGTTATGTGTGG  19
             ||||||||||||||||||
Sbjct  2335  GTGAAATGGTTATGTGTGG  2353
```

PREDICTED: Manis javanica consortin, connexin sorting protein (CNST), transcript variant X1, mRNA
Sequence ID: XM_037013617.1 Length: 4929 Number of Matches: 1
Range 1: 2918 to 2936 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
38.2 bits(19)	2.3	19/19(100%)	0/19(0%)	Plus/Plus

```
Query  1     GTGAAATGGTTATGTGTGG  19
             ||||||||||||||||||
Sbjct  2918  GTGAAATGGTTATGTGTGG  2936
```

PREDICTED: Manis pentadactyla consortin, connexin sorting protein (CNST), transcript variant X3, mRNA
Sequence ID: XM_036918394.1 Length: 4496 Number of Matches: 1
Range 1: 2496 to 2514 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
38.2 bits(19)	2.3	19/19(100%)	0/19(0%)	Plus/Plus

```
Query  1     GTGAAATGGTTATGTGTGG  19
             ||||||||||||||||||
Sbjct  2496  GTGAAATGGTTATGTGTGG  2514
```

PREDICTED: Egretta garzetta zinc finger protein 639 (ZNF639), transcript variant X3, mRNA
Sequence ID: XM_035898451.1 Length: 4342 Number of Matches: 1
Range 1: 1327 to 1345 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
38.2 bits(19)	2.3	19/19(100%)	0/19(0%)	Plus/Plus

```
Query  1     GTGAAATGGTTATGTGTGG  19
             ||||||||||||||||||
Sbjct  1327  GTGAAATGGTTATGTGTGG  1345
```

PREDICTED: Egretta garzetta zinc finger protein 639 (ZNF639), transcript variant X2, mRNA
Sequence ID: XM_035898450.1 Length: 4199 Number of Matches: 1
Range 1: 1327 to 1345 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
38.2 bits(19)	2.3	19/19(100%)	0/19(0%)	Plus/Plus

```
Query  1     GTGAAATGGTTATGTGTGG  19
             ||||||||||||||||||
Sbjct  1327  GTGAAATGGTTATGTGTGG  1345
```

PREDICTED: Egretta garzetta zinc finger protein 639 (ZNF639), transcript variant X1, mRNA
Sequence ID: XM_035898448.1 Length: 4394 Number of Matches: 1
Range 1: 1327 to 1345 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
38.2 bits(19)	2.3	19/19(100%)	0/19(0%)	Plus/Plus

```
Query  1     GTGAAATGGTTATGTGTGG  19
             ||||||||||||||||||
Sbjct  1327  GTGAAATGGTTATGTGTGG  1345
```

Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/WA-S2917/2020 ORF1ab polyprotein (ORF1ab), ORF1a polyprotein (ORF1ab), surface glycoprotein (S), ORF3a protein (ORF3a), envelope protein (E), membrane glycoprotein (M), ORF6 protein (ORF6), ORF7a protein (ORF7a), ORF7b (ORF7b), ORF8 protein (ORF8), nucleocapsid phosphoprotein (N), and ORF10 protein (ORF10) genes, complete cds

Sequence ID: MW444836.1 Length: 29813 Number of Matches: 1

Range 1: 15393 to 15414 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
36.2 bits(18)	9.2	21/22(95%)	0/22(0%)	Plus/Plus

```
CDS: Putative 1       1          E   M   V   M   C   G   G
Query                 1      GTGAAATGGTTATGTGTGGCGG   22
                             |||||||||| ||||||||||
Sbjct                 15393  GTGAAATGGTCATGTGTGGCGG   15414
CDS:ORF1ab polyprote  5056   S   E   M   V   M   C   G   G
```

5.2 BLAST Research on the RNA SEQUENCE "GTGAGATGGTCATGTGTGGCGG" of the RdRp Gene of the "Bat SARS-like coronavirus isolate bat-SL-CoVZC45, complete genome" – (Hitlist size: 100)

Query: **GTGAGATGGTCATGTGTGGCGG**

Search Parameters	
Program	blastn
Word size	7
Expect value	1000
Hitlist size	100
Match/Mismatch scores	1,-3
Gapcosts	5,2
Filter string	F
Genetic Code	1

100 sequences selected:

Accession Seq	Description	Comomm Name	Max Score	Total Score	Query C.	E value	Per. Id.	Acc. Len
MN312859.1	Bat SARS-like coronavirus isolate SZ140289 RNA-dependent RNA polymerase (RdRp) gene, partial cds	Bat SARS-like coronavirus	44.1	44.1	100%	0.037	100.00%	402
MN312858.1	Bat SARS-like coronavirus isolate SZ140288 RNA-dependent RNA polymerase (RdRp) gene, partial cds	Bat SARS-like coronavirus	44.1	44.1	100%	0.037	100.00%	402
MN312856.1	Bat SARS-like coronavirus isolate NL140494 RNA-dependent RNA polymerase (RdRp) gene, partial cds	Bat SARS-like coronavirus	44.1	44.1	100%	0.037	100.00%	402
MN312846.1	Bat SARS-like coronavirus isolate NL140400 RNA-dependent RNA polymerase (RdRp) gene, partial cds	Bat SARS-like coronavirus	44.1	44.1	100%	0.037	100.00%	407
MN312845.1	Bat SARS-like coronavirus isolate NL140391 RNA-dependent RNA polymerase (RdRp) gene, partial cds	Bat SARS-like coronavirus	44.1	44.1	100%	0.037	100.00%	404
MN312844.1	Bat SARS-like coronavirus isolate NL140352 RNA-dependent RNA polymerase (RdRp) gene, partial cds	Bat SARS-like coronavirus	44.1	44.1	100%	0.037	100.00%	407
MN312843.1	Bat SARS-like coronavirus isolate NL140346 RNA-dependent RNA polymerase (RdRp) gene, partial cds	Bat SARS-like coronavirus	44.1	44.1	100%	0.037	100.00%	399
MN312831.1	Bat SARS-like coronavirus isolate HZ13488 RNA-dependent RNA polymerase (RdRp) gene, partial cds	Bat SARS-like coronavirus	44.1	44.1	100%	0.037	100.00%	433
MN312830.1	Bat SARS-like coronavirus isolate HZ13484 RNA-dependent RNA polymerase (RdRp) gene, partial cds	Bat SARS-like coronavirus	44.1	44.1	100%	0.037	100.00%	442
MN312829.1	Bat SARS-like coronavirus isolate HZ13479 RNA-dependent RNA polymerase (RdRp) gene, partial cds	Bat SARS-like coronavirus	44.1	44.1	100%	0.037	100.00%	423
MN312821.1	Bat SARS-like coronavirus isolate 9408 RNA-dependent RNA polymerase (RdRp) gene, partial cds	Bat SARS-like coronavirus	44.1	44.1	100%	0.037	100.00%	358
MN312819.1	Bat SARS-like coronavirus isolate 9403 RNA-dependent RNA polymerase (RdRp) gene, partial cds	Bat SARS-like coronavirus	44.1	44.1	100%	0.037	100.00%	360
MN312818.1	Bat SARS-like coronavirus isolate 9400 RNA-dependent RNA polymerase (RdRp) gene, partial cds	Bat SARS-like coronavirus	44.1	44.1	100%	0.037	100.00%	364
MN312742.1	Bat SARS-like coronavirus isolate 8794 RNA-dependent RNA polymerase (RdRp) gene, partial cds	Bat SARS-like coronavirus	44.1	44.1	100%	0.037	100.00%	343
MN312741.1	Bat SARS-like coronavirus isolate 8586 RNA-dependent RNA polymerase (RdRp) gene, partial cds	Bat SARS-like coronavirus	44.1	44.1	100%	0.037	100.00%	377
MN312739.1	Bat SARS-like coronavirus isolate 8561 RNA-dependent RNA polymerase (RdRp) gene, partial cds	Bat SARS-like coronavirus	44.1	44.1	100%	0.037	100.00%	376
MN312738.1	Bat SARS-like coronavirus isolate 8548 RNA-dependent RNA polymerase (RdRp) gene, partial cds	Bat SARS-like coronavirus	44.1	44.1	100%	0.037	100.00%	377

MN312711.1	Bat SARS-like coronavirus isolate 151569 RNA-dependent RNA polymerase (RdRp) gene, partial cds	Bat SARS-like coronavirus	44.1	44.1	100%	0.037	100.00%	402
MN312710.1	Bat SARS-like coronavirus isolate 151514 RNA-dependent RNA polymerase (RdRp) gene, partial cds	Bat SARS-like coronavirus	44.1	44.1	100%	0.037	100.00%	411
MN312708.1	Bat SARS-like coronavirus isolate 151512 RNA-dependent RNA polymerase (RdRp) gene, partial cds	Bat SARS-like coronavirus	44.1	44.1	100%	0.037	100.00%	411
MN312707.1	Bat SARS-like coronavirus isolate 151507 RNA-dependent RNA polymerase (RdRp) gene, partial cds	Bat SARS-like coronavirus	44.1	44.1	100%	0.037	100.00%	409
MN312706.1	Bat SARS-like coronavirus isolate 151503 RNA-dependent RNA polymerase (RdRp) gene, partial cds	Bat SARS-like coronavirus	44.1	44.1	100%	0.037	100.00%	409
MN312705.1	Bat SARS-like coronavirus isolate 151493 RNA-dependent RNA polymerase (RdRp) gene, partial cds	Bat SARS-like coronavirus	44.1	44.1	100%	0.037	100.00%	410
MN312704.1	Bat SARS-like coronavirus isolate 151491 RNA-dependent RNA polymerase (RdRp) gene, partial cds	Bat SARS-like coronavirus	44.1	44.1	100%	0.037	100.00%	421
MN312703.1	Bat SARS-like coronavirus isolate 151388 RNA-dependent RNA polymerase (RdRp) gene, partial cds	Bat SARS-like coronavirus	44.1	44.1	100%	0.037	100.00%	408
MN312702.1	Bat SARS-like coronavirus isolate 151334 RNA-dependent RNA polymerase (RdRp) gene, partial cds	Bat SARS-like coronavirus	44.1	44.1	100%	0.037	100.00%	411
MN312700.1	Bat SARS-like coronavirus isolate 151270 RNA-dependent RNA polymerase (RdRp) gene, partial cds	Bat SARS-like coronavirus	44.1	44.1	100%	0.037	100.00%	411
MN312698.1	Bat SARS-like coronavirus isolate 151267 RNA-dependent RNA polymerase (RdRp) gene, partial cds	Bat SARS-like coronavirus	44.1	44.1	100%	0.037	100.00%	408
MN312697.1	Bat SARS-like coronavirus isolate 151262 RNA-dependent RNA polymerase (RdRp) gene, partial cds	Bat SARS-like coronavirus	44.1	44.1	100%	0.037	100.00%	410
MN312696.1	Bat SARS-like coronavirus isolate 151256 RNA-dependent RNA polymerase (RdRp) gene, partial cds	Bat SARS-like coronavirus	44.1	44.1	100%	0.037	100.00%	409
MN312695.1	Bat SARS-like coronavirus isolate 151239 RNA-dependent RNA polymerase (RdRp) gene, partial cds	Bat SARS-like coronavirus	44.1	44.1	100%	0.037	100.00%	411
MN312694.1	Bat SARS-like coronavirus isolate 151200 RNA-dependent RNA polymerase (RdRp) gene, partial cds	Bat SARS-like coronavirus	44.1	44.1	100%	0.037	100.00%	408
MN312693.1	Bat SARS-like coronavirus isolate 151199 RNA-dependent RNA polymerase (RdRp) gene, partial cds	Bat SARS-like coronavirus	44.1	44.1	100%	0.037	100.00%	407
MN312692.1	Bat SARS-like coronavirus isolate 151162 RNA-dependent RNA polymerase (RdRp) gene, partial cds	Bat SARS-like coronavirus	44.1	44.1	100%	0.037	100.00%	408
MN312688.1	Bat SARS-like coronavirus isolate 8363 RNA-dependent RNA polymerase (RdRp) gene, partial cds	Bat SARS-like coronavirus	44.1	44.1	100%	0.037	100.00%	388
MN312664.1	Bat SARS-like coronavirus isolate 141600 RNA-dependent RNA polymerase (RdRp) gene, partial cds	Bat SARS-like coronavirus	44.1	44.1	100%	0.037	100.00%	382
MN312663.1	Bat SARS-like coronavirus isolate 141599 RNA-dependent RNA polymerase (RdRp) gene, partial cds	Bat SARS-like coronavirus	44.1	44.1	100%	0.037	100.00%	370
MN312662.1	Bat SARS-like coronavirus isolate 141598 RNA-dependent RNA polymerase (RdRp) gene, partial cds	Bat SARS-like coronavirus	44.1	44.1	100%	0.037	100.00%	373
MN312661.1	Bat SARS-like coronavirus isolate 141582 RNA-dependent RNA polymerase (RdRp) gene, partial cds	Bat SARS-like coronavirus	44.1	44.1	100%	0.037	100.00%	382
MN312660.1	Bat SARS-like coronavirus isolate 141579 RNA-dependent RNA polymerase (RdRp) gene, partial cds	Bat SARS-like coronavirus	44.1	44.1	100%	0.037	100.00%	383
MN312659.1	Bat SARS-like coronavirus isolate 141577 RNA-dependent RNA polymerase (RdRp) gene, partial cds	Bat SARS-like coronavirus	44.1	44.1	100%	0.037	100.00%	383
MN312657.1	Bat SARS-like coronavirus isolate 141570 RNA-dependent RNA polymerase (RdRp) gene, partial cds	Bat SARS-like coronavirus	44.1	44.1	100%	0.037	100.00%	384
MN312656.1	Bat SARS-like coronavirus isolate 141569 RNA-dependent RNA polymerase (RdRp) gene, partial cds	Bat SARS-like coronavirus	44.1	44.1	100%	0.037	100.00%	382
MN312655.1	Bat SARS-like coronavirus isolate 141567 RNA-dependent RNA polymerase (RdRp) gene, partial cds	Bat SARS-like coronavirus	44.1	44.1	100%	0.037	100.00%	370
MN312654.1	Bat SARS-like coronavirus isolate	Bat SARS-like	44.1	44.1	100%	0.037	100.00%	364

	141470 RNA-dependent RNA polymerase (RdRp) gene, partial cds Bat SARS-like coronavirus isolate	coronavirus						
MN312653.1	141464 RNA-dependent RNA polymerase (RdRp) gene, partial cds Bat SARS-like coronavirus isolate	Bat SARS-like coronavirus	44.1	44.1	100%	0.037	100.00%	374
MN312652.1	141460 RNA-dependent RNA polymerase (RdRp) gene, partial cds Bat SARS-like coronavirus isolate	Bat SARS-like coronavirus	44.1	44.1	100%	0.037	100.00%	327
MN312651.1	141456 RNA-dependent RNA polymerase (RdRp) gene, partial cds Bat SARS-like coronavirus isolate	Bat SARS-like coronavirus	44.1	44.1	100%	0.037	100.00%	366
MN312650.1	141455 RNA-dependent RNA polymerase (RdRp) gene, partial cds Bat SARS-like coronavirus isolate	Bat SARS-like coronavirus	44.1	44.1	100%	0.037	100.00%	371
MN312645.1	141376 RNA-dependent RNA polymerase (RdRp) gene, partial cds Bat SARS-like coronavirus isolate	Bat SARS-like coronavirus	44.1	44.1	100%	0.037	100.00%	379
MN312644.1	141375 RNA-dependent RNA polymerase (RdRp) gene, partial cds Bat SARS-like coronavirus isolate	Bat SARS-like coronavirus	44.1	44.1	100%	0.037	100.00%	362
MN312643.1	141369 RNA-dependent RNA polymerase (RdRp) gene, partial cds Bat SARS-like coronavirus isolate	Bat SARS-like coronavirus	44.1	44.1	100%	0.037	100.00%	362
MN312642.1	141352 RNA-dependent RNA polymerase (RdRp) gene, partial cds Bat SARS-like coronavirus isolate	Bat SARS-like coronavirus	44.1	44.1	100%	0.037	100.00%	349
MN312641.1	141350 RNA-dependent RNA polymerase (RdRp) gene, partial cds Bat SARS-like coronavirus isolate	Bat SARS-like coronavirus	44.1	44.1	100%	0.037	100.00%	367
MN312640.1	141344 RNA-dependent RNA polymerase (RdRp) gene, partial cds Bat SARS-like coronavirus isolate	Bat SARS-like coronavirus	44.1	44.1	100%	0.037	100.00%	367
MN312639.1	141342 RNA-dependent RNA polymerase (RdRp) gene, partial cds Bat SARS-like coronavirus isolate	Bat SARS-like coronavirus	44.1	44.1	100%	0.037	100.00%	346
MN312638.1	141341 RNA-dependent RNA polymerase (RdRp) gene, partial cds Bat SARS-like coronavirus isolate	Bat SARS-like coronavirus	44.1	44.1	100%	0.037	100.00%	359
MN312637.1	141335 RNA-dependent RNA polymerase (RdRp) gene, partial cds Bat SARS-like coronavirus isolate	Bat SARS-like coronavirus	44.1	44.1	100%	0.037	100.00%	376
MN312606.1	7330 RNA-dependent RNA polymerase (RdRp) gene, partial cds Bat SARS-like coronavirus isolate	Bat SARS-like coronavirus	44.1	44.1	100%	0.037	100.00%	402
MN312605.1	7326 RNA-dependent RNA polymerase (RdRp) gene, partial cds Bat SARS-like coronavirus isolate	Bat SARS-like coronavirus	44.1	44.1	100%	0.037	100.00%	403
MN312604.1	7325 RNA-dependent RNA polymerase (RdRp) gene, partial cds Bat SARS-like coronavirus isolate	Bat SARS-like coronavirus	44.1	44.1	100%	0.037	100.00%	405
MN312603.1	6533 RNA-dependent RNA polymerase (RdRp) gene, partial cds Bat SARS-like coronavirus isolate	Bat SARS-like coronavirus	44.1	44.1	100%	0.037	100.00%	406
MN312602.1	6530 RNA-dependent RNA polymerase (RdRp) gene, partial cds Bat SARS-like coronavirus isolate	Bat SARS-like coronavirus	44.1	44.1	100%	0.037	100.00%	399
MN312601.1	6500 RNA-dependent RNA polymerase (RdRp) gene, partial cds Bat SARS-like coronavirus isolate	Bat SARS-like coronavirus	44.1	44.1	100%	0.037	100.00%	400
MN312599.1	6301 RNA-dependent RNA polymerase (RdRp) gene, partial cds Bat SARS-like coronavirus isolate	Bat SARS-like coronavirus	44.1	44.1	100%	0.037	100.00%	371
MN312598.1	6278 RNA-dependent RNA polymerase (RdRp) gene, partial cds Bat SARS-like coronavirus isolate	Bat SARS-like coronavirus	44.1	44.1	100%	0.037	100.00%	378
MN312597.1	6266 RNA-dependent RNA polymerase (RdRp) gene, partial cds Bat SARS-like coronavirus isolate	Bat SARS-like coronavirus	44.1	44.1	100%	0.037	100.00%	402
MN312596.1	6255 RNA-dependent RNA polymerase (RdRp) gene, partial cds Bat SARS-like coronavirus isolate	Bat SARS-like coronavirus	44.1	44.1	100%	0.037	100.00%	402
MN312595.1	6233 RNA-dependent RNA polymerase (RdRp) gene, partial cds	Bat SARS-like coronavirus	44.1	44.1	100%	0.037	100.00%	375
MT434760.1	**Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/IND/nimh-1598/2020 ORF1ab polyprotein (ORF1ab), ORF1a polyprotein (ORF1ab), surface**	Severe acute respiratory syndrome coronavirus 2	44.1	44.1	100%	0.037	100.00%	29833

	glycoprotein (S), ORF3a protein (ORF3a), envelope protein (E), membrane glycoprotein (M), ORF6 protein (ORF6), ORF7a protein (ORF7a), ORF7b (ORF7b), ORF8 protein (ORF8), nucleocapsid phosphoprotein (N), and ORF10 protein (ORF10) genes, complete cds							
MT308984.1	***Mutant SARS coronavirus Urbani clone SARS-Urbani-MA_SHC014-spike, complete genome***	SARS coronavirus Urbani	44.1	44.1	100%	0.037	100.00%	29730
MK211378.1	Coronavirus BtRs-BetaCoV/YN2018D, complete genome	Coronavirus BtRs-BetaCoV/YN2018D	44.1	44.1	100%	0.037	100.00%	30213
MK211377.1	Coronavirus BtRs-BetaCoV/YN2018C, complete genome	Coronavirus BtRs-BetaCoV/YN2018C	44.1	44.1	100%	0.037	100.00%	29689
MK211376.1	Coronavirus BtRs-BetaCoV/YN2018B, complete genome	Coronavirus BtRs-BetaCoV/YN2018B	44.1	44.1	100%	0.037	100.00%	30256
MK211375.1	Coronavirus BtRs-BetaCoV/YN2018A, complete genome	Coronavirus BtRs-BetaCoV/YN2018A	44.1	44.1	100%	0.037	100.00%	29698
MK211374.1	Coronavirus BtRl-BetaCoV/SC2018, complete genome	Coronavirus BtRl-BetaCoV/SC2018	44.1	44.1	100%	0.037	100.00%	29648
MK062184.1	SARS coronavirus Urbani isolate icSARS-C7-MA, complete genome	SARS coronavirus Urbani	44.1	44.1	100%	0.037	100.00%	29874
MK062183.1	SARS coronavirus Urbani isolate icSARS-C7, complete genome	SARS coronavirus Urbani	44.1	44.1	100%	0.037	100.00%	29874
MK062182.1	SARS coronavirus Urbani isolate icSARS-C3-MA, complete genome	SARS coronavirus Urbani	44.1	44.1	100%	0.037	100.00%	29727
MK062181.1	SARS coronavirus Urbani isolate icSARS-C3, complete genome	SARS coronavirus Urbani	44.1	44.1	100%	0.037	100.00%	29727
MK062180.1	SARS coronavirus Urbani isolate icSARS-MA, complete genome	SARS coronavirus Urbani	44.1	44.1	100%	0.037	100.00%	29727
MK062179.1	SARS coronavirus Urbani isolate icSARS, complete genome	SARS coronavirus Urbani	44.1	44.1	100%	0.037	100.00%	29727
MG772934.1	Bat SARS-like coronavirus isolate bat-SL-CoVZXC21, complete genome	Bat SARS-like coronavirus	44.1	44.1	100%	0.037	100.00%	29732
MG772933.1	**Bat SARS-like coronavirus isolate bat-SL-CoVZC45, complete genome**	Bat SARS-like coronavirus	44.1	44.1	100%	0.037	100.00%	29802
MG772932.1	Bat SARS-like coronavirus isolate bat-SL-CoVZXC1 RNA-dependent RNA polymerase gene, partial cds	Bat SARS-like coronavirus	44.1	44.1	100%	0.037	100.00%	437
MG772931.1	Bat SARS-like coronavirus isolate bat-SL-CoVZXC2 RNA-dependent RNA polymerase gene, partial cds	Bat SARS-like coronavirus	44.1	44.1	100%	0.037	100.00%	381
MG772930.1	Bat SARS-like coronavirus isolate bat-SL-CoVZXC4 RNA-dependent RNA polymerase gene, partial cds	Bat SARS-like coronavirus	44.1	44.1	100%	0.037	100.00%	405
MG772929.1	Bat SARS-like coronavirus isolate bat-SL-CoVZXC5 RNA-dependent RNA polymerase gene, partial cds	Bat SARS-like coronavirus	44.1	44.1	100%	0.037	100.00%	438
MG772928.1	Bat SARS-like coronavirus isolate bat-SL-CoVZXC6 RNA-dependent RNA polymerase gene, partial cds	Bat SARS-like coronavirus	44.1	44.1	100%	0.037	100.00%	430
MG772927.1	Bat SARS-like coronavirus isolate bat-SL-CoVZXC7 RNA-dependent RNA polymerase gene, partial cds	Bat SARS-like coronavirus	44.1	44.1	100%	0.037	100.00%	407
MG772924.1	Bat SARS-like coronavirus isolate bat-SL-CoVZXC12 RNA-dependent RNA polymerase gene, partial cds	Bat SARS-like coronavirus	44.1	44.1	100%	0.037	100.00%	437
MG772921.1	Bat SARS-like coronavirus isolate bat-SL-CoVZXC21 RNA-dependent RNA polymerase gene, partial cds	Bat SARS-like coronavirus	44.1	44.1	100%	0.037	100.00%	428
MG772920.1	Bat SARS-like coronavirus isolate bat-SL-CoVZXC22 RNA-dependent RNA polymerase gene, partial cds	Bat SARS-like coronavirus	44.1	44.1	100%	0.037	100.00%	418
MG772913.1	Bat SARS-like coronavirus isolate bat-SL-CoVZXC29 RNA-dependent RNA polymerase gene, partial cds	Bat SARS-like coronavirus	44.1	44.1	100%	0.037	100.00%	405
MG772911.1	Bat SARS-like coronavirus isolate bat-SL-CoVZXC33 RNA-dependent RNA polymerase gene, partial cds	Bat SARS-like coronavirus	44.1	44.1	100%	0.037	100.00%	438

MG772910.1	Bat SARS-like coronavirus isolate bat-SL-CoVZXC34 RNA-dependent RNA polymerase gene, partial cds	Bat SARS-like coronavirus	44.1	44.1	100%	0.037	100.00%	418
MT782115.1	***Mutant Bat SARS coronavirus HKU3 isolate icHKU3-SRBD-MA, complete genome***	Bat SARS coronavirus HKU3	44.1	44.1	100%	0.037	100.00%	29781
MT782114.1	***Mutant Bat SARS coronavirus HKU3 isolate icHKU3-SRBD, complete genome***	Bat SARS coronavirus HKU3	44.1	44.1	100%	0.037	100.00%	29781
MT439595.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/IND/nimh-4376/2020 ORF1ab polyprotein (ORF1ab), ORF1a polyprotein (ORF1ab), surface glycoprotein (S), ORF3a protein (ORF3a), envelope protein (E), membrane glycoprotein (M), ORF6 protein (ORF6), ORF7a protein (ORF7a), ORF7b (ORF7b), ORF8 protein (ORF8), nucleocapsid phosphoprotein (N), and ORF10 protein (ORF10) genes, complete cds	Severe acute respiratory syndrome coronavirus 2	42.1	42.1	95%	0.15	100.00%	29814
MW276628.1	**Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/CA-CZB-11679/2020 ORF1ab polyprotein (ORF1ab), ORF1a polyprotein (ORF1ab), surface glycoprotein (S), ORF3a protein (ORF3a), envelope protein (E), membrane glycoprotein (M), ORF6 protein (ORF6), ORF7a protein (ORF7a), ORF7b (ORF7b), ORF8 protein (ORF8), nucleocapsid phosphoprotein (N), and ORF10 protein (ORF10) genes, complete cds**	Severe acute respiratory syndrome coronavirus 2	40.1	40.1	100%	0.59	95.45%	29741

ALIGNMENT VIEW FLAT QUERY-ANCHORED WITH LETTERS FOR IDENTITIES

Query range 1: 1 to 22

Query	1	GTGAGATGGTCATGTGTGGCGG	22
MN312859.1	122	GTGAGATGGTCATGTGTGGCGG	143
MN312858.1	122	GTGAGATGGTCATGTGTGGCGG	143
MN312856.1	123	GTGAGATGGTCATGTGTGGCGG	144
MN312846.1	128	GTGAGATGGTCATGTGTGGCGG	149
MN312845.1	124	GTGAGATGGTCATGTGTGGCGG	145
MN312844.1	124	GTGAGATGGTCATGTGTGGCGG	145
MN312843.1	121	GTGAGATGGTCATGTGTGGCGG	142
MN312831.1	152	GTGAGATGGTCATGTGTGGCGG	173
MN312830.1	152	GTGAGATGGTCATGTGTGGCGG	173
MN312829.1	142	GTGAGATGGTCATGTGTGGCGG	163
MN312821.1	117	GTGAGATGGTCATGTGTGGCGG	138
MN312819.1	117	GTGAGATGGTCATGTGTGGCGG	138
MN312818.1	117	GTGAGATGGTCATGTGTGGCGG	138
MN312742.1	122	GTGAGATGGTCATGTGTGGCGG	143
MN312741.1	130	GTGAGATGGTCATGTGTGGCGG	151
MN312739.1	129	GTGAGATGGTCATGTGTGGCGG	150
MN312738.1	129	GTGAGATGGTCATGTGTGGCGG	150
MN312711.1	122	GTGAGATGGTCATGTGTGGCGG	143
MN312710.1	130	GTGAGATGGTCATGTGTGGCGG	151
MN312708.1	130	GTGAGATGGTCATGTGTGGCGG	151
MN312707.1	128	GTGAGATGGTCATGTGTGGCGG	149
MN312706.1	128	GTGAGATGGTCATGTGTGGCGG	149
MN312705.1	130	GTGAGATGGTCATGTGTGGCGG	151
MN312704.1	140	GTGAGATGGTCATGTGTGGCGG	161
MN312703.1	128	GTGAGATGGTCATGTGTGGCGG	149
MN312702.1	131	GTGAGATGGTCATGTGTGGCGG	152
MN312700.1	130	GTGAGATGGTCATGTGTGGCGG	151
MN312698.1	127	GTGAGATGGTCATGTGTGGCGG	148
MN312697.1	130	GTGAGATGGTCATGTGTGGCGG	151
MN312696.1	128	GTGAGATGGTCATGTGTGGCGG	149
MN312695.1	130	GTGAGATGGTCATGTGTGGCGG	151
MN312694.1	128	GTGAGATGGTCATGTGTGGCGG	149
MN312693.1	128	GTGAGATGGTCATGTGTGGCGG	149
MN312692.1	128	GTGAGATGGTCATGTGTGGCGG	149
MN312688.1	147	GTGAGATGGTCATGTGTGGCGG	168
MN312664.1	141	GTGAGATGGTCATGTGTGGCGG	162
MN312663.1	130	GTGAGATGGTCATGTGTGGCGG	151
MN312662.1	130	GTGAGATGGTCATGTGTGGCGG	151
MN312661.1	141	GTGAGATGGTCATGTGTGGCGG	162
MN312660.1	142	GTGAGATGGTCATGTGTGGCGG	163
MN312659.1	142	GTGAGATGGTCATGTGTGGCGG	163
MN312657.1	142	GTGAGATGGTCATGTGTGGCGG	163
MN312656.1	141	GTGAGATGGTCATGTGTGGCGG	162
MN312655.1	129	GTGAGATGGTCATGTGTGGCGG	150
MN312654.1	131	GTGAGATGGTCATGTGTGGCGG	152
MN312653.1	141	GTGAGATGGTCATGTGTGGCGG	162
MN312652.1	128	GTGAGATGGTCATGTGTGGCGG	149
MN312651.1	125	GTGAGATGGTCATGTGTGGCGG	146
MN312650.1	127	GTGAGATGGTCATGTGTGGCGG	148
MN312645.1	131	GTGAGATGGTCATGTGTGGCGG	152
MN312644.1	131	GTGAGATGGTCATGTGTGGCGG	152
MN312643.1	130	GTGAGATGGTCATGTGTGGCGG	151
MN312642.1	131	GTGAGATGGTCATGTGTGGCGG	152

MN312641.1	130	GTGAGATGGTCATGTGTGGCGG	151
MN312640.1	130	GTGAGATGGTCATGTGTGGCGG	151
MN312639.1	129	GTGAGATGGTCATGTGTGGCGG	150
MN312638.1	130	GTGAGATGGTCATGTGTGGCGG	151
MN312637.1	128	GTGAGATGGTCATGTGTGGCGG	149
MN312606.1	122	GTGAGATGGTCATGTGTGGCGG	143
MN312605.1	122	GTGAGATGGTCATGTGTGGCGG	143
MN312604.1	127	GTGAGATGGTCATGTGTGGCGG	148
MN312603.1	127	GTGAGATGGTCATGTGTGGCGG	148
MN312602.1	120	GTGAGATGGTCATGTGTGGCGG	141
MN312601.1	126	GTGAGATGGTCATGTGTGGCGG	147
MN312599.1	129	GTGAGATGGTCATGTGTGGCGG	150
MN312598.1	128	GTGAGATGGTCATGTGTGGCGG	149
MN312597.1	122	GTGAGATGGTCATGTGTGGCGG	143
MN312596.1	122	GTGAGATGGTCATGTGTGGCGG	143
MN312595.1	126	GTGAGATGGTCATGTGTGGCGG	147
MT434760.1	15403	GTGAGATGGTCATGTGTGGCGG	15424
MT308984.1	15361	GTGAGATGGTCATGTGTGGCGG	15382
MK211378.1	15361	GTGAGATGGTCATGTGTGGCGG	15382
MK211377.1	15359	GTGAGATGGTCATGTGTGGCGG	15380
MK211376.1	15360	GTGAGATGGTCATGTGTGGCGG	15381
MK211375.1	15362	GTGAGATGGTCATGTGTGGCGG	15383
MK211374.1	15324	GTGAGATGGTCATGTGTGGCGG	15345
MK062184.1	15362	GTGAGATGGTCATGTGTGGCGG	15383
MK062183.1	15362	GTGAGATGGTCATGTGTGGCGG	15383
MK062182.1	15361	GTGAGATGGTCATGTGTGGCGG	15382
MK062181.1	15361	GTGAGATGGTCATGTGTGGCGG	15382
MK062180.1	15361	GTGAGATGGTCATGTGTGGCGG	15382
MK062179.1	15361	GTGAGATGGTCATGTGTGGCGG	15382
MG772934.1	15352	GTGAGATGGTCATGTGTGGCGG	15373
MG772933.1	15418	GTGAGATGGTCATGTGTGGCGG	15439
MG772932.1	144	GTGAGATGGTCATGTGTGGCGG	165
MG772931.1	87	GTGAGATGGTCATGTGTGGCGG	108
MG772930.1	124	GTGAGATGGTCATGTGTGGCGG	145
MG772929.1	144	GTGAGATGGTCATGTGTGGCGG	165
MG772928.1	136	GTGAGATGGTCATGTGTGGCGG	157
MG772927.1	124	GTGAGATGGTCATGTGTGGCGG	145
MG772924.1	145	GTGAGATGGTCATGTGTGGCGG	166
MG772921.1	134	GTGAGATGGTCATGTGTGGCGG	155
MG772920.1	124	GTGAGATGGTCATGTGTGGCGG	145
MG772913.1	124	GTGAGATGGTCATGTGTGGCGG	145
MG772911.1	144	GTGAGATGGTCATGTGTGGCGG	165
MG772910.1	124	GTGAGATGGTCATGTGTGGCGG	145
MT782115.1	15343	GTGAGATGGTCATGTGTGGCGG	15364
MT782114.1	15343	GTGAGATGGTCATGTGTGGCGG	15364
MT439595.1	15403	TGAGATGGTCATGTGTGGCGG	15423
MW276628.1	15377	GTGA**R**ATGGTCATGTGTGGCGG	15398

Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/IND/nimh-1598/2020 ORF1ab polyprotein (ORF1ab), ORF1a polyprotein (ORF1ab), surface glycoprotein (S), ORF3a protein (ORF3a), envelope protein (E), membrane glycoprotein (M), ORF6 protein ...

GenBank: MT434760.1

FASTA Graphics

```
LOCUS       MT434760                 22 bp    RNA     linear   VRL 21-OCT-2020
DEFINITION  Severe acute respiratory syndrome coronavirus 2 isolate
            SARS-CoV-2/human/IND/nimh-1598/2020 ORF1ab polyprotein (ORF1ab),
            ORF1a polyprotein (ORF1ab), surface glycoprotein (S), ORF3a protein
            (ORF3a), envelope protein (E), membrane glycoprotein (M), ORF6
            protein (ORF6), ORF7a protein (ORF7a), ORF7b (ORF7b), ORF8 protein
            (ORF8), nucleocapsid phosphoprotein (N), and ORF10 protein (ORF10)
            genes, complete cds.
ACCESSION   MT434760 REGION: 15403..15424
VERSION     MT434760.1
KEYWORDS    .
SOURCE      Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)
  ORGANISM  Severe acute respiratory syndrome coronavirus 2
            Viruses; Riboviria; Orthornavirae; Pisuviricota; Pisoniviricetes;
            Nidovirales; Cornidovirineae; Coronaviridae; Orthocoronavirinae;
            Betacoronavirus; Sarbecovirus.
REFERENCE   1  (bases 1 to 22)
  AUTHORS   Pattabiraman,C., Reddy,V., Harsha,P.K., Rasheed,R., Hameed,S.,
            Venkataswamy,M., Desai,A. and Vasanthapuram,R.
  TITLE     SARS-CoV-2 in Karnataka, India
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 22)
  AUTHORS   Pattabiraman,C., Reddy,V., Harsha,P.K., Rasheed,R., Hameed,S.,
            Venkataswamy,M., Desai,A. and Vasanthapuram,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-MAY-2020) Neurovirology, National Institute of Mental
            Health and Neurosciences, Hosur Road, Bangalore, Karnataka 560029,
            India
FEATURES             Location/Qualifiers
     source          1..22
                     /organism="Severe acute respiratory syndrome coronavirus
                     2"
                     /mol type="genomic RNA"
                     /isolate="SARS-CoV-2/human/IND/nimh-1598/2020"
                     /isolation source="nasopharyngeal/oropharyngeal swab"
                     /host="Homo sapiens"
                     /db_xref="taxon:2697049"
                     /country="India"
                     /collection_date="2020-04-17"
     gene            <1..>22
```

/gene="ORF1ab"

CDS <1..>22

/gene="**ORF1ab**"

/ribosomal slippage

/codon_start=3

/product="**ORF1ab polyprotein**"

/protein_id="QJQ27876.1"

/translation="MESLVPGFNEKTHVQLSLPVLQVRDVLVRGFGDSVEEVLSEARQ
HLKDGTCGLVEVEKGVLPQLEQPYVFIKRSDARTAPHGHVMVELVAELEGIQYGRSGE
TLGVLVPHVGEIPVAYRKVLLRKNGNKGAGGHSYGADLKSFDLGDELGTDPYEDFQEN
WNTKHSSGVTRELMRELNGGAYTRYVDNNFCGPDGYPLECIKDLLARAGKASCTLSEQ
LDFIDTKRGVYCCREHEHEIAWYTERSEKSYELQTPFEIKLAKKFDTFNGECPNFVFP
LNSIIKTIQPRVEKKKLDGFMGRIRSVYPVASPNECNQMCLSTLMKCDHCGETSWQTG
DFVKATCEFCGTENLTKEGATTCGYLPQNAVVKIYCPACHNSEVGPEHSLAEYHNESG
LKTILRKGGRTIAFGGCVFSYVGCHNKCAYWVPRASANIGCNHTGVVGEGSEGLNDNL
LEILQKEKVNINIVGDFKLNEEIAIILASFSASTSAFVETVKGLDYKAFKQIVESCGN
FKVTKGKAKKGAWNIGEQKSILSPLYAFASEAARVVRSIFSRTLETAQNSVRVLQKAA
ITILDGISQYSLRLIDAMMFTSDLATNNLVVMAYITGGVVQLTSQWLTNIFGTVYEKL
KPVLDWLEEKFKEGVEFLRDGWEIVKFISTCACEIVGGQIVTCAKEIKESVQTFFKLV
NKFLALCADSIIIGGAKLKALNLGETFVTHSKGLYRKCVKSREETGLLMPLKAPKEII
FLEGETLPTEVLTEEVVLKTGDLQPLEQPTSEAVEAPLVGTPVCINGLMLLEIKDTEK
YCALAPNMMVTNNTFTLKGGAPTKVTFGDDTVIEVQGYKSVNITFELDERIDKVLNEK
CSAYTVELGTEVNEFACVVADAVIKTLQPVSELLTPLGIDLDEWSMATYYLFDESGEF
KLASHMYCSFYPPDEDEEEGDCEEEEFEPSTQYEYGTEDDYQGKPLEFGATSAALQSE
EEQEEDWLDDDSQQTVGQQDGSEDNQTTTIQTIVEVQPQLEMELTPVVQTIEVNSFSG
YLKLTDNVYIKNADIVEEAKKVKPTVVVNAANVYLKHGGGVAGALNKATNNAMQVESD
DYIATNGPLKVGGSCVLSGHNLAKHCLHVVGPNVNKGEDIQLLKSAYENFNQHEVLLA
PLLSAGIFGADPIHSLRVCVDTVRTNVYLAVFDKNLYDKLVSSFLEMKSEKQVEQKIA
EIPKEEVKPFITESKPSVEQRKQDDKKIKACVEEVTTTLEETKFLTENLLLYIDINGN
LHPDSATLVSDIDITFLKKDAPYIVGDVVQEGVLTAVVIPTKKAGGTTEMLAKALRKV
PTDNYITTYPGQGLNGYTVEEAKTVLKKCKSAFYILPSIISNEKQEILGTVSWNLREM
LAHAEETRKLMPVCVETKAIVSTIQRKYKGIKIQEGVVDYGARFYFYTSKTTVASLIN
TLNDLNETLVTMPLGYVTHGLNLEEAARYMRSLKVPATVSVSSPDAVTAYNGYLTSSS
KTPEEHFIETISLAGSYKDWSYSGQSTQLGIEFLKRGDKSVYYTSNPTTFHLDGEVIT
FDNLKTLLSLREVRTIKVFTTVDNINLHTQVVDMSMTYGQQFGPTYLDGADVTKIKPH
NXHEGKTFYVLPNDDTLRVEAFEYYHTTDPSFLGRYMSALNHTKXXXXXXXXGLTSIK
WADNNCYLATALLTLQQIELKFNPPALQDAYYRARAGEAANFCALILAYCNKTVGELG
DVRETMSYLFQHANLDSCKRVLNVVCKTCGQQQTTLKGVEAVMYMGTLSYEQFKKGVQ
IPCTCGKQATKYLVQQESPFVMMSAPPAQYELKHGTFTCASEYTGNYQCGHYKHITSK
ETLYCIDGALLTKSSEYKGPITDVFYKENSYTTTIKPVTYKLDGVVCTEIDPKLDNYY
KKDNSYFTEQPIDLVPNQPYPNASFDNFKFVCDNIKFADDLNQLTGYKKPASRELKVT
FFPDLNGDVVAIDYKHYTPSFKKGAKLLHKPIVWHVNNATNKATYKPNTWCIRCLWST
KPVETSNSFDVLKSEDAQGMDNLACEDLKPVSEEVVENPTIQKDVLECNVKTTEVVGD
IILKPANNSLKITEEVGHTDLMAAYVDNSSLTIKKPNELSRVLGLKTLATHGLAAVNS

VPWDTIANYAKPFLNKVVSTTTNIVTRCLNRVCTNYMPYFFTLLLQLCTFTRSTNSRI
KASMPTTIAKNTVKSVGKFCLEASFNYLKSPNFSKLINIIIWFLLLSVCLGSLIYSTA
ALGVLMSNLGMPSYCTGYREGYLNSTNVTIATYCTGSIPCSVCLSGLDSLDTYPSLET
IQITISSFKWDLTAFGLVAEWFLAYILFTRFFYVLGLAAIMQLFFSYFAVHFISNSWL
MWLIINLVQMAPISAMVRMYIFFASFYYVWKSYVHVVDGCNSSTCMMCYKRNRATRVE
CTTIVNGVRRSFYVYANGGKGFCKLHNWNCVNCDTFCAGSTFISDEVARDLSLQFKRP
INPTDQSSYIVDSVTVKNGSIHLYFDKAGQKTYERHSLSHFVNLDNLRANNTKGSLPI
NVIVFDGKSKCEESSAKSASVYYSQLMCQPILLLDQALVSDVGDSAEVAVKMFDAYVN
TFSSTFNVPMEKLKTLVATAEAELAKNVSLDNVLSTFISAARQGFVDSDVETKDVVEC
LKLSHQSDIEVTGDSCNNYMLTYNKVENMTPRDLGACIDCSARHINAQVAKSHNIALI
WNVKDFMSLSEQLRKQIRSAAKKNNLPFKLTCATTRQVVNVVTTKIALKGGKIVNNWL
KQLIKVTLVFLFVAAIFYLITPVHVMSKHTDFSSEIIGYKAIDGGVTRDIASTDTCFA
NKHADFDTWFSQRGGSYTNDKACPLIAAVITREVGFVVPGLPGTILRTTNGDFLHFLP
RVFSAVGNICYTPSKLIEYTDFATSACVLAAECTIFKDASGKPVPYCYDTNVLEGSVA
YESLRPDTRYVLMDGSIIQFPNTYLEGSVRVVTTFDSEYCRHGTCERSEAGVCVSTSG
RWVLNNDYYRSLPGVFCGVDAVNLLTNMFTPLIQPIGALDISASIVAGGIVAIVVTCL
AYYFMRFRRAFGEYSHVVAFNTLLFLMSFTVLCLTPVYSFLPGVYSVIYLYLTFYLTN
DVSFLAHIQWMVMFTPLVPFWITIAYIICISTKHFYWFFSNYLKRRVVFNGVSFSTFE
EAALCTFLLNKEMYLKLRSDVLLPLTQYNRYLALYNKYKYFSGAMDTTSYREAACCHL
AKALNDFSNSGSDVLYQPPQTSITSAVLQSGFRKMAFPSGKVEGCMVQVTCGTTTLNG
LWLDDVVYCPRHVICTSEDMLNPNYEDLLIRKSNHNFLVQAGNVQLRVIGHSMQNCVL
KLKVDTANPKTPKYKFVRIQPGQTFSVLACYNGSPSGVYQCAMRPNFTIKGSFLNGSC
GSVGFNIDYDCVSFCYMHHMELPTGVHAGTDLEGNFYGPFVDRQTAQAAGTDTTITVN
VLAWLYAAVINGDRWFLNRFTTTLNDFNLVAMKYNYEPLTQDHVDILGPLSAQTGIAV
LDMCASLKELLQNGMNGRTILGSALLEDEFTPFDVVRQCSGVTFQSAVKRTIKGTHHW
LLLTILTSLLVLVQSTQWSLFFFXYENAFLPFAMGIIAMSAFAMMFVKHKHAFLCLFL
LPSLATVAYFNMVYMPASWVMRIMTWLDMVDTSLSGFKLKDCVMYASAVVLLILMTAR
TVYDDGARRVWTLMNVLTLVYKVYYGNALDQAISMWALIISVTSNYSGVVTTVMFLAR
GIVFMCVEYCPIFFITGNTLQCIMLVYCFLGYFCTCYFGLFCLLNRYFRLTLGVYDYL
VSTQEFRYMNSQGLLPPKNSIDAFKLNIKLLGVGGKPCIKVATVQSKMSDVKCTSVVL
LSVLQQLRVESSSKLWAQCVQLHNDILLAKDTTEAFEKMVSLLSVLLSMQGAVDINKL
CEEMLDNRATLQAIASEFSSLPSYAAFATAQEAYEQAVANGDSEVVLKKLKKSLNVAK
SEFDRDAAMQRKLEKMADQAMTQMYKQARSEDKRAKVTSAMQTMLFTMLRKLDNDALN
NIINNARDGCVPLNIIPLTTAAKLMVVIPDYNTYKNTCDGTTFTYASALWEIQQVVDA
DSKIVQLSEISMDNSPNLAWPLIVTALRANSAVKLQNNELSPVALRQMSCAAGTTQTA
CTDDNALAYYNTTKGGRFVLALLSDLQDLKWARFPKSDGTGTIYTELEPPCRFVTDTP
KGPKVKYLYFIKGLNNLNRGMVLGSLAATVRLQAGNATEVPANSTVLSFCAFAVDAAK
AYKDYLASGGQPITNCVKMLCTHTGTGQAITVTPEANMDQESFGGASCCLYCRCHIDH
PNPKGFCDLKGKYVQIPTTCANDPVGFTLKNTVCTVCGMWKGYGCSCDQLREPMLQSA
DAQSFLNRVCGVSAARLTPCGTGTSTDVVYRAFDIYNDKVAGFAKFLKTNCCRFQEKD
EDDNLIDSYFVVKRHTFSNYQHEETIYNLLKDCPAVVKHDFFKFRIDGDMVPHISRQR
LTKYTMADLVYALRHFDEGNCDTLKEILVTYNCCDDDYFNKKDWYDFVENPDILRVYA
NLGERVRQALLKTVQFCDAMRNAGIVGVLTLDNQDLNGNWYDFGDFIQTTPGSGVPVV
DSYYSLLMPILTLTRALTAESHVDTDLTKPYIKWDLLKYDFTEERLKLFDRYFKYWDQ
TYHPNCVNCLDDRCILHCANFNVLFSTVFPPTSFGPLVRKIFVDGVPFVVSTGYHFRE

LGVVHNQDVNLHSSRLSFKELLVYAADPAMHAASGNLLLDKRTTCFSVAALTNNVAFQ
TVKPGNFNKDFYDFAVSKGFFKEGSSVELKHFFFAQDGNAAISDYDYYRYNLPTMCDI
RQLLFVVEVVDKYFDCYDGGCINANQVIVNNLDKSAGFPFNKWGKARLYYDSMSYEDQ
DALFAYTKRNVIPTITQMNLKYAISAKNRARTVAGVSICSTMTNRQFHQKLLKSIAAT
RGATVVIGTSKFYGGWHNMLKxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
xxxxxxxxxxxxxxxxxxxxxxXEMVMCGGSLYVKPGGTSSGDATTAYANSVFNICQ
AVTANVNALLSTDGNKIADKYVRNLQHRLYECLYRNRDVDTDFVNEFYAYLRKHFSMM
ILSDDAVVCFNSTYASQGLVASIKNFKSVLYYQNNVFMSEAKCWTETDLTKGPHEFCS
QHTMLVKQGDDYVYLPYPDPSRILGAGCFVDDIVKTDGTLMIERFVSLAIDAYPLTKH
PNQEYADVFHLYLQYIRKLHDELTGHMLDMYSVMLTNDNTSRYWEPEFYEAMYTPHTV
LQAVGACVLCNSQTSLRCGACIRRPFLCCKCCYDHVISTSHKLVLSVNPYVCNAPGCD
VTDVTQLYLGGMSYYCKSHKPPISFPLCANGQVFGLYKNTCVGSDNVTDFNAIATCDW
TNAGDYILANTCTERLKLFAAETLKATEETFKLSYGIATVREVLSDRELHLSWEVGKP
RPPLNRNYVFTGYRVTKNSKVQIGEYTFEKGDYGDAVVYRGTTTYKLNVGDYFVLTSH
TVMPLSAPTLVPQEHYVRITGLYPTLNISDEFSSNVANYQKVGMQKYSTLQGPPGTGK
SHFAIGLALYYPSARIVYTACSHAAVDALCEKALKYLPIDKCSRIIPARARVECFDKF
KVNSTLEQYVFCTVNALPETTADIVVFDEISMATNYDLSVVNARLRAKHYVYIGDPAQ
LPAPRTLLTKGTLEPEYFNSVCRLMKTIGPDMFLGTCRRCPAEIVDTVSALVYDNKLK
AHKDKSAQCFKMFYKGVITHDVSSAINRPQIGVVREFLTRNPAWRKAVFISPYNSQNA
VASKILGLPTQTVDSSQGSEYDYVIFTQTTETAHSCNVNRFNVAITRAKVGILCIMSD
RDLYDKLQFTSLEIPRRNVATLQAENVTGLFKDCSKVITGLHPTQAPTHLSVDTKFKT
EGLCVDIPGIPKDMTYRRLISMMGFKMNYQVNGYPNMFITREEAIRHVRAWIGFDVEG
CHATREAVGTNLPLQLGFSTGVNLVAVPTGYVDTPNNTDFSRVSAKPPPGDQFKHLIP
LMYKGLPWNVVRIKIVQMLSDTLKNLSDRVVFVLWAHGFELTSMKYFVKIGPERTCCL
CDRRATCFSTASDTYACWHHSIGFDYVYNPFMIDVQQWGFTGNLQSNHDLYCQVHGNA
HVASCDAIMTRCLAVHECFVKRVDWTIEYPIIGDELKINAACRKVQHMVVKAALLADK
FPVLHDIGNPKAIKCVPQADVEWKFYDAQPCSDKAYKIEELFYSYATHSDKFTDGVCL
FWNCNVDRYPANSIVCRFDTRVLSNLNLPGCDGGSLYVNKHAFHTPAFDKSAFVNLKQ
LPFFYYSDSPCESHGKQVVSDIDYVPLKSATCITRCNLGGAVCRHHANEYRLYLDAYN
MMISAGFSLWVYKQFDTYNLWNTFTRLQSLENVAFNVVNKGHFDGQQGEVPVSIINNT
VYTKVDGVDVELFENKTTLPVNVAFELWAKRNIKPVPEVKILNNLGVDIAANTVIWDY
KRDAPAHISTIGVCSMTDIAKKPTETICAPLTVFFDGRVDGQVDLFRNARNGVLITEG
SVKGLQPSVGPKQASLNGVTLIGEAVKTQFNYYKKVDGVVQQLPETYFTQSRNLQEFK
PRSQMEIDFLELAMDEFIERYKLEGYAFEHIVYGDFSHSQLGGLHLLIGLAKRFKESP
FELEDFIPMDSTVKNYFITDAQTGSSKCVCSVIDLLLDDFVEIIKSQDLSVVSKVVKV
TIDYTEISFMLWCKDGHVETFYPKLQSSQAWQPGVAMPNLYKMQRMLLEKCDLQNYGD
SATLPKGIMMNVAKYTQLCQYLNTLTLAVPYNMRVIHFGAGSDKGVAPGTAVLRQWLP
TGTLLVDSDLNDFVSDADSTLIGDCATVHTANKWDLIISDMYDPKTKNVTKENDSKEG
FFTYICGFIQQKLALGGSVAIKITEHSWNADLYKLMGHFAWWTAFVTNVNASSSEAFL
IGCNYLGKPREQIDGYVMHANYIFWRNTNPIQLSSYSLFDMSKFPLKLRGTAVMSLKE
GQINDMILSLLSKGRLIIRENNRVVISSDVLVNN"

mat_peptide <1..>22

/gene="**ORF1ab**"

/product="**RNA-dependent RNA polymerase**"

ORIGIN

```
  1 gtgagatggt catgtgtggc gg
//
```

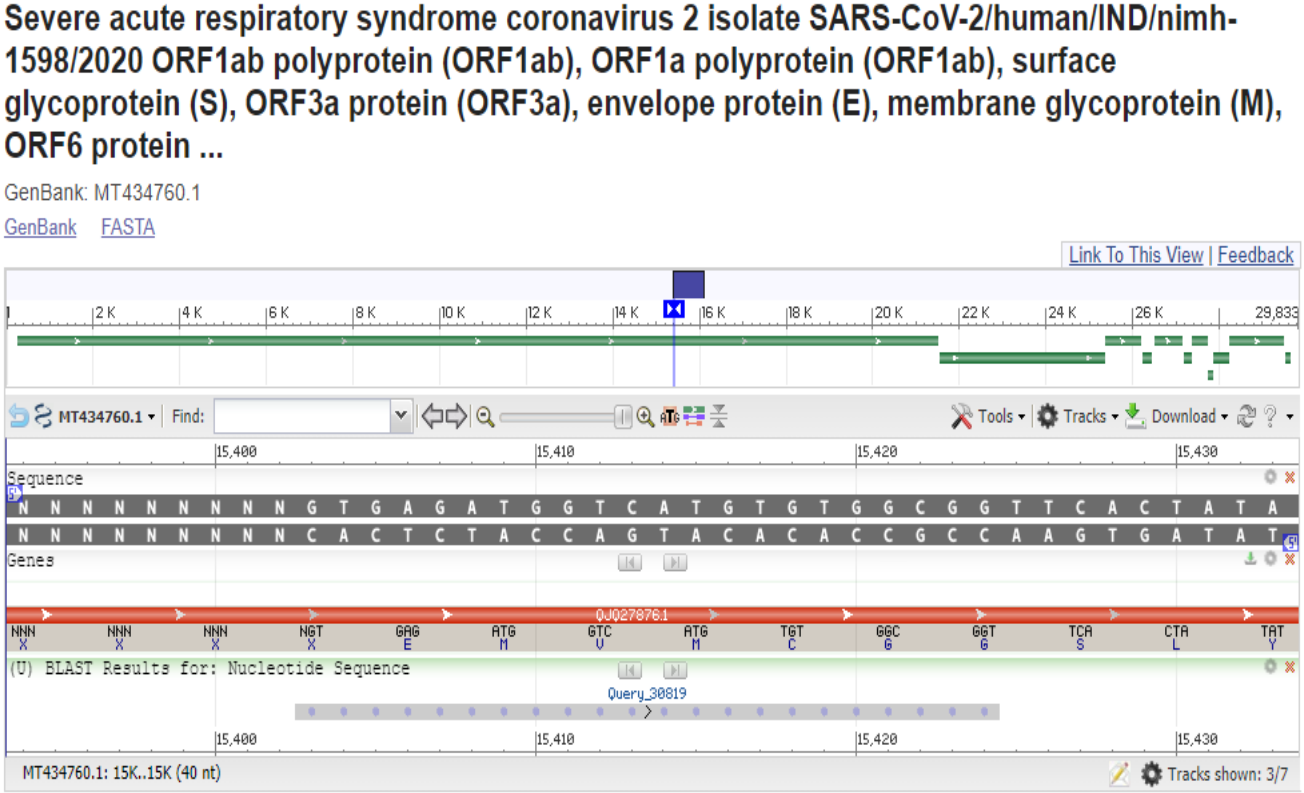
Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/IND/nimh-1598/2020 ORF1ab polyprotein (ORF1ab), ORF1a polyprotein (ORF1ab), surface glycoprotein (S), ORF3a protein (ORF3a), envelope protein (E), membrane glycoprotein (M), ORF6 protein (ORF6), ORF7a protein (ORF7a), ORF7b (ORF7b), ORF8 protein (ORF8), nucleocapsid phosphoprotein (N), and ORF10 protein (ORF10) genes, complete cds

Sequence ID: MT434760.1 Length: 29833 Number of Matches: 1

Range 1: 15403 to 15424 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
44.1 bits(22)	0.037	22/22(100%)	0/22(0%)	Plus/Plus

```
CDS: Putative 1          1          E  M  V  M  C  G  G
Query                    1      GTGAGATGGTCATGTGTGGCGG   22
                                |||||||||||||||||||||||
Sbjct                    15403  GTGAGATGGTCATGTGTGGCGG   15424
CDS:ORF1ab polyprote     5056   X  E  M  V  M  C  G  G
```



Mutant SARS coronavirus Urbani clone SARS-Urbani-MA_SHC014-spike, complete genome
GenBank: MT308984.1

FASTA Graphics

Go to:

LOCUS MT308984 22 bp RNA linear SYN **02-MAY-2020**

DEFINITION Mutant SARS coronavirus Urbani clone SARS-Urbani-MA_SHC014-spike, complete genome.

ACCESSION MT308984 REGION: 15361..15382

VERSION MT308984.1

KEYWORDS .

SOURCE SARS coronavirus Urbani

ORGANISM SARS coronavirus Urbani
Viruses; Riboviria; Orthornavirae; Pisuviricota; Pisoniviricetes; Nidovirales; Cornidovirineae; Coronaviridae; Orthocoronavirinae; Betacoronavirus; Sarbecovirus.

REFERENCE 1 (bases 1 to 22)

AUTHORS Menachery,V.D., Yount,B.L. Jr., Debbink,K., Agnihothram,S., Gralinski,L.E., Plante,J.A., Graham,R.L., Scobey,T., Ge,X.-Y., Donaldson,E.F., Randell,S.H., Lanzavecchia,A., Marasco,W.A., Shi,Z.-L. and Baric,R.S.

TITLE Direct Submission

JOURNAL **Submitted (03-APR-2020) Epidemiology, University of North Carolina at Chapel Hill, 2107 McGavran-Greenberg CB7435, Chapel Hill, NC 27599, USA**

COMMENT ##Assembly-Data-START##
Sequencing Technology :: Sanger dideoxy sequencing
##Assembly-Data-END##

FEATURES Location/Qualifiers

source 1..22
/organism="SARS coronavirus Urbani"
/mol type="genomic RNA"
/host="mouse"
/db_xref="taxon:228330"
/clone="SARS-Urbani-MA SHC014-spike"
/note="mouse-adapted; spike protein derived from bat coronavirus SHC014"

CDS <1..>22
/ribosomal slippage
/codon_start=3
/product="ORF1ab"
/protein_id="QJE50587.1"
/translation="MESLVLGVNEKTHVQLSLPVLQVRDVLVRGFGDSVEEALSEARE
HLKNGTCGLVELEKGVLPQLEQPYVFIKRSDALSTNHGHKVVELVAEMDGIQYGRSGI
TLGVLVPHVGETPIAYRNVLLRKNGNKGAGGHSYGIDLKSYDLGDELGTDPIEDYEQN
WNTKHGSGALRELTRELNGGAVTRYVDNNFCGPDGYPLDCIKDFLARAGKSMCTLSEQ

LDYIESKRGVYCCRDHEHEIAWFTERSDKSYEHQTPFEIKSAKKFDTFKGECPKFVFP
LNSKVKVIQPRVEKKKTEGFMGRIRSVYPVASPQECNNMHLSTLMKCNHCDEVSWQTC
DFLKATCEHCGTENLVIEGPTTCGYLPTNAVVKMPCPACQDPEIGPEHSVADYHNHSN
IETRLRKGGRTRCFGGCVFAYVGCYNKRAYWVPRASADIGSGHTGITGDNVETLNEDL
LEILSRERVNINIVGDFHLNEEVAIILASFSASTSAFIDTIKSLDYKSFKTIVESCGN
YKVTKGKPVKGAWNIGQQRSVLTPLCGFPSQAAGVIRSIFARTLDAANHSIPDLQRAA
VTILDGISEQSLRLVDAMVYTSDLLTNSVIIMAYVTGGLVQQTSQWLSNLLGTTVEKL
RPIFEWIEAKLSAGVEFLKDAWEILKFLITGVFDIVKGQIQVASDNIKDCVKCFIDVV
NKALEMCIDQVTIAGAKLRSLNLGEVFIAQSKGLYRQCIRGKEQLQLLMPLKAPKEVT
FLEGDSHDTVLTSEEVVLKNGELEALETPVDSFTNGAIVGTPVCVNGLMLLEIKDKEQ
YCALSPGLLATNNVFRLKGGAPIKGVTFGEDTVWEVQGYKNVRITFELDERVDKVLNE
KCSVYTVESGTEVTEFACVVAEAVVKTLQPVSDLLTNMGIDLDEWSVATFYLFDDAGE
ENFSSRMYCSFYPPDEEEEDDAECEEEEIDETCEHEYGTEDDYQGLPLEFGASAETVR
VEEEEEEDWLDDTTEQSEIEPEPEPTPEEPVNQFTGYLKLTDNVAIKCVDIVKEAQSA
NPMVIVNAANIHLKHGGGVAGALNKATNGAMQKESDDYIKLNGPLTVGGSCLLSGHNL
AKKCLHVVGPNLNAGEDIQLLKAAYENFNSQDILLAPLLSAGIFGAKPLQSLQVCVQT
VRTQVYIAVNDKALYEQVVMDYLDNLKPRVEAPKQEEPPNTEDSKTEEKSVVQKPVDV
KPKIKACIDEVTTTLEETKFLTNKLLLFADINGKLYHDSQNMLRGEDMSFLEKDAPYM
VGDVITSGDITCVVIPSKKAGGTTEMLSRALKKVPVDEYITTYPGQGCAGYTLEEAKT
ALKKCKSAFYVLPSEAPNAKEEILGTVSWNLREMLAHAEETRKLMPICMDVRAIMATI
QRKYKGIKIQEGIVDYGVRFFFYTSKEPVASIITKLNSLNEPLVTMPIGYVTHGFNLE
EAARCMRSLKAPAVVSVSSPDAVTTYNGYLTSSSKTSEEHFVETVSLAGSYRDWSYSG
QRTELGVEFLKRGDKIVYHTLESPVEFHLDGEVLSLDKLKSLLSLREVKTIKVFTTVD
NTNLHTQLVDMSMTYGQQFGPTYLDGADVTKIKPHVNHEGKTFFVLPSDDTLRSEAFE
YYHTLDESFLGRYMSALNHTKKWKFPQVGGLTSIKWADNNCYLSSVLLALQQLEVKFN
APALQEAYYRARAGDAANFCALILAYSNKTVGELGDVRETMTHLLQHANLESAKRVLN
VVCKHCGQKTTTLTGVEAVMYMGTLSYDNLKTGVSIPCVCGRDATQYLVQQESSFVMM
SAPPAEYKLQQGTFLCANEYTGNYQCGHYTHITAKETLYRIDGAHLTKMSEYKGPVTD
VFYKETSYTTTIKPVSYKLDGVTYTEIEPKLDGYYKKDNAYYTEQPIDLVPTQPLPNA
SFDNFKLTCSNTKFADDLNQMTGFTKPASRELSVTFFPDLNGDVVAIDYRHYSASFKK
GAKLLHKPIVWHINQATTKTTFKPNTWCLRCLWSTKPVDTSNSFEVLAVEDTQGMDNL
ACESQQPTSEEVVENPTIQKEVIECDVKTTEVVGNVILKPSDEGVKVTQELGHEDLMA
AYVENTSITIKKPNELSLALGLKTIATHGIAAINSVPWSKILAYVKPFLGQAAITTSN
CAKRLAQRVFNNYMPYVFTLLFQLCTFTKSTNSRIRASLPTTIAKNSVKSVAKLCLDA
GINYVKSPKFSKLFTIAMWLLLLSICLGSLICVTAAFGVLLSNFGAPSYCNGVRELYL
NSSNVTTMDFCEGSFPCSICLSGLDSLDSYPALETIQVTISSYKLDLTILGLAAEWVL
AYMLFTKFFYLLGLSAIMQVFFGYFASHFISNSWLMWFIISIVQMAPVSAMVRMYIFF
ASFYYIWKSYVHIMDGCTSSTCMMCYKRNRATRVECTTIVNGMKRSFYVYANGGRGFC
KTHNWNCLNCDTFCTGSTFISDEVARDLSLQFKRPINPTDQSSYIVDSVAVKNGALHL
YFDKAGQKTYERHPLSHFVNLDNLRANNTKGSLPINVIVFDGKSKCDESASKSASVYY
SQLMCQPILLLDQVLVSDVGDSTEVSVKMFDAYVDTFSATFSVPMEKLKALVATAHSE
LAKGVALDGVLSTFVSAARQGVVDTDVDTKDVIECLKLSHHSDLEVTGDSCNNFMLTY
NKVENMTPRDLGACIDCNARHINAQVAKSHNVSLIWNVKDYMSLSEQLRKQIRSAAKK
NNIPFRLTCATTRQVVNVITTKISLKGGKIVSTCFKLMLKATLLCVLAALVCYIVMPV
HTLSIHDGYTNEIIGYKAIQDGVTRDIISTDDCFANKHAGFDAWFSQRGGSYKNDKSC

PVVAAIITREIGFIVPGLPGTVLRAINGDFLHFLPRVFSAVGNICYTPSKLIEYSDFA
TSACVLAAECTIFKDAMGKPVPYCYDTNLLEGSISYSELRPDTRYVLMDGSIIQFPNT
YLEGSVRVVTTFDAEYCRHGTCERSEVGICLSTSGRWVLNNEHYRALSGVFCGVDAMN
LIANIFTPLVQPVGALDVSASVVAGGIIAILVTCAAYYFMKFRRVFGEYNHVVAANAL
LFLMSFTILCLVPAYSFLPGVYSVFYLYLTFYFTNDVSFLAHLQWFAMFSPIVPFWIT
AIYVFCISLKHCHWFFNNYLRKRVMFNGVTFSTFEEAALCTFLLNKEMYLKLRSETLL
PLTQYNRYLALYNKYKYFSGALDTTSYREAACCHLAKALNDFSNSGADVLYQPPQTSI
TSAVLQSGFRKMAFPSGKVEGCMVQVTCGTTTLNGLWLDDTVYCPRHVICTAEDMLNP
NYEDLLIRKSNHSFLVQAGNVQLRVIGHSMQNCLLRLKVDTSNPKTPKYKFVRIQPGQ
TFSVLACYNGSPSGVYQCAMRPNYTIKGSFLNGSCGSVGFNIDYDCVSFCYMHHMELP
TGVHAGTDLEGKFYGPFVDRQTAQAAGTDTTITLNVLAWLYAAVINGDRWFLNRFTTT
LNDFNLVAMKYNYEPLTQDHVDILGPLSAQTGIAVLDMCAALKALLQNGMNGRTILGS
TILEDEFTPFDVVRQCSGVTFQGKFKKIVKGTHHWMLLTFLTSLLILVQSTQWSLFFF
VYENAFLPFTLGIMAIAACAMLLVKHKHAFLCLFLLPSLATVAYFNMVYMPASWVMRI
MTWLELADTSLSGYRLKDCVMYASALVLLILMTARTVYDDAARRVWTLMNVITLVYKV
YYGNALDQAISMWALVISVTSNYSGVVTTIMFLARAIVFVCVEYYPLLFITGNTLQCI
MLVYCFLGYCCCCYFGLFCLLNRYFRLTLGVYDYLVSTQEFRYMNSQGLLPPKSSIDA
FKLNIKLLGIGGKPCIKVATVQSKMSDVKCTSVVLLSVLQQLRVESSSKLWAQCVQLH
NDILLAKDTTEAFEKMVSLLSVLLSMQGAVDINRLCEEMLDNRATLQAIASEFSSLPS
YAAYATAQEAYEQAVANGDSEVVLKKLKKSLNVAKSEFDRDAAMQRKLEKMADQAMTQ
MYKQARSEDKRAKVTSAMQTMLFTMLRKLDNDALNNIINNARDGCVPLNIIPLTTAAK
LMVVVPDYGTYKNTCDGNTFTYASALWEIQQVVDADSKIVQLSEINMDNSPNLAWPLI
VTALRANSAVKLQNNELSPVALRQMSCAAGTTQTACTDDNALAYYNNSKGGRFVLALL
SDHQDLKWARFPKSDGTGTIYAELEPPCRFVTDTPKGPKVKYLYFIKGLNNLNRGMVL
GSLAATVRLQAGNATEVPANSTVLSFCAFAVDPAKAYKDYLASGGQPITNCVKMLCTH
TGTGQAITVTPEANMDQESFGGASCCLYCRCHIDHPNPKGFCDLKGKYVQIPTTCAND
PVGFTLRNTVCTVCGMWKGYGCSCDQLREPLMQSADASTFLNRVCGVSAARLTPCGTG
TSTDVVYRAFDIYNEKVAGFAKFLKTNCCRFQEKDEEGNLLDSYFVVKRHTMSNYQHE
ETIYNLVKDCPAVAVHDFFKFRVDGDMVPHISRQRLTKYTMADLVYALRHFDEGNCDT
LKEILVTYNCCDDDYFNKKDWYDFVENPDILRVYANLGERVRQSLLKTVQFCDAMRDA
GIVGVLTLDNQDLNGNWYDFGDFVQVAPGCGVPIVDSYYSLLMPILTLTRALAAESHM
DADLAKPLIKWDLLKYDFTEERLCLFDRYFKYWDQTYHPNCINCLDDRCILHCANFNV
LFSTVFPPTSFGPLVRKIFVDGVPFVVSTGYHFRELGVVHNQDVNLHSSRLSFKELLV
YAADPAMHAASGNLLLDKRTTCFSVAALTNNVAFQTVKPGNFNKDFYDFAVSKGFFKE
GSSVELKHFFFAQDGNAAISDYDYYRYNLPTMCDIRQLLFVVEVVDKYFDCYDGGCIN
ANQVIVNNLDKSAGFPFNKWGKARLYYDSMSYEDQDALFAYTKRNVIPTITQMNLKYA
ISAKNRARTVAGVSICSTMTNRQFHQKLLKSIAATRGATVVIGTSKFYGGWHNMLKTV
YSDVETPHLMGWDYPKCDRAMPNMLRIMASLVLARKHNTCCNLSHRFYRLANECAQVL
SSLYVKPGGTSSGDATTAYANSVFNICQAVTANVNALLSTDGNKIADKYVR
SEMVMCGG
NLQHRLYECLYRNRDVDHEFVDEFYAYLRKHFSMMILSDDAVVCYNSNYAAQGLVASI
KNFKAVLYYQNNVFMSEAKCWTETDLTKGPHEFCSQHTMLVKQGDDYVYLPYPDPSRI
LGAGCFVDDIVKTDGTLMIERFVSLAIDAYPLTKHPNQEYADVFHLYLQYIRKLHDEL
TGHMLDMYSVMLTNDNTSRYWEPEFYEAMYTPHTVLQAVGVCVLCNSQTSLRCGACIR
RPFLCCKCCYDHVISTSHKLVLSVNPYVCNAPGCDVTDVTQLYLGGMSYYCKSHKPPI
SFPLCANGQVFGLYKNTCVGSDNVTDFNAIATCDWTNAGDYILANTCTERLKLFAAET

LKATEETFKLSYGIATVREVLSDRELHLSWEVGKPRPPLNRNYVFTGYRVTKNSKVQI
GEYTFEKGDYGDAVVYRGTTTYKLNVGDYFVLTSHTVMPLSAPTLVPQEHYVRITGLY
PTLNISDEFSSNVANYQKVGMQKYSTLQGPPGTGKSHFAIGLALYYPSARIVYTACSH
AAVDALCEKALKYLPIDKCSRIIPARARVECFDKFKVNSTLEQYVFCTVNALPETTAD
IVVFDEISMATNYDLSVVNARLRAKHYVYIGDPAQLPAPRTLLTKGTLEPEYFNSVCR
LMKTIGPDMFLGTCRRCPAEIVDTVSALVYDNKLKAHKDKSAQCFKMFYKGVITHDVS
SAINRPQIGVVREFLTRNPAWRKAVFISPYNSQNAVASKILGLPTQTVDSSQGSEYDY
VIFTQTTETAHSCNVNRFNVAITRAKIGILCIMSDRDLYDKLQFTSLEIPRRNVATLQ
AENVTGLFKDCSKIITGLHPTQAPTHLSVDIKFKTEGLCVDIPGIPKDMTYRRLISMM
GFKMNYQVNGYPNMFITREEAIRHVRAWIGFDVEGCHATRDAVGTNLPLQLGFSTGVN
LVAVPTGYVDTENNTEFTRVNAKPPPGDQFKHLIPLMYKGLPWNVVRIKIVQMLSDTL
KGLSDRVVFVLWAHGFELTSMKYFVKIGPERTCCLCDKRATCFSTSSDTYACWNHSVG
FDYVYNPFMIDVQQWGFTGNLQSNHDQHCQVHGNAHVASCDAIMTRCLAVHECFVKRV
DWSVEYPIIGDELRVNSACRKVQHMVVKSALLADKFPVLHDIGNPKAIKCVPQAEVEW
KFYDAQPCSDKAYKIEELFYSYATHHDKFTDGVCLFWNCNVDRYPANAIVCRFDTRVL
SNLNLPGCDGGSLYVNKHAFHTPAFDKSAFTNLKQLPFFYYSDSPCESHGKQVVSDID
YVPLKSATCITRCNLGGAVCRHHANEYRQYLDAYNMMISAGFSLWIYKQFDTYNLWNT
FTRLQSLENVAYNVVNKGHFDGHAGEAPVSIINNAVYTKVDGIDVEIFENKTTLPVNV
AFELWAKRNIKPVPEIKILNNLGVDIAANTVIWDYKREAPAHVSTIGVCTMTDIAKKP
TESACSSLTVLFDGRVEGQVDLFRNARNGVLITEGSVKGLTPSKGPAQASVNGVTLIG
ESVKTQFNYFKKVDGIIQQLPETYFTQSRDLEDFKPRSQMETDFLELAMDEFIQRYKL
EGYAFEHIVYGDFSHGQLGGLHLMIGLAKRSQDSPLKLEDFIPMDSTVKNYFITDAQT
GSSKCVCSVIDLLLDDFVEIIKSQDLSVISKVVKVTIDYAEISFMLWCKDGHVETFYP
KLQASQAWQPGVAMPNLYKMQRMLLEKCDLQNYGENAVIPKGIMMNVAKYTQLCQYLN
TLTLAVPYNMRVIHFGAGSDKGVAPGTAVLRQWLPTGTLLVDSDLNDFVSDADSTLIG
DCATVHTANKWDLIISDMYDPRTKHVTKENDSKEGFFTYLCGFIKQKLALGGSIAVKI
TEHSWNADLYKLMGHFSWWTAFVTNVNASSSEAFLIGANYLGKPKEQIDGYTMHANYI
FWRNTNPIQLSSYSLFDMSKFPLKLRGTAVMSLKENQINDMIYSLLEKGRLIIRENNR
VVVSSDILVNN"

ORIGIN

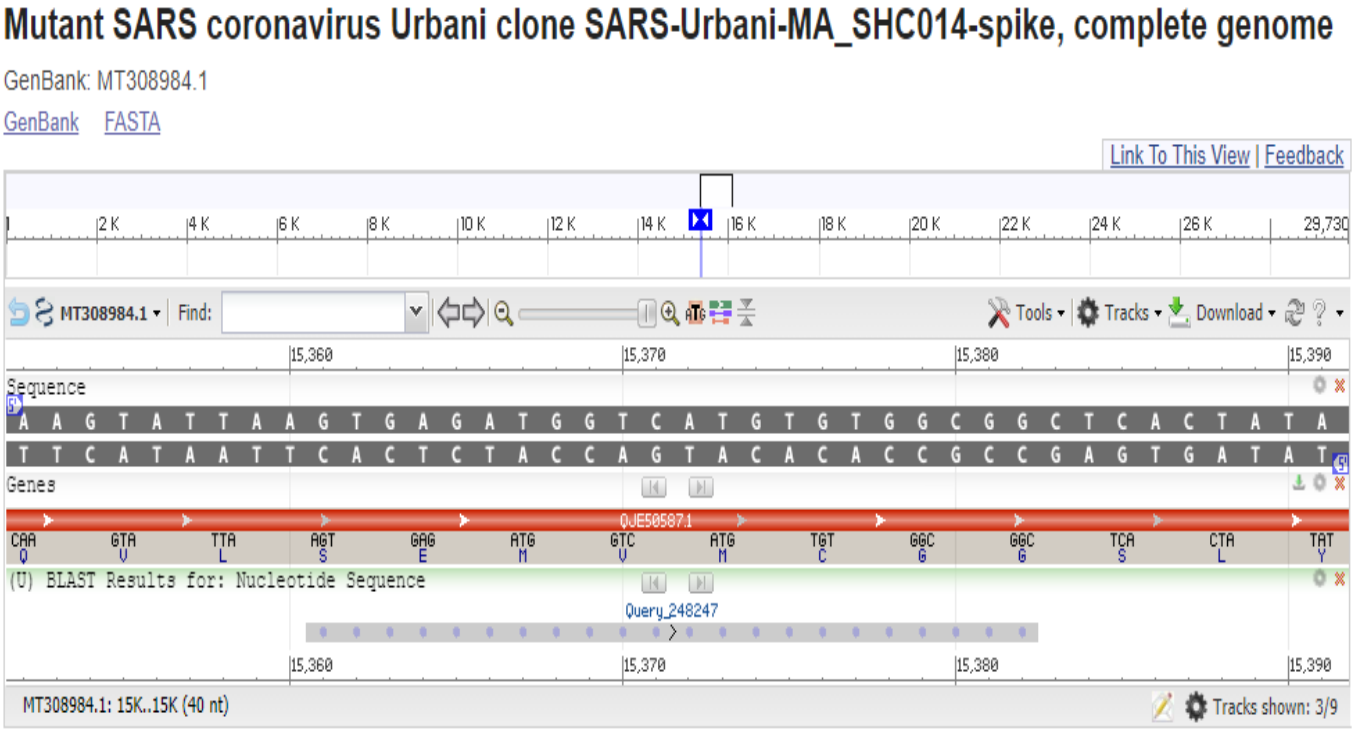
1 gtgagatggt catgtgtggc gg

//

Mutant SARS coronavirus Urbani clone SARS-Urbani-MA_SHC014-spike, complete genome
Sequence ID: MT308984.1 Length: 29730 Number of Matches: 1
Range 1: 15361 to 15382 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
44.1 bits(22)	0.038	22/22(100%)	0/22(0%)	Plus/Plus

```
CDS: Putative 1        1          E   M   V   M   C   G   G
Query                  1      GTGAGATGGTCATGTGTGGCGG   22
                              ||||||||||||||||||||||
Sbjct                  15361  GTGAGATGGTCATGTGTGGCGG   15382
CDS:ORF1ab [SARS cor   5033   S   E   M   V   M   C   G   G
```



Bat SARS-like coronavirus isolate bat-SL-CoVZC45, complete genome

GenBank: MG772933.1

FASTA Graphics

Go to:

```
LOCUS       MG772933                 22 bp    RNA     linear   VRL 29-SEP-2020
DEFINITION  Bat SARS-like coronavirus isolate bat-SL-CoVZC45, complete genome.
ACCESSION   MG772933 REGION: 15418..15439
VERSION     MG772933.1
KEYWORDS    .
SOURCE      Bat SARS-like coronavirus
  ORGANISM  Bat SARS-like coronavirus
            Viruses; Riboviria; Orthornavirae; Pisuviricota; Pisoniviricetes;
            Nidovirales; Cornidovirineae; Coronaviridae; Orthocoronavirinae;
            Betacoronavirus; Sarbecovirus.
REFERENCE   1  (bases 1 to 22)
  AUTHORS   Hu,D., Zhu,C., Ai,L., He,T., Wang,Y., Ye,F., Yang,L., Ding,C.,
            Zhu,X., Lv,R., Zhu,J., Hassan,B., Feng,Y., Tan,W. and Wang,C.
  TITLE     Genomic characterization and infectivity of a novel SARS-like
            coronavirus in Chinese bats
  JOURNAL   Emerg Microbes Infect 7 (1), 154 (2018)
   PUBMED   30209269
  REMARK    Publication Status: Online-Only
REFERENCE   2  (bases 1 to 22)
  AUTHORS   Hu,D.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-JAN-2018) Institute of Military Medicine Nanjing
            Command, Nanjing, Institute of Military Medicine Nanjing Command,
            Nanjing, NO. 293 East Zhongshan Road, Nanjing, JangSu 210002, China
COMMENT     ##Assembly-Data-START##
            Sequencing Technology :: Sanger dideoxy sequencing
            ##Assembly-Data-END##
FEATURES             Location/Qualifiers
     source          1..22
                     /organism="Bat SARS-like coronavirus"
                     /mol_type="genomic RNA"
                     /isolate="bat-SL-CoVZC45"
                     /host="Rhinolophus pusillus"
                     /db_xref="taxon:1508227"
                     /country="China"
                     /collection_date="Feb-2017"
     CDS             <1..>22
                     /ribosomal_slippage
                     /codon_start=3
                     /product="non-structural polyprotein 1ab"
                     /protein_id="AVP78030.1"
```

/translation="MESLVPGFNEKTHGPLSLPVLQVRDVLVRGFGDSVEEALSEARQ
HLKDGTCGLVEVEKGVLPQLEQPYVFIKRSDARTAPHGHVMVELVAELDGIQYGRSGE
TLGVLVPHVGEVPVAYRKVLLRKNGNKGAGGHSYGADLKSFDLGDELGTDPIEDFQEN
WNTKHGSGVTRELKRELNGGAYTRYVDNNFCGPDGYPLECIKDLLARAGKASYALSEQ
LDFIDTKRGVYCCREHEHEIAWYTERSEKSYELQTPFEIKLAKKFDTFNGECPNFVFP
LNSTIKTIQPRVEKKKLDGFMGRIRSVYPVASPNECNQMCLSILMKCDHCGETSWQTG
DFVRATCEFCGTENLTKEGATTCGYLPQNAVVKLYCPACHNPEVGPEHSLAEYHNESG
LKTVLRKGGRTIAYGGCVFAYVGCYNKCAYWVPRASANIGCNHTGVVGEGSESLNDNL
LEILQKEKVNINIVGDFKLNEEIAIILASFSASISAFVETVKGLDYKTFKQIVESCGN
FKVTKGKAKKGAWNIGEQSSILSPLYAFPSDAAHVVRSIFSRTLETAHHSVHVLQKAA
IIILDGISQYSLRLIDAMMFTSDLVTNNLVVMAYITGGVVQMTSQWLTNIFGTVYEKL
KPVLDWLEEKFKEGIEFLRDGWEIVKFISTCACEIVGGQIVTCAKEIKESVQTFFKLV
NKFLALCADSIIIGGAKLKALNLGETFVTHSKGLYRKCVRSREETGLLLPLKAPKEII
FLEGETLPTEVLTEEVVLKTGVLQPLEQPTNEAVEAPLIGTPVCINGLMLLEIKDTEK
YCALAPNMMVTNNTFTLKGGAPTKVTFGDDTVIEVQGYKSVNITFELDERIDKVLNEK
CSNYTVELGTEVNEFACVVADAVIKTLQPVSELIIPLGIDLDEWSMATYYLFDESGEF
KLSSHMYCSFYPPEDEGEDDCEEGQCEPSTQYEYGTEDDYQGKPLEFGATSFSSSSQE
EEQEEDWLESDSQDGQETAVEENKIPSVEVPPVLQVESTPVVTETSEQNNFTGYLKLT
DNVFIKNADIVEEAKKVKPTVVVNAANVYLKHGGGVAGALNKATNNAMQVESDKYITT
NGPLIVGGGCVLSGHNLAKNCLHVVGPNVNRGEDIQLLKNAYENFNQHEILLAPLLSA
GIFGADPVHSLRVCVETVHTNVYLVVFDKNLYDKLVSSFLEMKSEKQVEQKVAENPKE
EVKPFFTENKPSVEQRQQAEEKKIKASIEEVTTTLEETKFLTENLLLYIDINGNLHPD
SATLVKDIDTTFLKKDVPYIVGDVIKEGALTAVVIPTKKAGGTTEMLAKALRKVPTDN
YITTYPGQGLNGYTVEEAKTVLKKCKSAFYILPSIISNAKQEILGTVSWNLREMLAHA
EETRKLMPVCMETKAIVSTIQRKYKGIKIQEGVVDYGARFYFYTSKTTVASLINTLNN
LNETLVTMPLGYVTHGLNLEEAARYMRSLKVPATVSVSSPDAVTAYNGYLTSSSKTPE
EHFIETVSLAGSYKDWSYSGQSTQLGIEFLKRGDKSVYYTSNPTTFHLDGETITFDNL
KTLLSLREVRNIKVFTTVDNINLHTQVVDMSMTYGQQFGPIYLDGADVTKIKPHNSHE
GKTFYVLPNDDTLRAEAFEYYHTTDPSFLGRYMSALNHTKKWKYPQVNGLTSIKWADN
NCYLATALLTIQQIELKFNPPALQDAYYRARAGEAANFCALILAYCNKTVGELGDVRE
TMNYLFQHANLDSCKRVLNVVCKTCGQQQTTLKGVEAVMYMGTLSYEQLKKGVQIPCM
CGKQATQYLVQQESPFVMMSAPPAQYELKHGTFVCASEYTGNYQCGHYKHITSKETLY
CIDGALLTKSSEYKGSITDVFYKENSYTTTIKPVTYKLDGVVCTEIDPKLDGYYKKDN
SYFTEQPIDLVPNQPYPNASFDNFKFVCDNTKFADDLNQLSGYKKPASRELKVTFFPD
LNGDVVAIDYKHYTPSFKKGAKLLHKPIVWHVNNATNKATYKPNIWCIRCLWSTKPVE
TSNSFDALELGDTQGMDNLACEVLKPVSEEVVENPTIQKDILECNVKTTEVVGDIILK
PASDGLKITKEVGHTDLMAAYVDNSSLTIKKPNELSRVLGLKTLATHGLAAINSVPWD
TIANYVKPFLNKVVSTTTNIVTRCLNRVCTNYMPYLFTLLLQLCTFTKSTNSRIRASM
PTTIAKNTVKSVGKFCIEASFNYLKSPNFSKLINIVIWFLLLSVCLGSLIYSTAALGV
LMSNLGMPSYCTVYREGYLNSTNVTTATYCTGSIPCSVCLSGLDSLDTYPSLETIQIT
ISSFKWDLTAFGLVAEWFLAYILFTRFFYVLGLAAIMQLFFSYFAVHFISNSWLMWLI
INLVQMAPISAMVRMYIFFASFYYVWKSYVHVVDGCTSSTCMMCYKRNRATRVECTTI
VNGVRRSFYVYANGGKGFCKLHNWNCINCDTFCAGSTFISDEVARDLSLQFKRPINPT
DQSSYIVDSVTVKNGSIHLYFDKAGQKTYERHSLSHFVNLDNLRANNTKGSLPINVIV
FDGKSKCEESSAKSASVYYSQLMCQPILLLDQALVSDVGDSAEVAVKMFDAYVNIFSS

TFNVPMEKLKALVATAEAELAKNVSLDNVLSTFISAARQGFVDSDVETKDVVECLKLS
HQSDIEVTGDSCNNYMLTYNKVENMTPRDLGACIDCSARHINAQVAKSHNIALIWNVK
DFMSLSEQLRKQIRSAAKKNNLPFRLTCATTRQVVNVVTTKIALKGGKIVNNWLKQLI
KVTLVFLFITVIFYLITPVHVMFKHNDFSSEIIGYKAIDGGVTRDIASTDTCFANKHA
DFDSWFSQRGGSYTNDKACPLVAAVITREVGFVVPGLPGTILRTINGDFLHFLPRVFS
AVGNICYTPSKLIEYTDFATSACVLAAECTIFKDASGKPVPYCYDTNVLEGSVAYESL
RPDTRYVLMDGSIIQFPNTYLEGSVRVVTTFDSEYCRHGTCERSEAGICVSTSGRWVL
NNDYYRSLPGVFCGVDAVNLLTNMFTPLIQPIGALDISASIVAGGVVAIIVTCLAYYF
MRFRRAFGEYSHVVAFNTLLFFMSFTVLCLTPVYSFLPGVYSVIYLYLTFYLTNDVSF
LAHIQWMVMFTPLVPFWMTIVYVICISTKHFYWFFSNYLKRRVVFNGVSFSTFEEAAL
CTFLLNKEMYLKLRSDVLLPLTQYNRYLALYNKYKYFSGAMDTTSYREAACCHLAKAI
NDFSNSGSDVLYQPPQTSITSAVLQSGFRKMAFPSGKVEGCMVQVTCGTTTLNGLWLD
DVVYCPRHVICTSEDMLNPNYEDLLIRKSNHNFLVQAGNVQLRVVGHSMQNCVLKLKV
DTANPKTPKYKFVRIQPGQTFSVLACYNGSPSGVYQCAMRPNFTIKGSFLNGSCGSVG
FNIDYDCVSFCYMHHMELPTGVHAGTDLEGTFYGPFVDRQTAQAAGTDTTITVNVLAW
LYAAVINGDRWFLNRFTTTLNDFNLVAMKYNYEPLTQDHLDILGPLSAQTGIAVLDMC
ASLKELLQNGMNGRTILGSALLEDEFTPFDVVRQCSGVTFQSAVKRTIKGTHHWLLLT
VLTSLLVLVQSTQWSLFFFVYENAFMPFAMGIIAMSAFAMMFVKHKHAFLCLFLLPSL
ATVAYFNMVYMPASWVMRIMTWLDIVDTSLSGFKLKDCVMYASAVVLLILMTARTVYD
DGARRVWTLMNVLTLVYKVYYGNALDQAISMWALIISVTSNYSGVVTTVMFLARGIVF
MCVEYCPIFFITGNTLQCIMLVYCFLGYFCTCYFGLFCLLNRYFRLTLGVYDYLVSTQ
EFRYMNSQGLLPPKSSIDAFKLNIKLLGIGGKPCIKVATVQSKMSDVKCTSVVLLSVL
QQLRVESSSKLWAQCVQLHNDILLAKDTTEAFEKMVSLLSVLLSMQGAVDINKLCEEM
LDNRATLQAIASEFSSLPSYAAYATAQEAYEQAVANGDSEVVLKKLKKSLNVAKSEFD
RDAAMQRKLEKMADQAMTQMYKQARSEDKRAKVTSAMQTMLFTMLRKLDNDALNNIIN
NARDGCVPLNIIPLTTAAKLMVVVPDYGTYKNTCDGNTFTYASALWEIQQVVDADSKI
VQLSEINMDNSPNLAWPLIVTALRANSAVKLQNNELSPVALRQMSCAAGTTQTACTDD
NALAYYNNSKGGRFVLALLSDHQDLKWARFPKSDGTGTIYTELEPPCRFVTDTPKGPK
VKYLYFIKGLNNLNRGMVLGSLAATVRLQAGNATEVPANSTVLSFCAFAVDPAKAYKD
YLASGGQPITNCVKMLCTHTGTGQAITVTPEANMDQESFGGASCCLYCRCHIDHPNPK
GFCDLKGKYVQIPTTCVNDPVGFTLRNTVCTVCGMWKGYGCSCDQLREPMMQSADAST
FLNRVCGVSAARLTPCGTGTSTDVVYRAFDIYNEKVAGFAKFLKTNCCRFQEKDEEGN
LLDSYFVVKRHTMSNYQHEETIYNLVKGCPAVAVHDFFKFRVDGDMVPHISRQRLTKY
TMADLVYALRHFDEGNCDTLKEILVTYKCCDDNYFNKKDWYDFVENPDVLRVYANLGE
RVRQALLKTVQFCDAMRDAGIVGVLTLDNQDLNGNWYDFGDFVQVAPGCGVPIVDSYY
SLLMPILTLTRALAAESHMDADLAKPLIKWDLLKYDFTEERLCLFDRYFKYWDQTYHP
NCINCLDDRCILHCANFNVLFSTVFPPTSFGPLVRKIFVDGVPFVVSTGYHFRELGVV
HNQDVNLHNSRLSFKELLVYAADPAMHAASGNLLLDKRTTCFSVAALTNNVAFQTVKP
GNFNKDFYDFAVSKGFFKEGSSVELKHFFFAQDGNAAISDYDYYRYNLPTMCDIRQLL
FVVEVVDKYFDCYDGGCINANQVIVNNLDKSAGFPFNKWGKARLYYDSMSYEDQDALF
AYTKRNVIPTITQMNLKYAISAKNRARTVAGVSICSTMTNRQFHQKLLKSIAATRGAT
VVIGTSKFYGGWHNMLKTVYSDVESPHLMGWDYPKCDRAMPNMLRIMASLVLARKHST
CCNLSHRFYRLANECAQVL**S**EMVMCGGSLYVKPGGTSSGDATTAYANSVFNICQAVTA
NVNALLSTDGNKIADKYVRNLQHRLYECLYRNRDVDQEFVDEFYAYLRKHFSMMILSD
DAVVCYNSNYAAQGLVASIKNFKAVLYYQNNVFMSEAKCWTETDLTKGPHEFCSQHTM

LVKQGDDYVYLPYPDPSRILGAGCFVDDIVKTDGTLMIERFVSLAIDAYPLTKHPNQE
YADVFHLYLQYIRKLHDELTGHMLDMYSVMLTNDNTSRYWEPEFYEAMYTPHTVLQAV
GACVLCNSQTSLRCGACIRRPFLCCKCCYDHVISTSHKLVLSVNPYVCNAPGCDVTDV
TQLYLGGMSYYCKSHKPPISFPLCANGQVFGLYKNTCVGSDNVTDFNAIATCDWTNAG
DYILANTCTERLKLFAAETLKATEETFKLSYGIATVREVLSDRELHLSWEVGKPRPPL
NRNYVFTGYRVTKNSKVQIGEYTFEKGDYGDAVVYRGTTTYKLNVGDYFVLTSHTVMP
LSAPTLVPQEHYVRITGLYPTLNISDEFSSNVANYQKVGMQKYSTLQGPPGTGKSHFA
IGLALYYPSARIVYTACSHAAVDALCEKALKYLPIDKCSRIIPARARVECFDKFKVNS
TLEQYVFCTVNALPETTADIVVFDEISMATNYDLSVVNARLRAKHYVYIGDPAQLPAP
RTLLTKGTLEPEYFNSVCRLMKTIGPDMFLGTCRRCPAEIVDTVSALVYDNKLKAHKE
KSAQCFKMFYKGVITHDVSSAINRPQIGVVREFLTRNPAWRKAVFISPYNSQNAVASK
ILGLPTQTVDSSQGSEYDYVIFTQTTETAHSCNVNRFNVAITRAKIGILCIMSDRDLY
DKLQFMSLEVPRRNVATLQAENVTGLFKDCSKIITGLHPTQAPTHLSVDTKFKTEGLC
VDIPGIPKDMTYRRLISMMGFKMNYQVNGYPNMFITREEAIRHVRAWIGFDVEGCHAT
RDAVGTNLPLQLGFSTGVNLVAVPTGYVDTEHSTEFTRVNAKPPPGDQFKHLIPLMYK
GLPWNVVRIKIVQMLSDTLKGLSDRVVFVLWAHGFELTSMKYFVKIGPERTCCLCDKR
ATCFSTSSDTYACWNHSVGFDYVYNPFMIDVQQWGFTGNLQSNHDQHCQVHGNAHVAS
CDAIMTRCLAVHECFVKRVDWSVEYPIIGDELKINAACRKVQHMVVKSALLADKFPVL
HDIGNPKAIRCVPQSEVDWKFYDAQPCSDKAYKIEELFYSYATHHDKFTDGVCLFWNC
NVDRYPANAIVCRFDTRVLSNLNLPGCDGGSLYVNKHAFHTPAFDKSAFTHLKQLPFF
YYSDSPCESHGKQVVSDIDYVPLKSATCITRCNLGGAVCRHHANEYRQYLDAYNMMIS
AGFSLWIYKQFDTYNLWNTFTKLQSLENVAYNVINKGHFDGQNGEAPVSIVNNAVYTK
LDGVDVEIFENKTTLPVNVAFELWAKRNIKPVPEIKILNNLGVDIAANTVIWDYKREA
PAHVSTIGVCTMTDIAKKPTESACSSLTVLFDGRVEGQVDLFRNARNGVLITEGSVKG
LTPSKGPAQASVNGVTLIGESVKTQFNYFKKVDGIIQQLPETYFTQSRDLEDFKPRSQ
METDFLELAMDEFIERYKLDGYAFEHIVYGDFSHGQLGGLHLLIGLAKRSQDSLLKLE
DFIPMDSTVKNYFITDAQTGSSKCVCSVIDLLLDDFVEIIKSQDLSVVSKVVKVTIDY
TEISFMLWCKDGHVETFYPKLQSSQAWQPGVAMPNLYKMQRMLLEKCDLQNYGDSAIL
PKGIMMNVAKYTQLCQYLNTLTLAVPYNMRVIHFGAGSDKGVAPGTAVLRQWLPTGTL
LVDSDLNDFVSDADSTLIGHCATVHTANKWDLIISDMYDPKTKNVTKENDSKEGFFTY
ICGFIQQKLALGGSVAVKITEHSWNADLYKLMGHFAWWTAFVTNVNASSSEAFLIGCN
YLGKPREQIDGYVMHANYIFWRNTNPIQLSSYSLFDMSKFPLKLRGTAVMSLKDGQIN
DMILSLLSKGRLIVRENNRVIISSDVLVNN"

ORIGIN

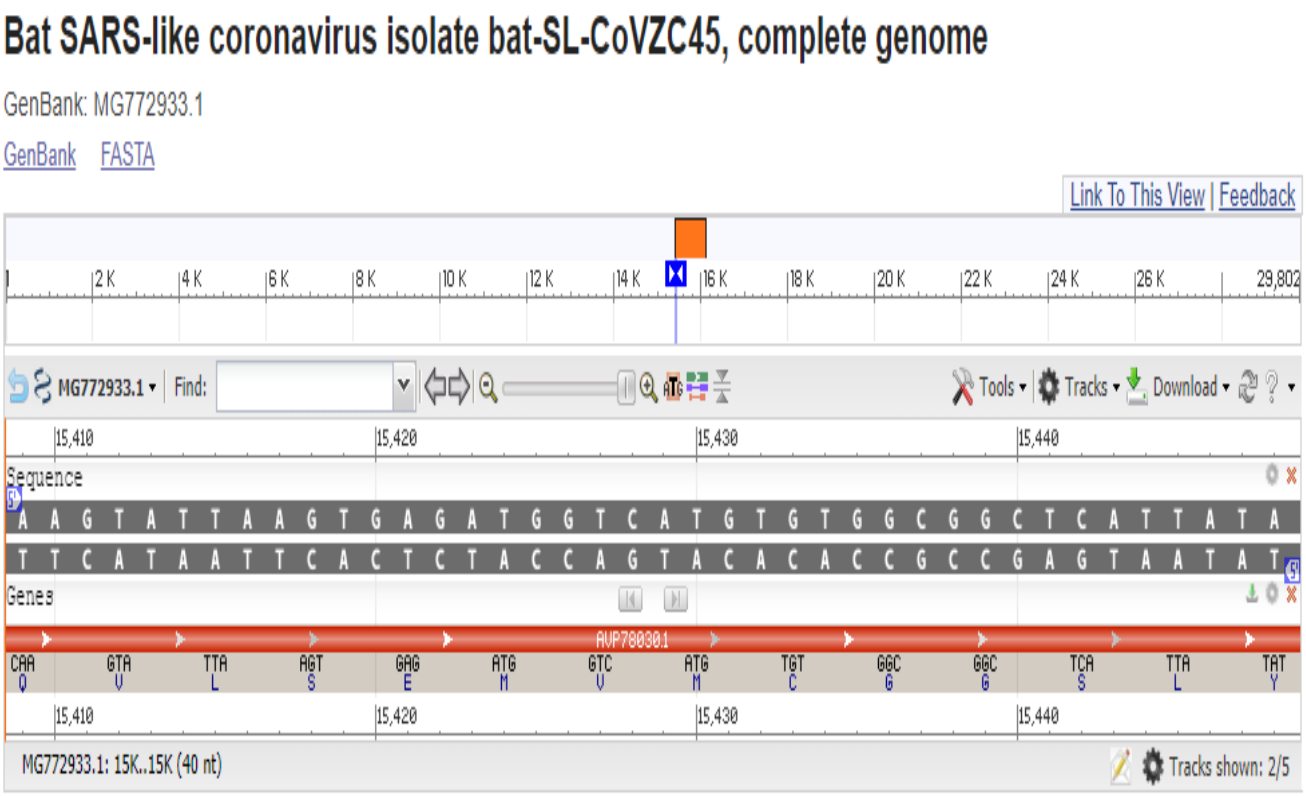
1 gtgagatggt catgtgtggc gg

//

Bat SARS-like coronavirus isolate bat-SL-CoVZC45, complete genome
Sequence ID: MG772933.1 **Length:** 29802 **Number of Matches:** 1
Range 1: 15418 to 15439 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
44.1 bits(22)	0.036	22/22(100%)	0/22(0%)	Plus/Plus

```
CDS: Putative 1          1          E   M   V   M   C   G   G
Query                    1      GTGAGATGGTCATGTGTGGCGG   22
                                |||||||||||||||||||||||
Sbjct                    15418  GTGAGATGGTCATGTGTGGCGG   15439
CDS:non-structural p     5052   S   E   M   V   M   C   G   G
```



Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/CA-CZB-11679/2020 ORF1ab polyprotein (ORF1ab), ORF1a polyprotein (ORF1ab), surface glycoprotein (S), ORF3a protein (ORF3a), envelope protein (E), membrane glycoprotein (M), ORF6 prote...

GenBank: MW276628.1

FASTA Graphics

Go to:

LOCUS MW276628 22 bp RNA linear VRL **20-NOV-2020**

DEFINITION Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/CA-CZB-11679/2020 ORF1ab polyprotein (ORF1ab), ORF1a polyprotein (ORF1ab), surface glycoprotein (S), ORF3a protein (ORF3a), envelope protein (E), membrane glycoprotein (M), ORF6 protein (ORF6), ORF7a protein (ORF7a), ORF7b (ORF7b), ORF8 protein (ORF8), nucleocapsid phosphoprotein (N), and ORF10 protein (ORF10) genes, complete cds.

ACCESSION MW276628 REGION: 15377..15398

VERSION MW276628.1

DBLINK BioSample: SAMN16690370

KEYWORDS .

SOURCE Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)

ORGANISM Severe acute respiratory syndrome coronavirus 2
Viruses; Riboviria; Orthornavirae; Pisuviricota; Pisoniviricetes; Nidovirales; Cornidovirineae; Coronaviridae; Orthocoronavirinae; Betacoronavirus; Sarbecovirus.

REFERENCE 1 (bases 1 to 22)

AUTHORS Bhatt,K.D.

TITLE Direct Submission

JOURNAL **Submitted (19-NOV-2020) Infectious diseases, Chan Zuckerberg Biohub, 499 Illinois st., San Francisco, CA 94158, USA**

COMMENT ##Assembly-Data-START##
Assembly Method :: Minimap v. 2.17; ivar v. 1.2; samtools v. 1.9
Sequencing Technology :: Illumina
##Assembly-Data-END##

FEATURES Location/Qualifiers

source 1..22
/organism="Severe acute respiratory syndrome coronavirus 2"
/mol type="genomic RNA"
/isolate="SARS-CoV-2/human/USA/CA-CZB-11679/2020"
/isolation source="Nasopharyngeal/oropharyngeal swab"
/host="**Homo sapiens**"
/db_xref="taxon:2697049"
/country="USA: CA, Santa Clara County"
/collection_date="2020-09-24"

gene <1..>22

/gene="ORF1ab"

CDS <1..>22

/gene="ORF1ab"

/ribosomal slippage

/codon_start=3

/product="ORF1ab polyprotein"

/protein_id="QPF55728.1"

/translation="MESLVPGFNEKTHVQLSLPVLQVRDVLVRGFGDSVEXXXSXXRQ
XLKDGTCGLVEVEKGVLPQLEQPYVXIKRSDARTAPXGHVMVXXXXXXXXXXXXXXXX
XXXXXXPHVGEIXXXXXXXXXXXXXXXXXXXHSYGADLKSFDLGDELGTDPYEDFQEN
WNTKHSSGVTRELMRELNGGAYTRYVDNNFCGPDGYPLECIKDLLARAGKASCTLSEQ
LDFIDTKRGVYCCREHEHEIAWYTERSEKSYELQTPFEIKLAKKFDTFNGECPNFVFP
LNSIIKTIQPRVEKKKLDGFMGRIRSVYPVASPNECNQMCLSTLMKXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXTCGYLPXNAVVKIYCPACHNSEVGPEHSLAEYHNESG
LKTILRKGGRTIAFGGCVFSYVGCHNKCAYWVPRASANIGCNHTGVVGEGXEGLNDNL
LEILQKEKVNINIVGDFKLNEEIAIILASFSASTSAFVETVKGLDYKAFKQIVESCGN
FKVTKGKAKKGAWNIGEQKSILSPLYAFASEAARVVRSIFSRTLETAQNSVRVLQKAA
ITILDGISQYSLRLIDAMMFTSDLATNNLVVMAYITGGVVQLTSQWLTNIFGTVYEKL
KPVLDWLEEKFKEGVEFLRDGWEIVKFISTCAXXXXXXXXXXXXXXXXXXXXFKLV
NKXXXXXXXXXIIGGAKLKALNLGETFVTHSKGLYRKCVKSREETGLLMPLKAPKEII
FLEGETLPTEXLTXXVVLKTGDLQPLEQPTSEAVEAPLVGTPVCINGLMLLEIKDTEK
YCALAPNMMVTNNTFTLKGGAPTKVTFGDDTVIEVQGYKSVNITFELDERIDKVLNEK
CSAYTVELGTEVNEFACVVADAVIKTLQPVSELLTPLGIDLDEWSMATYYLFDESGEF
KLASHMYCSFYPPDEDEEEGDCEEEEFEPSTQYEYGTEDDYQGKPLEFGATSAALQPE
EEQEEDWLDDDSQQTVGQQDGSEDNQTTTIQTIVEVQPQLEMELTPVVQTIEVNSFSG
YLKLTDNVYIKNADIVEEAKKVKPTVVVNAANVYLKHGGGVAGALNKATNNAMQVESD
DXIATNGPLKVGGSCVLSGHNLAKHCLHVVGPNVNKGEDIQLLKSAYENFNQHEVLLA
PLLSAGIFGADPIHSLRVCVDTXXXXXYXAXFDKNLYDKLVSSFLEMKSEKQVEQKIA
EIPKEEVKPFITESKPSVEQRKQDDKKIKACVEEVTTTLEETKFLTENLLLYIDINGN
LHPDSATLVSDIDITFLKKDAPYIVGDVVQEGVLTAVVIPTKKAGGTTEMLAKALRKV
PTDNYITTYPGQGLNGYTVEEAKTVLKKCKSAFYILPSIISNEKQEILGTVSWNLREM
LAHAEETRKLMPVCVETKAIVSTIQRKYKGIKIQEGVVDYGARFYFYTSKTTVASLIN
TLNDLNETLVTMPLGYVTHGLNLEEAARYMRSLKVPATVSVSSPDAVTAYNGYLTSSS
KTPEEHFIETISLAGSYKDWSYSGQSTQLGIEFLKRGDKSVYYTSNPTTFHLDGEVIT
FDNLKTLLSLREVRTIKVFTTVDNINLHTQVVDMSMTYGQQFGPTYLDGADVTKIKPH
NSHEGKTFYVLPNDDTLRVEAFEYYHTTDPSFLGRYMSALNHTKKWKYPQVNGLTSIK
WADNNCYLATALLTLQQIELKFNPPALQDAYYRARAGEAANFCALILAYCNKTVGELG
DVRETMSYLFQHANLDSCKRVLNVVCKTCGQQQTTLKGVEAVMYMGTLSYEQFKKGVQ
IPCTCGKQATKYLVQQESPFVMMSAPPAQYELKHGTFTCASEYTGNYQCGHYKHITSK
ETLYCIDGALLTKSSEYKGPITDVFYKENSYTTTIKPVTYKLDGVVCTEIDPKLDNYY
KKDNSYFTEQPIDLVPNQPYPNASFDNFKFVCDNIKFADDLNQLTGYKKPASRELKVT
FFPDLNGDVVAIDYKHYTPSFKKGAKLLHKPIVWHVNNATNKATYKPNTWCIRCLWST
KPVETSNSFDVLKSEDAQGMDNLACEDLKPVSEEVVENPTIQKDVLECNVKTTEVVGD
IILKPANNSLKITEEVGHTDLMAAYVDNSSLTIKKPNELSRVLGLKTLATHGLAAVNS

VPWDTIANYAKPFLNKVVSTTTNIVTRCLNRVCTNYMPYFFTLLLQLCTFTRSTNSRI
KASMPTTIAKNTVKSVGKFCLEASFNYLKSPNFSKLINIIIWFLLLSVCLGSLIYSTA
ALGVLMSNLGMPSYCTGYREGYLNSTNVTIATYCTGSIPCSVCLSGLDSLDTYPSLET
IQITISSFKWDLTAFGLVAEWFLAYILFTRFFYVLGLAAIMQLFFSYFAVHFISNSWL
MWLIINLVQMAPISAMVRMYIFFASFYYVWKSYVHVVDGCNSSTCMMCYKRNRATRVE
CTTIVNGVRRSFYVYANGGKGFCKLHNWNCVNCDTFCAGSTFISDEVARDLSLQFKRP
INPTDQSSYIVDSVTVKNGSIHLYFDKAGQKTYERHSLSHFVNLDNLRANNTKGSLPI
NVIVFDGKSKCEESSAKSASVYYSQLMCQPILLLDQALVSDVGDSAEVAVKMFDAYVN
TFSSTFNVPMEKLKTLVATAEAELAKNVSLDNVLSTFISAARQGFVDSDVETKDVVEC
LKLSHQSDIEVTGDSCNNYMLTYNKVENMTPRDLGACIDCSARHINAQVAKSHNIALI
WNVKDFMSLSEQLRKQIRSAAKKNNLPFKLTCATTRQVVNVVTTKIALKGGKIVNNWL
KQLIKVTLVFLFVXXIFYLITPVHVMSKHTDFSSEIIGYKAIDGGVTRDIASTDTCFA
NKHADFDTWFSQRGGSYTNDKACPLIAAVITREVGFVVPGLPGTILRTTNGDFLHFLP
RVFSAVGNICYTPSKLIEYTDFATSACVLAAECTIFKDASGKPVPYCYDTNVLEGSVA
YESLRPDTRYVLMDGSIIQFPNTYLEGSVRVVTTFDSEYCRHGTCERSEVGVCVSTSG
RWVLNNDYYRSLPGVFCGVDAVNLLTNMFTPLIQPIGALDISASIVAGGIVAIVVTCL
AYYFMRFRRAFGEYSHVVAFNTLLFLMSFTVLCLTPVYSFLPGVYSVIYLYLTFYLTN
DVSFLAHIQWMVMFTPLVPFWITIAYIICISTKHFYWFFSNYLKRRVVFNGVSFSTFE
EAALCTFLLNKEMYLKLRSDVLLPLTQYNRYLALYNKYKYFSGAMDTTSYREAACCHL
AKALNDFSNSGSDVLYQPPQTSITSAVLQSGFRKMAFPSGKVEGCMVQVTCGTTTLNG
LWLDDVVYCPRHVICTSEDMLNPNYEDLLIRKSNHNFLVQAGNVQLRVIGHSMQNCVL
KLKVDTANPKTPKYKFVRIQPGQTFSVLACYNGSPSGVYQCAMRPNFTIKGSFLNGSC
GSVGFNIDYDCVSFCYMHHMELPTGVHAGTDLEGNFYGPFVDRQTAQAAGTDTTITVN
VLAWLYAAVINGDRWFLNRFTTTLXXXXLVAMKYNYEPLTQDHVDILGPLSAQTGIAV
LDMCASLKELLQNGMNGRTILGSALLEDEFTPFDVVRQCSGVTFQSAVKRTIKGTHHW
XXXXILTSLLVLVQSTQWSLFFFLYENAFLPFAMGIIAMSAFAMMFVKHKHAFLCLFL
LPSLATVAYFNMVYMPASWVMRIMTWLDMVDTSLSGFKLKDCVMYASAVVLLILMTAR
TVYDDGARRVWTLMNVLTLVYKVYYGNALDQAISMWALIISVTSNYSGVVTTVMFLAR
GIVFMCVEYCPIFFITGNTLQCIMLVYCFLGYFCTCYFGLFCLLNRYFRLTLGVYDYL
VSTQEFRYMNSQGLLPPKNSIDAFKLNIKLLGVGGKPCIKVATVQSKMSDVKCTSVVL
LSVLQQLRVESSSKLWAQCVQLHNDILLAKDTTEAFEKMVSLLSVLLSMQGAVDINKL
CEEMLDNRATLQAIASEFSSLPSYAAFATAQEAYEQAVANGDSEVVLKKLKKSLNVAK
SEFDRDAAMQRKLEKMADQAMTQMYKQARSEDKRAKVTSAMQTMLFTMLRKLDNDALN
NIINNARDGCVPLNIIPLTTAAKLMVVIPDYNTYKNTCDGTTFTYASALWEIQQVVDA
DSKIVQLSEISMDNSPNLAWPLIVTALRANSAVKLQNNELSPVALRQMSCAAGTTQTA
CTDDNALAYYNTTKGGRFVLALLSDLQDLKWARFPKSDGTGTIYTELEPPCRFVTDTP
KGPKVKYLYFIKGLNNLNRGMVLGSLAATVRLQAGNATEVPANSTVLSFCAFAVDAAK
AYKDYLASGGQPITNCVKMLCTHTGTGQAITVTPEANMDQESFGGASCCLYCRCHIDH
PNPKGFCDLKGKYVQIPTTCANDPVGFTLKNTVCTVCGMWKGYGCSCDQLREPMLQSA
DAQSFLNRVCGVSAARLTPCGTGTSTDVVYRAFDIYNDKVAGFAKFLKTNCCRFQEKD
EDDNLIDSYFVVKRHTFSNYQHEETIYNLLKDCPAVAKHDFFKFRIDGDMVPHISRQR
LTKYTMADLVYALRHFDEGNCDTLKEILVTYNCCDDDYFNKKDWYDFVENPDILRVYA
NLGERVRQALLKTVQFCDAMRNAGIVGVLTLDNQDLNGNWYDFGDFIQTTPGSGVPVV
DSYYSLLMPILTLTRALTAESHVDTDLTKPYIKWDLLKYDFTEERLKLFDRYFKYWDQ
TYHPNCVNCLDDRCILHCANFNVLFSTVFPLTSFGPLVRKIFVDGVPFVVSTGYHFRE

LGVVHNQDVNLHSSRLSFKELLVYAAXPAMHAASGNLLLDKRTTCFSVAALTNNVAFQ
TVKPGNFNKDFYDFAVSKGFFKEGSSVELKHFFFAQDGNAAISDYDYYRYNLPTMCDI
RQLLFVVEVVDKYFDCYDGGXINANQVIVNNLDKSAGFPFNKWGKARLYYDSMSYEDQ
DALFAYTKRNVIPTITQMNLKYAISAKNRARTVAGVSICSTMTNRQFHQKLLKSIAAT
RGATVVIGTSKFYGGWHNMLKTVYSDVENPHLMGWDYPKCDRAMPNMLRIMASLVLAR
KHTTCCSLSHRFYRLANECAQVL**S**EMVMCGGSLYVKPGGTSSGDATTAYANSVFNICQ
AVTANVNALLSTDGNKIADKYVRNLQHRLYECLYRNRDVDTDFVNEFYAYLRKHFSMM
ILSDDAVVCFNSTYASQGLVASIKNFKSVLYYQNNVFMSEAKCWTETDLTKGPHEFCS
QHTMLVKQGDDYVYLPYPDPSRILGAGCFVDDIVKTDGTLMIERFVSLAIDAYPLTKH
PNQEYADVFHLYLQYIRKLHDELTGHMLDMYSVMLTNDNTSRYWEPEFYEAMYTPHTV
LQAVGACVLCNSQTSLRCGACIRRPFLCCKCCYDHVISTSHKLVLSVNPYVCNAPGCD
VTDVTQLYLGGMSYYCKSHKPPISFPLCANGQVFGLYKNTCVGSDNVTDFNAIATCDW
TNAGDYILANTCTERLKLFAAETLKATEETFKLSYGIATVREVLSDRELHLSWEVGKP
RPPLNRNYVFTGYRVTKNSKVQIGEYTFEKGDYGDAVVYRGTTTYKLNVGDYFVLTSH
TVMPLSAPTLVPQEHYVRITGLYPTLNISDEFSSNVANYQKVGMQKYSTLQGPPGTGK
SHFAIGLALYYPSARIVYTACSHAAVDALCEKALKYLPIDKCSRIIPARARVECFDKF
KVNSTLEQYVFCTVNALPETTADIVVFDEISMATNYDLSVVNARLRAKHYVYIGDPAQ
LPAPRTLLTKGTLEPEYFNSVCRLMKTIGPDMFLGTCRRCPAEIVDTVSALVYDNKLK
AHKDKSAQCFKMFYKGVITHDVSSAINRPQIGVVREFLTRNPAWRKAVFISPYNSQNA
VASKILGLPTQTVDSSQGSEYDYVIFTQTTETAHSCNVNRFNVAITRAKVGILCIMSD
RDLYDKLQFTSLEIPRRNVATLQAENVTGLFKDCSKVITGLHPTQAPTHLSVDTKFKT
EGLCVDIPGIPKDMTYRRLISMMGFKMNYQVNGYPNMFITREEAIRHVRAWIGFDVEG
CHATREAVGTNLPLQLGFSTGVNLVAVPTGYVDTPNNTDFSRVSAKPPPGDQFKHLIP
LMYKGLPWNVVRIKIVQMLSDTLKNLSDRVVFVLWAHGFELTSMKYFVKIGPERTCCL
CDRRATCFSTASDTYACWHHSIGFDYVYNPFMIDVQQWGFTGNLQSNHDLYCQVHGNA
HVASCDAIMTRCLAVHECFVKRVDWTIEYPIIGDELKINAACRKVQHMVVKAALLADK
FPVLHDIGNPKAIKCVPQADVEWKFYDAQPCSDKAYKIEELFYSYATHSDKFTDGVCL
FWNCNVDRYPANSIVCRFDTRVLSNLNLPGCDGGSLYVNKHAFHTPAFDKSAFVNLKQ
LPFFYYSDSPCESHGKQVVSDIDYVPLKSATCITRCNLGGAVCRHHANEYRLYLDAYN
MMISAGFSLWVYKQFDTYNLWNTFTRLQSLENVAFNVVNKGHFDGQQGEVPVSIINNT
VYTKVDGVDVELFENKTTLPVNVAFELWAKRNIKPVPEVKILNNLGVDIAANTVIWDY
KRDAPAHISTIGVCSMTDIAKKPTETICAPLTVFFDGRVDGQVDLFRNARNGVLITEG
SVKGLQPSVGPKQASLNGVTLIGEAVKTQFNYYKKVDGVXQQXXXXXXTQSRNLQEFK
PRSQMEIDFLELAMDEFIERYKLEGYAFEHIVYGDFSHSQLGGLHLLIGLAKRFKESP
FELEDFIPMDSTVKNYFITDAQTGSSKCVCSVIDLLLDDFVEIIKSQDLSVVSKVVKV
TIDYTEISFMLWCKDGHVETFYPKLQSSQAWQPGVAMPNLYKMQRMLLEKCDLQNYGD
SATLPKGIMMNVAKYTQLCQYLNTLTLAVPYNMRVIHFGAGSDKGVAPGTAVLRQWLP
TGTLLVDSDLNDFVSDADSTLIGDCATVHTANKWDLIISDMYDPKTKNVTKENDSKEG
FFTYICGFIQQKLALGGSVAIKITEHSWNADLYKLMGHFAWWTAFVTNVNASSSEAFL
IGCNYLGKPREQIDGYVMHANYIFWRNTNPIQLSSYSLFDMSKFPLKLRGTAVMSLKE
GQINDMILSLLSKGRLIIRENNRVVISSDVLVNN"

mat_peptide <1..>22

/gene="ORF1ab"

/product="RNA-dependent RNA polymerase"

ORIGIN

1 gtgaratggt catgtgtggc gg

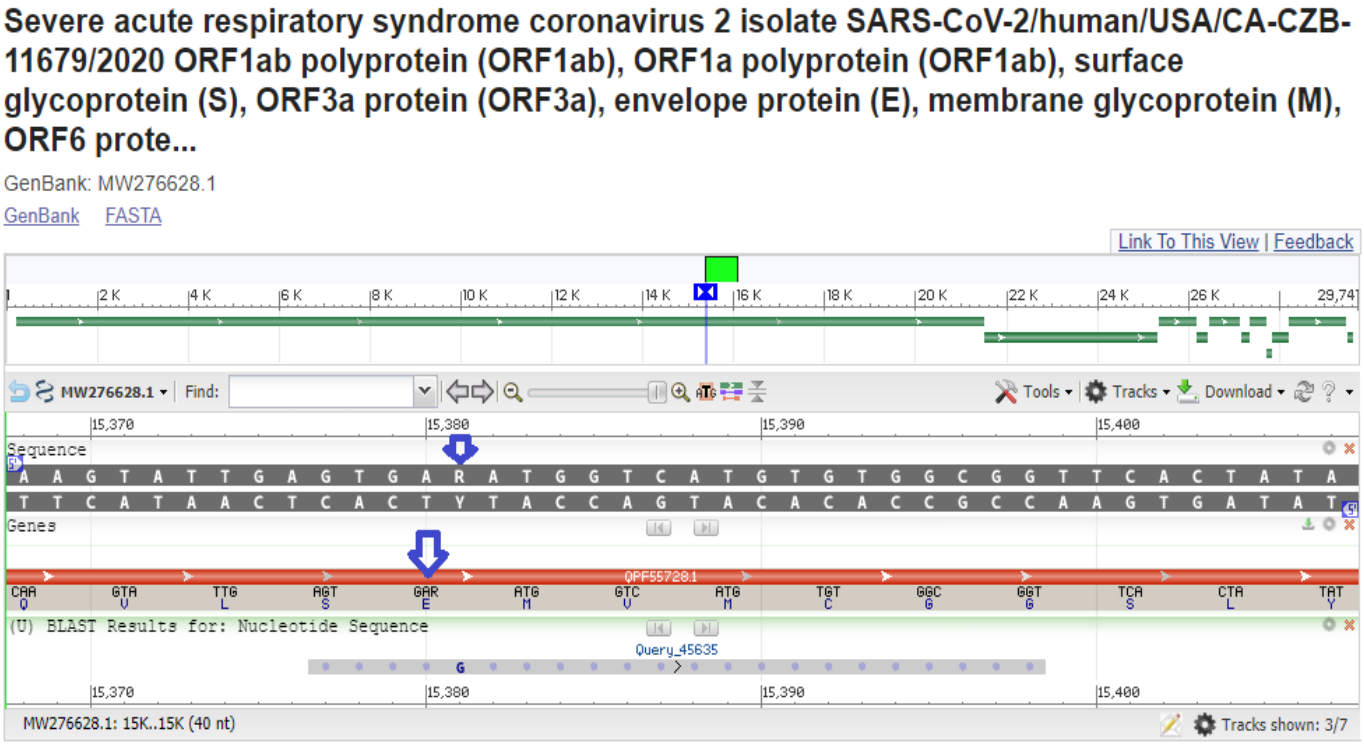
//

Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/CA-CZB-11679/2020 ORF1ab polyprotein (ORF1ab), ORF1a polyprotein (ORF1ab), surface glycoprotein (S), ORF3a protein (ORF3a), envelope protein (E), membrane glycoprotein (M), ORF6 protein (ORF6), ORF7a protein (ORF7a), ORF7b (ORF7b), ORF8 protein (ORF8), nucleocapsid phosphoprotein (N), and ORF10 protein (ORF10) genes, complete cds
Sequence ID: MW276628.1 Length: 29741 Number of Matches: 1
Range 1: 15377 to 15398 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
40.1 bits(20)	0.59	21/22(95%)	0/22(0%)	Plus/Plus

```
CDS: Putative 1          1          E   M   V   M   C   G   G
Query                    1      GTGAGATGGTCATGTGTGGCGG   22
                                |||| ||||||||||||||||||
Sbjct                    15377  GTGARATGGTCATGTGTGGCGG   15398
CDS:ORF1ab polyprote     5056   S   E   M   V   M   C   G   G
```

ATTENTION: R = A or G



5.3 BLAST Research on the RNA SEQUENCE "GTGAGATGGTCATGTGTGGCGG" of the RdRp Gene of the coronavirus isolate bat-SL-CoVZC45, complete genome" – (Hitlist size: 1000)

Query: **GTGAGATGGTCATGTGTGGCGG**

Search Parameters	
Program	blastn
Word size	7
Expect value	1000
Hitlist size	1000
Match/Mismatch scores	1,-3
Gapcosts	5,2
Filter string	F
Genetic Code	1

1000 sequences selected

Accession Seq	Description	Comomm Name	Max Score	Total Score	Query C.	E value	Per. Id.	Acc. Len
.....	.....	.....	.....	.....	.....	.....	.....	.....
KY775091.1	Rhinolophus pusillus coronavirus isolate CLR7 RNA-dependent RNA polymerase gene, partial cds	Rhinolophus pusillus coronavirus	44.1	44.1	100%	0.037	100.00%	438
KY775088.1	Rhinolophus pusillus coronavirus isolate SSR1 RNA-dependent RNA polymerase gene, partial cds	Rhinolophus pusillus coronavirus	44.1	44.1	100%	0.037	100.00%	414
KY775087.1	Rhinolophus pusillus coronavirus isolate SSR2 RNA-dependent RNA polymerase gene, partial cds	Rhinolophus pusillus coronavirus	44.1	44.1	100%	0.037	100.00%	405
KY775086.1	Rhinolophus pusillus coronavirus isolate SSR10 RNA-dependent RNA polymerase gene, partial cds	Rhinolophus pusillus coronavirus	44.1	44.1	100%	0.037	100.00%	435
KY775084.1	Rhinolophus pusillus coronavirus isolate ZSR4 RNA-dependent RNA polymerase gene, partial cds	Rhinolophus pusillus coronavirus	44.1	44.1	100%	0.037	100.00%	402
KY775083.1	Rhinolophus pusillus coronavirus isolate ZSR6 RNA-dependent RNA polymerase gene, partial cds	Rhinolophus pusillus coronavirus	44.1	44.1	100%	0.037	100.00%	426
KY775079.1	Rhinolophus pusillus coronavirus isolate ZSR31 RNA-dependent RNA polymerase gene, partial cds	Rhinolophus pusillus coronavirus	44.1	44.1	100%	0.037	100.00%	435
KY775078.1	Rhinolophus pusillus coronavirus isolate ZSR42 RNA-dependent RNA polymerase gene, partial cds	Rhinolophus pusillus coronavirus	44.1	44.1	100%	0.037	100.00%	427
.....	.....	.....	.....	.....	.....	.....	.....	.....
KT381914.1	Rhinolophus monoceros coronavirus isolate CYCU-R9/TW/2013 RNA-dependent RNA polymerase gene, partial cds	Rhinolophus monoceros coronavirus	44.1	44.1	100%	0.037	100.00%	375

.....	.....	.....	.....	.....	.....	.....	.....	.....
KF569996.1	**Rhinolophus affinis coronavirus isolate LYRa11, complete genome**	Rhinolophus affinis coronavirus	44.1	44.1	100%	0.037	100.00%	29805
KF569973.1	**Rhinolophus affinis coronavirus isolate LYRa3 RNA-dependent RNA polymerase gene, partial cds**	Rhinolophus affinis coronavirus	44.1	44.1	100%	0.037	100.00%	438
.....	.....	.....	.....	.....	.....	.....	.....	.....
CP016545.1	**Altererythrobacter namhicola strain JCM 16345, complete genome**	Altererythrobacter namhicola	38.2	38.2	86%	2.3	100.00%	2591679
.....	.....	.....	.....	.....	.....	.....	.....	.....
XR_001826207.2	**PREDICTED: Cebus imitator uncharacterized LOC108306613 (LOC108306613), transcript variant X2, ncRNA**	Panamanian white-faced capuchin	38.2	38.2	86%	2.3	100.00%	804
XR_001826208.2	**PREDICTED: Cebus imitator uncharacterized LOC108306613 (LOC108306613), transcript variant X1, ncRNA**	Panamanian white-faced capuchin	38.2	38.2	86%	2.3	100.00%	567
.....	.....	.....	.....	.....	.....	.....	.....	.....
OD593106.1	**4_Tbi_b3v08**	Timema bartmani	36.2	36.2	81%	9.1	100.00%	2568
.....	.....	.....	.....	.....	.....	.....	.....	.....
OE023157.1	**4_Tte_b3v08**	Timema tahoe	36.2	36.2	81%	9.1	100.00%	2611
.....	.....	.....	.....	.....	.....	.....	.....	.....

CHAPTER VI

BASE SUBSTITUTIONS in the RNA SEQUENCE "GTGAAATGGTCATGTGTGGCGG" of the RdRp GENE (RdRp_SARSr-F) and SUBSEQUENT BLAST RESEARCH

GOALS OF CHAPTER VI

While in the previous chapter BLAST research was carried out on the sequences **"GTGAAATGGTTATGTGTGGCGG"** and **"GTGAGATGGTCATGTGTGGCGG"** (the two sequences of the **RdRp gene** whose significant alignments of BLAST researches identify the ***Bat*** and the ***Pangolin***), in this chapter, in order not to neglect the impact of any other changes in nitrogenous bases in the same positions highlighted in blue in the sequence **"GTGAAATGGTCATGTGTGGCGG"** and identify any other "organisms" relevant for the purposes of this study, **all the possible base substitutions** were performed, obtaining the following **4 sequences** as a result:

"**GTGAAATGGTAATGTGTGGCGG**", "**GTGAAATGGTGATGTGTGGCGG**"

"**GTGATATGGTCATGTGTGGCGG**", "**GTGACATGGTCATGTGTGGCGG**"

These sequences have also been the object of **four *complete BLAST Researches***.
The following are the "**organisms**" that have been found in "**common**" between these *four sequences* and *other sequences* shown in subsequent chapters. To be more easily distinguishable, either by scrolling through the document or by using the "**Find**" function, the "**common organisms**" have been appropriately *colored differently from each other*.

1) Sequence "**GTGAAATGGTAATGTGTGGCGG**":

Gouania willdenowi, **Myripristis murdjan**, **Salmo trutta**, **Branchiostoma belcheri**, **Acomys russatus**, **3_Tce_b3v08** e **3_Tms_b3v08**.

2) Sequence "**GTGAAATGGTGATGTGTGGCGG**":

Coregonus sp. 'balchen', **Thalassophryne amazonica**, **Sparus aurata, Ipomoea triloba** e **trifida**, **Solanum pennellii**, **Oryza sativa Indica** e **2_Tsi_b3v08**.

3) Sequence "**GTGATATGGTCATGTGTGGCGG**":

Chanos chanos, **Corvus brachyrhynchos**, **1_Tps_b3v08** e **1_Tdi_b3v08.**

4) Sequence "**GTGACATGGTCATGTGTGGCGG**":

Human DNA.

6.1 BLAST Research on the modified RNA SEQUENCE "GTGAAATGGTAATGTGTGGCGG" of the RdRp Gene – (Hitlist size: 1000)

Query: **GTGAAATGGTAATGTGTGGCGG**

Search Parameters	
Program	blastn
Word size	7
Expect value	1000
Hitlist size	1000
Match/Mismatch scores	1,-3
Gapcosts	5,2
Filter string	F
Genetic Code	1

1000 sequences selected:

Accession Seq	Description	Comomm Name	Max Score	Total Score	Query C.	E value	Per. Id.	Acc. Len
LR131992.1	Gouania willdenowi genome assembly, chromosome: 16	blunt-snouted clingfish	40.1	40.1	90%	0.57	100.00%	40234402
LR597565.1	Myripristis murdjan genome assembly, chromosome: 16	pinecone soldierfish	38.2	100	90%	2.3	100.00%	34798541
LR584429.1	Salmo trutta genome assembly, chromosome: 20	river trout	38.2	38.2	86%	2.3	100.00%	55161844
LR584427.1	Salmo trutta genome assembly, chromosome: 19	river trout	38.2	70.4	86%	2.3	100.00%	56583805
XM_019775117.1	PREDICTED: Branchiostoma belcheri uncharacterized LOC109474756 (LOC109474756), mRNA	Belcher's lancelet	38.2	38.2	86%	2.3	100.00%	5179
XM_007315056.1	Serpula lacrymans var. lacrymans S7.9 hypothetical protein (SERLADRAFT_459811), mRNA	Serpula lacrymans var. lacrymans S7.9	38.2	38.2	86%	2.3	100.00%	1545
LR877218.1	Acomys russatus genome assembly, chromosome: 7	golden spiny mouse	38.2	68.4	86%	2.3	100.00%	69034057
MW444836.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/WA-S2917/2020 ORF1ab polyprotein (ORF1ab), ORF1a polyprotein (ORF1ab), surface glycoprotein (S), ORF3a protein (ORF3a), envelope protein (E), membrane glycoprotein (M), ORF6 protein (ORF6), ORF7a protein (ORF7a), ORF7b (ORF7b), ORF8 protein (ORF8), nucleocapsid phosphoprotein (N), and ORF10 protein (ORF10) genes, complete cds	Severe acute respiratory syndrome coronavirus 2	36.2	36.2	100%	9.2	95.45%	29813
.....	.....	.....	.....	.....	.....	.....	.....	.....
OC337859.1	3_Tce_b3v08	Timema cristinae	36.2	36.2	81%	9.2	100.00%	3065
.....	.....	.....	.....	.....	.....	.....	.....	.....
OB796542.1	3_Tms_b3v08	Timema monikensis	36.2	36.2	81%	9.2	100.00%	61900
.....	.....	.....	.....	.....	.....	.....	.....	.....

ALIGNMENT VIEW FLAT QUERY-ANCHORED WITH LETTERS FOR IDENTITIES

Query range 1: 1 to 22

Query	**1**	**GTGAAATGGTAATGTGTGGCGG**	**22**
LR131992.1	22487576	GTGAAATGGT**A**ATGTGTGGC	22487595
LR597565.1	10507530	TGAAATGGT**A**ATGTGTGGC	10507548
LR597565.1	33051678	GAAATGGT**A**ATGTGTG	33051663
LR597565.1	19180778	GTGAAATGGT**A**ATGT	19180764
LR584429.1	36663303	TGAAATGGT**A**ATGTGTGGC	36663285
LR584427.1	33634069	GTGAAATGGT**A**ATGTGTGG	33634051
LR584427.1	34417632	TGAAATGGT**A**ATGTGT	34417617
XM_019775117.1	4946	GTGAAATGGT**A**ATGTGTGG	4928
XM_007315056.1	618	TGAAATGGT**A**ATGTGTGGC	600
LR877218.1	10353639	GTGAAATGGT**A**ATGTGTGG	10353657
LR877218.1	1417218	AATGGT**A**ATGTGTGG	1417204
MW444836.1	15393	GTGAAATGGT**C**ATGTGTGGCGG	15414
OC337859.1	2280	TGAAATGGT**A**ATGTGTGG	2263
OB796542.1	42576	TGAAATGGT**A**ATGTGTGG	42593

6.2 BLAST Research on the modified RNA SEQUENCE "GTGAAATGGTGATGTGTGGCGG" of the RdRp Gene – (Hitlist size: 1000)

Query: **GTGAAATGGTGATGTGTGGCGG**

Search Parameters	
Program	blastn
Word size	7
Expect value	1000
Hitlist size	1000
Match/Mismatch scores	1,-3
Gapcosts	5,2
Filter string	F
Genetic Code	1

1000 Sequences selected:

Accession Seq	Description	Comomm Name	Max Score	Total Score	Query C.	E value	Per. Id.	Acc. Len
LR778259.1	Coregonus sp. 'balchen' genome assembly, chromosome: 7	Coregonus sp. 'balchen'	38.2	68.4	86%	2.3	100.00%	65391737
LR722978.1	Thalassophryne amazonica genome assembly, chromosome: 13	Thalassophryne amazonica	38.2	348	90%	2.3	100.00%	99659484
LR537138.1	Sparus aurata genome assembly, chromosome: 18	gilthead seabream	38.2	98.6	90%	2.3	100.00%	37104330
CP025673.1	Ipomoea triloba cultivar NCNSP0323 chromosome 14	trilobed morning glory	38.2	38.2	86%	2.3	100.00%	23651743
CP025657.1	Ipomoea trifida cultivar NCNSP0306 chromosome 14	Ipomoea trifida	38.2	38.2	86%	2.3	100.00%	19195807
XR_002546422.1	PREDICTED: Zootermopsis nevadensis piggyBac transposable element-derived protein 4-like (LOC110839295), transcript variant X3, misc_RNA	Zootermopsis nevadensis	38.2	38.2	86%	2.3	100.00%	2565
XM_022083378.1	PREDICTED: Zootermopsis nevadensis piggyBac transposable element-derived protein 4-like (LOC110839295), transcript variant X2, mRNA	Zootermopsis nevadensis	38.2	38.2	86%	2.3	100.00%	2569
XM_022083377.1	PREDICTED: Zootermopsis nevadensis piggyBac transposable element-derived protein 4-like (LOC110839295), transcript variant X1, mRNA	Zootermopsis nevadensis	38.2	38.2	86%	2.3	100.00%	3415
HG975450.1	Solanum pennellii chromosome ch11, complete genome	Solanum pennellii	38.2	70.4	86%	2.3	100.00%	66223686
MW444836.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/WA-S2917/2020 ORF1ab polyprotein (ORF1ab), ORF1a polyprotein (ORF1ab), surface glycoprotein (S), ORF3a protein (ORF3a), envelope protein (E), membrane glycoprotein (M), ORF6 protein (ORF6), ORF7a protein	Severe acute respiratory syndrome coronavirus 2	36.2	36.2	100%	9.2	95.45%	29813

	(ORF7a), ORF7b (ORF7b), ORF8 protein (ORF8), nucleocapsid phosphoprotein (N), and ORF10 protein (ORF10) genes, complete cds							
.....	.....	.....	.....	.....	.....	.....	.....	.....
CP054686.1	Oryza sativa Indica Group cultivar Minghui 63 chromosome 11	Oryza sativa Indica Group	36.2	134	95%	9.2	100.00%	34100580
.....	.....	.....	.....	.....	.....	.....	.....	.....
OC005272.1	2_Tsi_b3v08	Timema shepardi	36.2	36.2	81%	9.2	100.00%	52995
.....	.....	.....	.....	.....	.....	.....	.....	.....

ALIGNMENT VIEW FLAT QUERY-ANCHORED WITH LETTERS FOR IDENTITIES

Query range 1: 1 to 22

Query	1	GTGAAATGGTGATGTGTGGCGG	22
LR778259.1	43904207	GTGAAATGGTGATGTGTGG	43904189
LR778259.1	23381906	TGAAATGGTGATGTG	23381920
LR722978.1	59908563	TGAAATGGTGATGTGTGGC	59908581
LR722978.1	33466936	GAAATGGTGATGTGTGG	33466952
LR722978.1	52466582	AAATGGTGATGTGTGG	52466597
LR722978.1	88735259	ATGGTGATGTGTGGCG	88735274
LR722978.1	34565888	AAATGGTGATGTGTG	34565902
LR722978.1	43597785	TGAAATGGTGATGTG	43597799
LR722978.1	48408969	GAAATGGTGATGTGT	48408955
LR722978.1	51014259	GAAATGGTGATGTGT	51014245
LR722978.1	60018449	TGAAATGGTGATGTG	60018463
LR722978.1	81496758	GAAATGGTGATGTGT	81496744
LR722978.1	91961262	TGAAATGGTGATGTG	91961276
LR537138.1	35644928	TGAAATGGTGATGTGTGGC	35644946
LR537138.1	2734107	TGGTGATGTGTGGCG	2734093
LR537138.1	28927756	AATGGTGATGTGTGG	28927770
CP025673.1	4838415	AAATGGTGATGTGTGGCGG	4838433
CP025657.1	4115448	AAATGGTGATGTGTGGCGG	4115466
XR_002546422.1	2426	GTGAAATGGTGATGTGTGG	2444
XM_022083378.1	2430	GTGAAATGGTGATGTGTGG	2448
XM_022083377.1	3276	GTGAAATGGTGATGTGTGG	3294
HG975450.1	59839601	GAAATGGTGATGTGTGGCG	59839619
HG975450.1	41067225	AAATGGTGATGTGTGG	41067240
MW444836.1	15393	GTGAAATGGTCATGTGTGGCGG	15414
CP054686.1	24319927	GTGAAATGGTGATGTGTG	24319910
CP054686.1	5164724	GTGAAATGGTGATGTGT	5164740
CP054686.1	24402614	GTGAAATGGTGATGTGT	24402598
CP054686.1	10548306	TGGTGATGTGTGGCG	10548292
OC005272.1	4696	TGAAATGGTGATGTGTGG	4679

6.3 BLAST Research on the modified RNA SEQUENCE "GTGATATGGTCATGTGTGGCGG" of the RdRp Gene – (Hitlist size: 1000)

Query: **GTGATATGGTCATGTGTGGCGG**

Search Parameters	
Program	blastn
Word size	7
Expect value	1000
Hitlist size	1000
Match/Mismatch scores	1,-3
Gapcosts	5,2
Filter string	F
Genetic Code	1

1000 sequences selected:

Accession Seq	Description	Comomm Name	Max Score	Total Score	Query C.	E value	Per. Id.	Acc. Len
LR697110.1	Chanos chanos genome assembly, chromosome: 5	milkfish	38.2	98.6	86%	2.2	100.00%	53690977
XR_001863586.1	PREDICTED: Corvus brachyrhynchos uncharacterized LOC108447881 (LOC108447881), ncRNA	American crow	38.2	38.2	86%	2.2	100.00%	3419
MW444836.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/WA-S2917/2020 ORF1ab polyprotein (ORF1ab), ORF1a polyprotein (ORF1ab), surface glycoprotein (S), ORF3a protein (ORF3a), envelope protein (E), membrane glycoprotein (M), ORF6 protein (ORF6), ORF7a protein (ORF7a), ORF7b (ORF7b), ORF8 protein (ORF8), nucleocapsid phosphoprotein (N), and ORF10 protein (ORF10) genes, complete cds	Severe acute respiratory syndrome coronavirus 2	36.2	36.2	100%	9.2	95.45%	29813
.....	.....	.....	.....	.....	.....	.....	.....	.....
OD009150.1	1_Tps_b3v08	Timema poppensis	36.2	36.2	81%	9.2	100.00%	20538
OA566508.1	1_Tdi_b3v08	Timema douglasi	36.2	36.2	81%	8.9	100.00%	118698

ALIGNMENT VIEW FLAT QUERY-ANCHORED WITH LETTERS FOR IDENTITIES

Query range 1: 1 to 22

```
Query           1         GTGATATGGTCATGTGTGGCGG  22
LR697110.1      35827398  GTGATATGGTCATGTGTGG     35827416
LR697110.1      10569623   TGATATGGTCATGTG        10569637
LR697110.1      8521643    TGATATGGTCATGTG        8521629
XR_001863586.1  919       GTGATATGGTCATGTGTGG     901
MW444836.1      15393     GTGAAATGGTCATGTGTGGCGG  15414
OD009150.1      17979     GTGATATGGTCATGTGTG      17996
OA566508.1      78403     GTGATATGGTCATGTGTG      78420
```

6.4 BLAST Research on the modified RNA SEQUENCE "GTGACATGGTCATGTGTGGCGG" of the RdRp Gene – (Hitlist size: 100)

Query: **GTGACATGGTCATGTGTGGCGG**

Search Parameters	
Program	Blastn
Word size	7
Expect value	1000
Hitlist size	100
Match/Mismatch scores	1,-3
Gapcosts	5,2
Filter string	F
Genetic Code	1

100 sequences selected:

Accession Seq	Description	Comomm Name	Max Score	Total Score	Query C.	E value	Per. Id.	Acc. Len
AL162393.15	**Human DNA sequence from clone RP11-534I8 on chromosome 9, complete sequence**	Human	38.2	38.2	86%	2.2	100.00%	184302
MW444836.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/WA-S2917/2020 ORF1ab polyprotein (ORF1ab), ORF1a polyprotein (ORF1ab), surface glycoprotein (S), ORF3a protein (ORF3a), envelope protein (E), membrane glycoprotein (M), ORF6 protein (ORF6), ORF7a protein (ORF7a), ORF7b (ORF7b), ORF8 protein (ORF8), nucleocapsid phosphoprotein (N), and ORF10 protein (ORF10) genes, complete cds	Severe acute respiratory syndrome coronavirus 2	36.2	36.2	100%	9.2	95.45%	29813
.....	.....	.....	.....	.....	.....	.....	.....	.....

ALIGNMENT VIEW FLAT QUERY-ANCHORED WITH LETTERS FOR IDENTITIES

Query range 1: 1 to 22

```
Query       1       GTGACATGGTCATGTGTGGCGG  22
AL162393.15 144164  GTGACATGGTCATGTGTGG     144182
MW444836.1  15393   GTGAAATGGTCATGTGTGGCGG  15414
```

CHAPTER VII

PROCEDURE FOR THE MODIFICATION OF THE ORIGINAL SEQUENCE AND THE GENERATION OF NEW SEQUENCES

GOALS OF CHAPTER VII

This chapter deals with the important and revolutionary topic of DNA or RNA sequence analysis and modification of based on the principles of **TRICHROMATIC THEORY OF EQUILIBRIUM OF SYSTEMS** (**T.T.E.S.**): the **procedure** through which one can **modify the Original Sequence** and **generate new DNA or RNA sequences**.
This procedure introduces a profound review of the current way of understanding **the constitutive structure of a DNA or RNA** sequence and the **relationship** between itself and other different and possible DNA or RNA sequences.
In particular, the concept of "**Non-Obvious Trend**" of a specific DNA or RNA sequence is explained and the mode through which it multiplies in numerous "**Non-Obvious Trends**".
The **original sequence** differs from **all new sequences generated** by at least **70% of bases** and therefore refers **to the very different biological material**.
This data is of fundamental importance because this paper states that, despite the considerable difference in bases between the **original sequence** and the **new sequences generated**, they have **common** "characteristics" precisely because their TRENDS are STRUCTURALLY IDENTICAL.
The chapter concludes with a specific proposal to apply the **binary code** to the reading of the **genetic code** through which the **complexity of the genetic code** increases considerably, thus ***highlighting currently unknown similarities and differences between sequences***, but also significantly increasing the body of information that could be obtained if we had a *database* that took into account **Trends** and everything related to them.

7.1 PARTIAL MODIFICATION OF AN ORIGINAL SEQUENCE AND GENERATION OF NEW SEQUENCES

As mentioned in the **Introduction**, the **complete procedure** for analyzing and modifying a DNA (or RNA) sequence involves **NINE PHASES**.
(http://www.ttesystems.eu/Cms/files/Pdf/Introduzione%20DNA%20ENG.pdf).
Due to the short time available for the drafting of this document and the huge amount of data presented, some *Phases* have been omitted and, among them, the **First Phase**, the **Analysis of the Original Sequence**.

This paragraph explains the procedure for the realization of the **Second Phase**, the **Modification of the Original Sequence**.

In this document, the **Original Sequence** object of an in-depth study is the *RdRp Gene (RdRp_SARSr-F) RNA sequence* **"GTGAAATGGTCATGTGTGGCGG"**.

After performing the **graphic analysis** of the **original** DNA or RNA **sequence**, the **T.T.E.S.** software offers also the possibility to modify the **original sequence** and to generate numerous and different **new** DNA or RNA **sequences** that faithfully respect the numerous and different "*non-obvious trends*" of the **original sequence**.
The "**non-obvious trends**" are *trends*, not clearly identifiable, present in all DNA and RNA sequences.

Generating many different **new** DNA or RNA **sequences** is possible for **two reasons**:

1) Firstly, because every specific DNA or RNA sequence (*original sequence*) can be «transformed» into many different *new sequences*, following the many different "non-obvious trends" of the specific *original sequence* (Fig.7);
2) Secondly, because every "non-obvious trend" of the *original sequence* can generate many different new sequences (Fig.7).

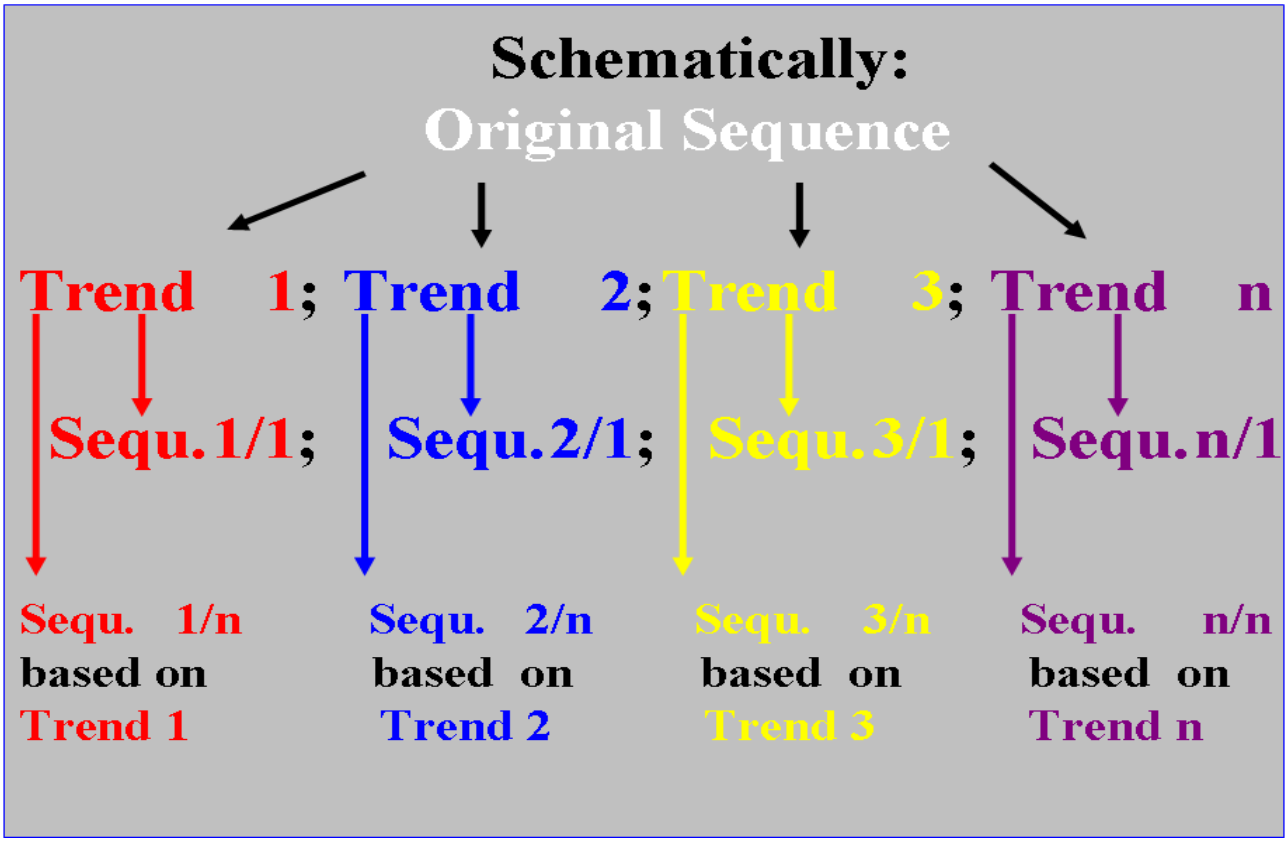


Fig. 7

Starting from **each** of the **possible "Non-Obvious Trend"** of an **original sequence**, it is possible to generate an indeterminate number of "new sequences" (as schematized in Fig. 7). Therefore, from **Trend n°1** it is possible to generate sequences n°1/1, 1/2, 1/3, etc. Similarly, from **Trend n°2** it is possible to generate sequences n°2/1, 2/2, 2/3, etc. and so on.
The identification of "**non-obvious trends**" is the fundamental prerequisite for DNA and RNA sequences modifications performed through the **T.T.E.S.** software.
For the **COMPLETE MODIFICATION OF AN ORIGINAL SEQUENCE** and the generation of **ALL NEW SEQUENCES** it is necessary to **identify all possible NON-OBVIOUS TRENDS** (beyond **Trend n°1** already identified, analyzed and graphically represented in the **first phase**) of the **original sequence**.

It is fundamental to underline that, usually, an **original sequence** differs from **all new sequences generated** by at least **70% of bases** (it is therefore a **very different biological material**).

7.2 THE TWENTY-FOUR "NON-OBVIOUS TRENDS"

The main "**NON-OBVIOUS TRENDS**" are **TWENTY-FOUR**.
They have been identified by attributing a number from **0** to **3** to the **four nitrogenous bases** (**T**, **C**, **G** and **A**) and calculating the **number** of **non-repeating Permutations** of **n** (**4**) **elements** according to the formula **P(k) = n!** [**P(4) = 4!**].
Below the 24 "**Non-Obvious Trends**" are presented.

Trend n° 1 (T=0; C=1; G=2; A=3)

Trend n° 2 (T=3; C=2; G=1; A=0)

Trend n° 3 (T=2; C=0; G=3; A=1) = Trend n° 4 (T=1; C=3; G=0; A=2)

Trend n° 5 (T=3; C=1; G=2; A=0) = Trend n° 22 (T=0; C=2; G=1; A=3)

Trend n° 6 (T=2; C=3; G=0; A=1)

Trend n° 7 (T=2; C=1; G=0; A=3)

Trend n° 8 (T=1; C=0; G=3; A=2)

Trend n° 9 (T=2; C=0; G=1; A=3)

Trend n° 10 (T=2; C=1; G=3; A=0)

Trend n° 11 (T=3; C=0; G=1; A=2) = Trend n° 20 (T=0; C=3; G=2; A=1)

Trend n° 12 (T=1; C=2; G=0; A=3)

Trend n° 13 (T=3; C=0; G=2; A=1) = Trend n° 21 (T=0; C=3; G=1; A=2)

Trend n° 14 (T=1; C=3; G=2; A=0)

Trend n° 15 (T=3; C=2; G=0; A=1)

Trend n° 16 (T=3; C=1; G=0; A=2) = Trend n° 23 (T=0; C=2; G=3; A=1)

Trend n° 17 (T=2; C=3; G=1; A=0) = Trend n° 24 (T=1; C=0; G=2; A=3)

Trend n° 18 (T=0; C=1; G=3; A=2)

Trend n° 19 (T=1; C=2; G=3; A=0)

ATTENTION:
As shown above and seen in the next chapter, some "**Non-Obvious Trends**", even if they correspond to **different permutations**, during the **modifications of the original sequence**, will generate the **same new sequences**.

7.3 PROCEDURE FOR PARTIAL MODIFICATION OF THE ORIGINAL SEQUENCE

Paragraph 7.1 states that for the COMPLETE **MODIFICATION OF THE ORIGINAL SEQUENCE** and the generation of **ALL NEW SEQUENCES** it is necessary **to identify all possible NON-OBVIOUS TRENDS** of the **original sequence**.
In this example, we'll consider **only** one of the possible **Non-Obvious Trends**. For this reason, the modification of the original sequence is called "**partial**".

Suppose you want to edit a **sequence of 22 bases**.
We define this sequence as "**Original Sequence X**".

Original Sequence X	**T**	**A**	**T**	**C**	**C**	**C**	**A**	**T**	**T**	**A**	**C**	**A**	**T**	**G**	**A**	**T**	**A**	**T**	**T**	**A**	**T**	**T**

Suppose you want to change the **Original Sequence X** to **Trend n°3**, one of the **24 Non-Obvious Trends** identified in the previous paragraph.

Modification of the "Original Sequence X" following Trend n° 3 (T=2; C=0; G=3; A=1)

Original Sequence X	**T**	**A**	**T**	**C**	**C**	**C**	**A**	**T**	**T**	**A**	**C**	**A**	**T**	**G**	**A**	**T**	**A**	**T**	**T**	**A**	**T**	**T**
Trend n° 3 (T=2; C=0; G=3; A=1)	**2**	**1**	**2**	**0**	**0**	**0**	**1**	**2**	**2**	**1**	**0**	**1**	**2**	**3**	**1**	**2**	**1**	**2**	**2**	**1**	**2**	**2**

As shown in the table above, the **first procedure** to perform is to **give each letter in the sequence the "specific number"** that has been assigned to it according to the selected *Non-Obvious Trends*.

Original Sequence X	**T**	**A**	**T**	**C**	**C**	**C**	**A**	**T**	**T**	**A**	**C**	**A**	**T**	**G**	**A**	**T**	**A**	**T**	**T**	**A**	**T**	**T**
Trend n° 3 (T=2; C=0; G=3; A=1)	**2**	**1**	**2**	**0**	**0**	**0**	**1**	**2**	**2**	**1**	**0**	**1**	**2**	**3**	**1**	**2**	**1**	**2**	**2**	**1**	**2**	**2**
Trend n° 3 (I i D d)	**-**	**D**	**I**	**D**	**D**	**d**	**I**	**I**	**i**	**D**	**D**	**I**	**I**	**I**	**D**	**I**	**D**	**I**	**i**	**D**	**I**	**i**

The **second procedure** to be performed is to **assign to each "specific number"** of the **sequence one of the following four letters**: "**I**", "**i**", "**D**" or "**d**".
The *four letters* are established by comparing a "**specific number**" with the "**specific number**" **preceding it in the sequence** and have the following meaning:

I = **Increase** (Increase over the previous **specific number** assigned)
i = **repeated increase** (repeated increase over the previous **specific number** assigned)
D = **Decrease** (Decrease over the previous **specific number** assigned)
d = **repeated decrease** (repeated decrease over the previous **specific number** assigned)

The **repeated increase** ("**i**") and **repeated decrease** ("**d**") definitions are used when, **after an Increment or after a Decrease**, the "**specific number**" is **repeated**, **equal to the**

previous one (as shown in the previous table and those shown in the following pages).
These four letters ("**I**", "**i**", "**D**" or "**d**") make **Obvious** the **Non-Obvious Trend associated with the four** "**specific numbers**" that have been assigned as determined by the selected *Non-Obvious Trend*.

Original Sequence X	**T**	**A**	**T**	**C**	**C**	**C**	**A**	**T**	**T**	**A**	**C**	**A**	**T**	**G**	**A**	**T**	**A**	**T**	**T**	**A**	**T**	**T**
Trend n° 3 (T=2; C=0; G=3; A=1)	**2**	**1**	**2**	**0**	**0**	**0**	**1**	**2**	**2**	**1**	**0**	**1**	**2**	**3**	**1**	**2**	**1**	**2**	**2**	**1**	**2**	**2**
Trend n° 3 (I i D d)	**-**	**D**	**I**	**D**	**D**	**d**	**I**	**I**	**i**	**D**	**D**	**I**	**I**	**I**	**D**	**I**	**D**	**I**	**i**	**D**	**I**	**i**
Trend n° 3/1	**3**	**2**	**3**	**1**	**1**	**1**	**2**	**3**	**3**	**2**	**0**	**1**	**2**	**3**	**2**	**3**	**2**	**3**	**3**	**0**	**3**	**3**

The **third procedure** to be performed is to create a **new sequence of "specific numbers"** different from the previous one, but absolutely reflective of **Trend n° 3**.
To achieve this result, it is sufficient **to replace** the "**specific numbers**" of **Trend n° 3** (those in the second row in the table shown above) with **different "specific numbers"** that respect **Trend n° 3 (I i D d)**, i. e. numbers whose **Increment** or **Decrease** strictly adheres to the sequence of "**I**", "**i**", "**D**" and "**d**" created during the **second procedure** (see tables above and below).

Original Sequence X	**T**	**A**	**T**	**C**	**C**	**C**	**A**	**T**	**T**	**A**	**C**	**A**	**T**	**G**	**A**	**T**	**A**	**T**	**T**	**A**	**T**	**T**
Trend n° 3 (T=2; C=0; G=3; A=1)	**2**	**1**	**2**	**0**	**0**	**0**	**1**	**2**	**2**	**1**	**0**	**1**	**2**	**3**	**1**	**2**	**1**	**2**	**2**	**1**	**2**	**2**
Trend n° 3 (I i D d)	**-**	**D**	**I**	**D**	**D**	**d**	**I**	**I**	**i**	**D**	**D**	**I**	**I**	**I**	**D**	**I**	**D**	**I**	**i**	**D**	**I**	**i**
Trend n° 3/1	**3**	**2**	**3**	**1**	**1**	**1**	**2**	**3**	**3**	**2**	**0**	**1**	**2**	**3**	**2**	**3**	**2**	**3**	**3**	**0**	**3**	**3**
Sequence n° 3/1 (81,82%)	**G**	**T**	**G**	**A**	**A**	**A**	**T**	**G**	**G**	**T**	**C**	**A**	**T**	**G**	**T**	**G**	**T**	**G**	**G**	**C**	**G**	**G**

The **fourth procedure** to perform is to generate the **new sequence** (sequence of letters coloured in red in the table shown above), **Sequence n° 3/1**.
To achieve this result, it is sufficient **to replace** the sequence of **different** "**specific numbers**" (those of the fourth row in the table shown above and corresponding to **Trend n° 3/1**) the corresponding letters attributed as determined by the selected *Non-Obvious Trend* (**Trend n° 3**), i. e. **T=2**, **C=0**, **G=3** e **A=1**.

In summary:

Original Sequence X of **22 bases**:

Original Sequence X	**T**	**A**	**T**	**C**	**C**	**C**	**A**	**T**	**T**	**A**	**C**	**A**	**T**	**G**	**A**	**T**	**A**	**T**	**T**	**A**	**T**	**T**

Sequence n° 3/1 of **22 bases**
generated by the modification of the **Original Sequence X**:

Sequence n° 3/1 (81,82%)	**G**	**T**	**G**	**A**	**A**	**A**	**T**	**G**	**G**	**T**	**C**	**A**	**T**	**G**	**T**	**G**	**T**	**G**	**G**	**C**	**G**	**G**

Sequence n° 3/1, the **new sequence generated** following **Trend n° 3/1** (the Trend that was generated faithfully following **Trend n° 3/1**), has **81.82%** bases (letters) different from the **Original Sequence X**.
Despite the considerable difference in bases between the two sequences (the **original sequence** and the **new sequence**), they have common "characteristics" precisely because their TRENDS are STRUCTURALLY IDENTICAL.

As we will see in the following chapters [and also in DNA or RNA Sequence Analysis and Modification through T_T_E_S_ (Chapter I° - First Part) and DNA or RNA Sequence Analysis and Modification through T_T_E_S_ (Chapter I° - Second Part)], some of the common "characteristics" (highlighted by the results of significant alignments of BLAST searches) between the **original sequence** and the **new sequences** generated, concern **common "organisms"**, **common amino acids** (although the bases are different) and sometimes "**products**" (proteins or other) whose functions can theoretically be correlated among the different sequences.

Returning to the example of the modified sequence in this paragraph, you can also perform the **REVERSE PROCESS**, that is, starting from **Sequence n° 3/1** of **22 bases** generated above, which we now call "**Original Sequence Y**", it is possible to generate the **Original Sequence X** of **22 bases**, which will constitute the new and different **Sequence n° 3/1**.

Original Sequence Y	**G**	**T**	**G**	**A**	**A**	**A**	**T**	**G**	**G**	**T**	**C**	**A**	**T**	**G**	**T**	**G**	**T**	**G**	**G**	**C**	**G**	**G**

Modification of the "Original Sequence Y"
following Trend n° 3 (T=2; C=0; G=3; A=1)

Original Sequence Y	**G**	**T**	**G**	**A**	**A**	**A**	**T**	**G**	**G**	**T**	**C**	**A**	**T**	**G**	**T**	**G**	**T**	**G**	**G**	**C**	**G**	**G**
Trend n° 3 (T=2; C=0; G=3; A=1)	**3**	**2**	**3**	**1**	**1**	**1**	**2**	**3**	**3**	**2**	**0**	**1**	**2**	**3**	**2**	**3**	**2**	**3**	**3**	**0**	**3**	**3**

Original Sequence Y	**G**	**T**	**G**	**A**	**A**	**A**	**T**	**G**	**G**	**T**	**C**	**A**	**T**	**G**	**T**	**G**	**T**	**G**	**G**	**C**	**G**	**G**
Trend n° 3 (T=2; C=0; G=3; A=1)	**3**	**2**	**3**	**1**	**1**	**1**	**2**	**3**	**3**	**2**	**0**	**1**	**2**	**3**	**2**	**3**	**2**	**3**	**3**	**0**	**3**	**3**
Trend n° 3 (I i D d)	**-**	**D**	**I**	**D**	**D**	**d**	**I**	**I**	**i**	**D**	**D**	**I**	**I**	**I**	**D**	**I**	**D**	**I**	**i**	**D**	**I**	**i**

Original Sequence Y	**G**	**T**	**G**	**A**	**A**	**A**	**T**	**G**	**G**	**T**	**C**	**A**	**T**	**G**	**T**	**G**	**T**	**G**	**G**	**C**	**G**	**G**
Trend n° 3 (T=2; C=0; G=3; A=1)	**3**	**2**	**3**	**1**	**1**	**1**	**2**	**3**	**3**	**2**	**0**	**1**	**2**	**3**	**2**	**3**	**2**	**3**	**3**	**0**	**3**	**3**
Trend n° 3 (I i D d)	**-**	**D**	**I**	**D**	**D**	**d**	**I**	**I**	**i**	**D**	**D**	**I**	**I**	**I**	**D**	**I**	**D**	**I**	**i**	**D**	**I**	**i**
Trend n° 3/1	**2**	**1**	**2**	**0**	**0**	**0**	**1**	**2**	**2**	**1**	**0**	**1**	**2**	**3**	**1**	**2**	**1**	**2**	**2**	**1**	**2**	**2**

Original Sequence Y	**G**	**T**	**G**	**A**	**A**	**A**	**T**	**G**	**G**	**T**	**C**	**A**	**T**	**G**	**T**	**G**	**T**	**G**	**G**	**C**	**G**	**G**
Trend n° 3 (T=2; C=0; G=3; A=1)	**3**	**2**	**3**	**1**	**1**	**1**	**2**	**3**	**3**	**2**	**0**	**1**	**2**	**3**	**2**	**3**	**2**	**3**	**3**	**0**	**3**	**3**
Trend n° 3 (I i D d)	**-**	**D**	**I**	**D**	**D**	**d**	**I**	**I**	**i**	**D**	**D**	**I**	**I**	**I**	**D**	**I**	**D**	**I**	**i**	**D**	**I**	**i**
Trend n° 3/1	**2**	**1**	**2**	**0**	**0**	**0**	**1**	**2**	**2**	**1**	**0**	**1**	**2**	**3**	**1**	**2**	**1**	**2**	**2**	**1**	**2**	**2**
Sequenza n° 3/1 (81,82%)	**T**	**A**	**T**	**C**	**C**	**C**	**A**	**T**	**T**	**A**	**C**	**A**	**T**	**G**	**A**	**T**	**A**	**T**	**T**	**A**	**T**	**T**

In summary:

Original Sequence Y of **22 bases**:

Original Sequence Y	**G**	**T**	**G**	**A**	**A**	**A**	**T**	**G**	**G**	**T**	**C**	**A**	**T**	**G**	**T**	**G**	**T**	**G**	**G**	**C**	**G**	**G**

Sequence n° 3/1 of **22 bases** generated by the modification of the **Original Sequence Y**:

Sequence n° 3/1 (81,82%)	**T**	**A**	**T**	**C**	**C**	**C**	**A**	**T**	**T**	**A**	**C**	**A**	**T**	**G**	**A**	**T**	**A**	**T**	**T**	**A**	**T**	**T**

Again, **Sequence n° 3/1**, the **new sequence generated** following **Trend n° 3/1** (Trend that was generated faithfully following **Trend n° 3**), has **81.82%** of bases (letters) different from the **Original Sequence Y**, but the **original sequence** and the **new sequence**, despite the differences in bases, have common "characteristics" precisely because their TRENDS are STRUCTURALLY IDENTICAL.

Now let's pick up the last table shown on the previous page.

Original Sequence Y	**G**	**T**	**G**	**A**	**A**	**A**	**T**	**G**	**G**	**T**	**C**	**A**	**T**	**G**	**T**	**G**	**T**	**G**	**G**	**C**	**G**	**G**
Trend n° 3 (T=2; C=0; G=3; A=1)	**3**	**2**	**3**	**1**	**1**	**1**	**2**	**3**	**3**	**2**	**0**	**1**	**2**	**3**	**2**	**3**	**2**	**3**	**3**	**0**	**3**	**3**
Trend n° 3 (I i D d)	**-**	**D**	**I**	**D**	**D**	**d**	**I**	**I**	**i**	**D**	**D**	**I**	**I**	**I**	**D**	**I**	**D**	**I**	**i**	**D**	**I**	**i**
Trend n° 3/1	**2**	**1**	**2**	**0**	**0**	**0**	**1**	**2**	**2**	**1**	**0**	**1**	**2**	**3**	**1**	**2**	**1**	**2**	**2**	**1**	**2**	**2**
Sequence n° 3/1 (81,82%)	**T**	**A**	**T**	**C**	**C**	**C**	**A**	**T**	**T**	**A**	**C**	**A**	**T**	**G**	**A**	**T**	**A**	**T**	**T**	**A**	**T**	**T**

I would like to draw attention to the **EQUAL letters and numbers highlighted in gray** in both sequences.
They are **highlighted in gray** to indicate that, in those **specific locations**, the letters of the **Original Sequence** have a **Trend** that CANNOT BE CHANGED.

As we can see, the letters "**CATG**" correspond to the numbers "**0123**".
In the example shown, the letters in the **Original Sequence** correspond to **progressively increasing numbers** (or may even be repeated *increasing numbers*, such as "0112223", which would be matched in the example given by the hypothetical letters "CAATTTG") of a unit, starting with the **smaller number** (i.e., from **0**) to the **largest number** (i.e., to **3**).

Under these **specific conditions**, COMPARED TO THE ORIGINAL SEQUENCES, LETTER MODIFICATIONS IN NEW SEQUENCES ARE **NEVER** ALLOWED.
The same applies if the letters in the **Original Sequence** corresponded to **progressively decreasing numbers** (i. e. "3201", or repeated decreasing numbers, such as "3221110", to which the hypothetical letters "GTTAAAC" would correspond) of a unit, starting from the **largest number** (i.e., from **3**) to the smaller number (i.e., the **0**).

It is plausible to hypothesize that **such particular groupings of letters** may assume an important biological significance as a signal or perhaps also because they may be particularly resistant to accidental modification during **transcription** processes.

The **Original Sequence Y**,

Original Sequence Y **G T G A A A T G G T C A T G T G T G G C G G**

corresponds to the now well-known *RdRp Gene RNA sequence (RdRp_SARSr-F)* found in SARS-CoV-2 (see the sequence delimited by black side lines and red top lines in the graph shown below in Fig. 7.2).
The green arrows in Fig. 7.2 indicate respectively the amino acids **Methionine** (Met/**M**, **ATG** codon) and **Valine** (**V**al/V, **GTC** codon), while the blue arrow indicates the amino acid **Methionine** (**ATG** codon).
Also, in Fig. 7.2, the letters delimited by the blue frame (those placed at the top) refer to the **particular grouping of letters** ("**CATG**") discussed on the previous page. Three of these letters encode the amino acid **M**, while the letter "C" corresponds to the **last** letter of amino acid **V preceding** the amino acid **M** indicated by the blue arrow.

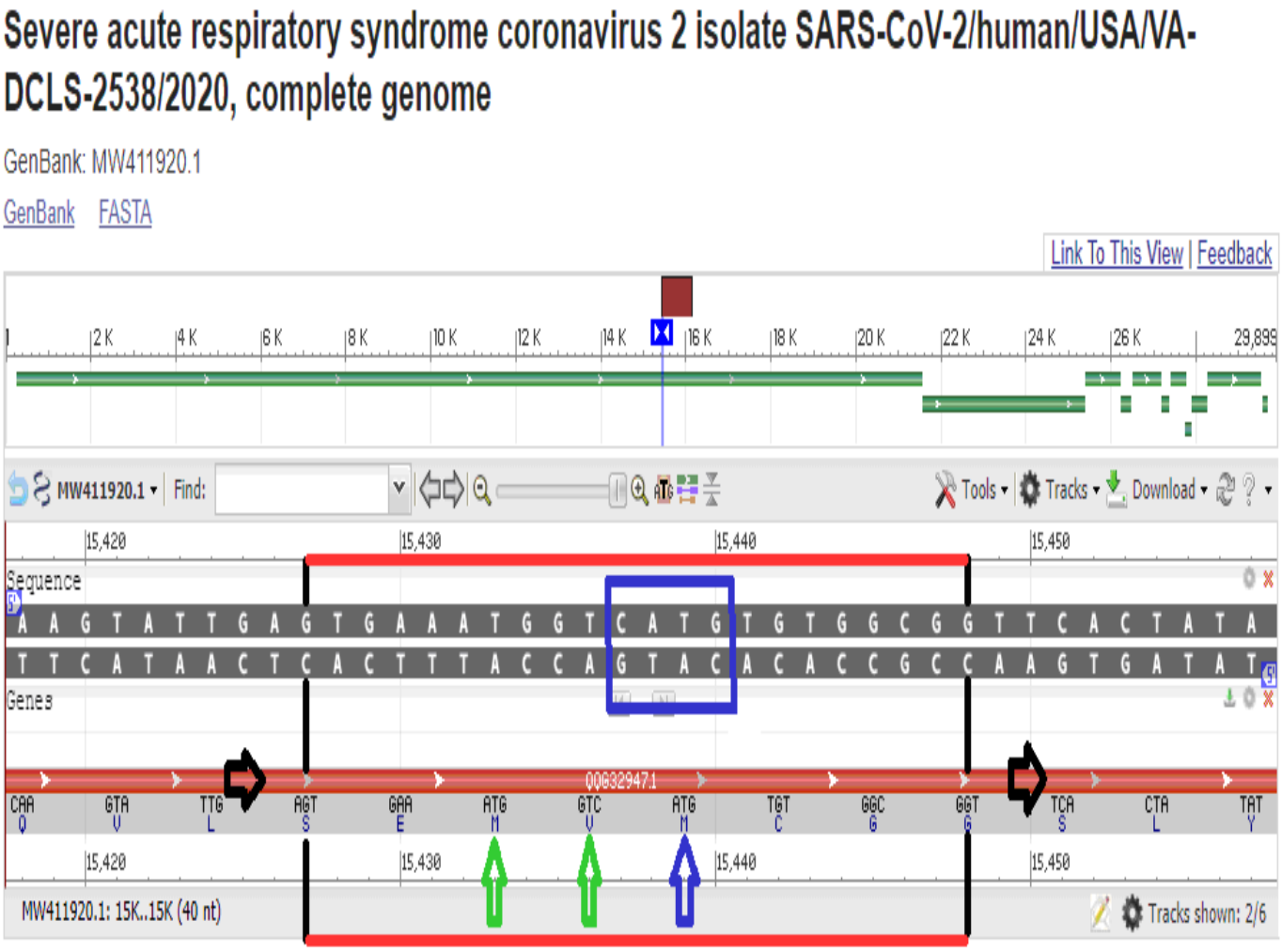


Fig. 7.2

As we shall see in the next chapter, **Sequence n° 3/1** of the last example shown (defined as The **Original Sequence X** in the example shown at the beginning of this paragraph),

Sequence n° 3/1 (81,82%) **T A T C C C A T T A C A T G A T A T T A T T**

really corresponds to **Sequence n° 3/1** (see the sequence delimited by red black and upper side lines in the graph shown below in Fig. 7.3), a **new sequence generated** (following **Trend n° 3/1**) starting with the *RdRp Gene RNA sequence (RdRp_SARSr-F)*, the **Original Sequence Y** above.

As can be seen in Fig. 7.3, among the "**organisms**" identified by the significant alignments (of the BLAST search) of this sequence is the ***Bacillus Anthracis*** (remember what has been said in the Foreword and see what has been written in Chapters X and XI, in General Discussion and Conclusions).
Interestingly, as shown in Fig. 7.2, the green arrows shown in Fig. 7.3 indicate respectively the amino acids **Methionine** (Met/**M**, **ATG** codon) and **Valine** (Val/**V**, **GTA** codon), while the blue arrow indicates a different amino acid, **Histidine** (His/**H**, **CAT** codon).

Incidentally, I would remind you that, as explained in section 2.1 of Chapter 2 (and to which I refer for further information), to decide the **sense of the codon reading** it is necessary to refer to the sense of reading established by the direction indicated by the **black Arrows** shown in Figs 7.2 and 7.3.
Thus, while in Fig. 7.2 the **codons** will be read in the direction 5'→ 3' (from left to right) in "**sequence**" (sequence at the top), in Fig.7.3 the **codons** will be read in the direction 5'→ 3' (from right to left) in "**genes**" (sequence below).

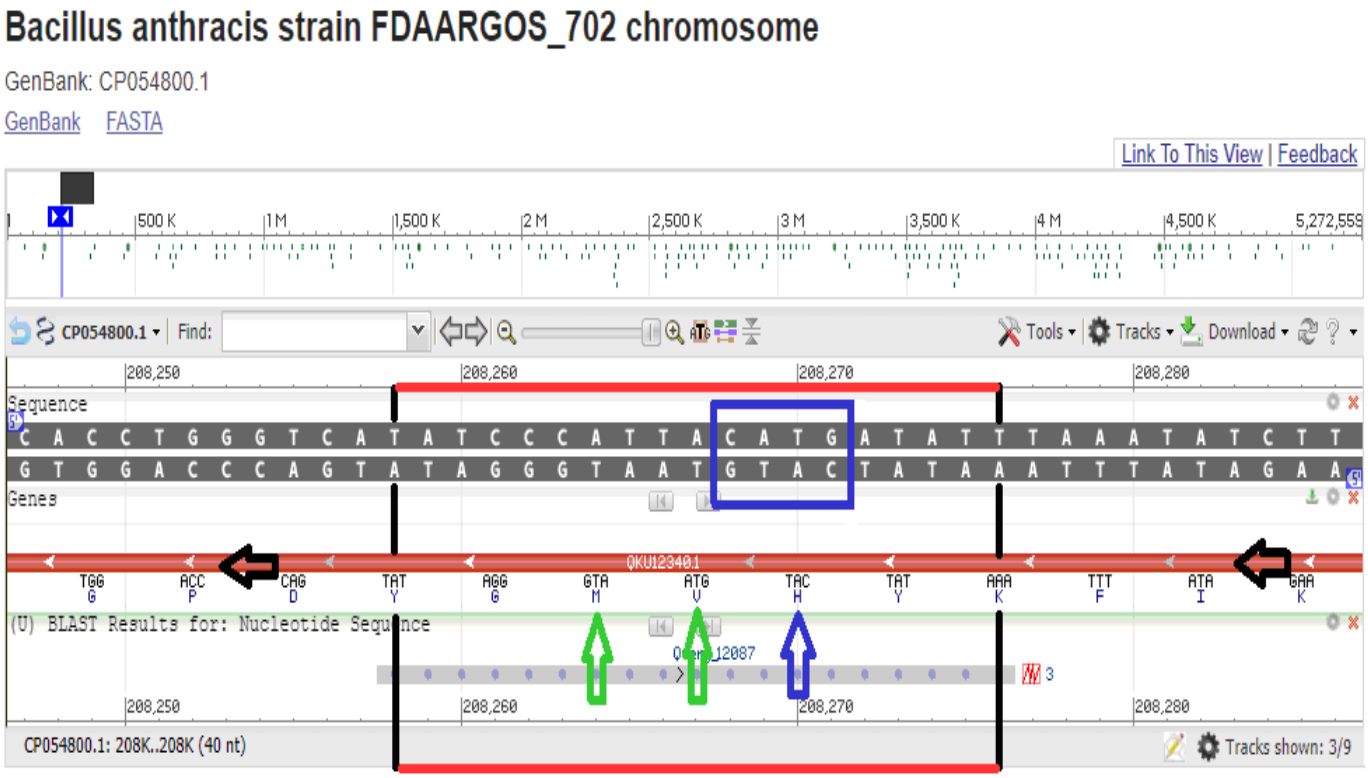


Fig. 7.3

Also in Fig. 7.3, the letters delimited by the blue frame (those at the top) refer to the **particular grouping of letters** ("**CATG**") already widely discussed in the previous pages. Three of these letters ("**ATG**") correspond to the three mirror letters "**TAC**" which, read from right to left, encode the amino acid **H** (differently from what showed in Fig. 7.2 in which the amino acid in question is **M**), while the letter "C" corresponds to the **first** mirror letter (the letter "**G**") of the amino acid **V** ("**G**TA") that **follows** the amino acid **H** indicated by the blue arrow.

Before concluding the chapter, let's take up again the last table shown on page 142, that of the **Modification of the "Original Sequence Y"** following **Trend n° 3 (T=2; C=0; G=3; A=1)**.

Original Sequence Y	**G**	**T**	**G**	**A**	**A**	**A**	**T**	**G**	**G**	**T**	**C**	**A**	**T**	**G**	**T**	**G**	**T**	**G**	**G**	**C**	**G**	**G**
Trend n° 3 (T=2; C=0; G=3; A=1)	**3**	**2**	**3**	**1**	**1**	**1**	**2**	**3**	**3**	**2**	**0**	**1**	**2**	**3**	**2**	**3**	**2**	**3**	**3**	**0**	**3**	**3**
Trend n° 3 (I i D d)	**-**	**D**	**I**	**D**	**d**	**d**	**I**	**I**	**i**	**D**	**D**	**I**	**I**	**I**	**D**	**I**	**D**	**I**	**i**	**D**	**I**	**i**
Trend n° 3/1	**2**	**1**	**2**	**0**	**0**	**0**	**1**	**2**	**2**	**1**	**0**	**1**	**2**	**3**	**1**	**2**	**1**	**2**	**2**	**1**	**2**	**2**
Sequence n° 3/1 (81,82%)	**T**	**A**	**T**	**C**	**C**	**C**	**A**	**T**	**T**	**A**	**C**	**A**	**T**	**G**	**A**	**T**	**A**	**T**	**T**	**A**	**T**	**T**

As mentioned in section 1.1 of Chapter I, in **bioinformatics** genetic information is treated in a **binary way**, i. e. as *a series of 0 and 1*, and transformed into **algorithms**.

If we wanted to apply the **binary code** and use **a series of 0 and 1** instead of the numbers **0, 1, 2** and **3** in the **first** and **third procedure** of **modifying an original sequence** (as it was shown in the previous pages), simply assign the *binary code* "**00**" to the **0**, "**01**" to the number **1**, "**10**" to the number **2** and finally, the binary code "**11**" at number **3**.

Following the application of the **binary code**, **Trend n° 3** would become

(T=10; C=00; G=11; A=01).

Original Sequence Y	**G**	**T**	**G**	**A**	**A**	**A**	**T**	**G**	**G**	**T**	**C**	**A**	**T**	**G**
Trend n° 3 (T=10; C=00; G=11; A=01)	**11**	**10**	**11**	**01**	**01**	**01**	**10**	**11**	**11**	**10**	**00**	**01**	**10**	**11**
Trend n° 3 (I i D d)	**-**	**D**	**I**	**D**	**d**	**d**	**I**	**I**	**I**	**D**	**D**	**I**	**I**	**I**
Trend n° 3/1	**10**	**01**	**10**	**00**	**00**	**00**	**01**	**10**	**10**	**01**	**00**	**01**	**10**	**11**
Sequence n° 3/1 (81,82%)	**T**	**A**	**T**	**C**	**C**	**C**	**A**	**T**	**T**	**A**	**C**	**A**	**T**	**G**

With this new codification, *amino acids* will be defined by **six numbers**.

Thus, in **Sequence n° 3/1**, at the "**ATG**" codon (highlighted in dark gray in the table shown above), mirrored "**TAC**" (fig. 7.3, letters delimited by the blue frame, those at the bottom) and coding "**CAT**" (right-to-left bed codon in fig. 7.3), will correspond to the numbers "**000110**" (indeed, for **Trend n° 3**, **C=00**, **A=01** and **T=10**) associated with the amino acid **H** (indicated by the blue arrow in Fig. 7.3 on the previous page).

ATTENTION: if, differently from what showed above, we had taken into account the same **original sequence** (the **Original Sequence Y**), but used for modifying **Trend n° 1 (T=00; C=01; G=10; A=11)** instead of Trend n° 3 (T=10; C=00; G=11; A=01), any "**CAT**" codon would have corresponded to **6** different numbers, i.e. "**011100**" rather than the numbers "000110".
The 6 numbers "**011100**" would always have been associated with the amino acid **H**, but they would have had a different meaning, because they would have been associated with a **sequence** (**sequence n° 1/1**) that must have had "characteristics" in **common** with the **original sequence** (because **Trend n°1** of the **original sequence** and **Trend n°1/1** of **sequence n° 1/1** are STRUCTURALLY IDENTICAL) and also with the **new sequence generated** following **Trend n° 3/1** (simply because it had been generated from the **original sequence**), but also "characteristics" **not in common** with the latter sequence, because it would have been generated following **Trend n° 3/1**, rather than **Trend n°1/1**.

From this simple example, we understand how, with a codification like the one proposed, the **complexity of the genetic code** increases considerably, thus ***highlighting currently unknown similarities and differences between sequences***, and also how the number of information that could be obtained would grow if we had a *database* that took into account **Trends** and everything related to them.

CHAPTER VIII

MODIFICATIONS OF THE RNA SEQUENCE “GTGAAATGGTCATGTGTGGCGG” OF THE RdRp GENE (RdRp_SARSr-F) AND GENERATION OF NEW SEQUENCES

GOALS OF CHAPTER VIII

This chapter is dedicated to **modifications applied** to the *RNA* sequence **"GTGAAATGGTCATGTGGCGG"** of the *RdRp Gene (RdRp_SARSr-F)*, one of the RNA sequences used as positive control of the rhinopharyngeal swab, along with other sequences of SARS-CoV-2 genes.
The goal of these modifications to the aforementioned RNA sequence (here called the **original sequence**) is to generate a number of **new sequences** that have at least **70%** different bases from the **original sequence**.
To achieve this result, the procedure explained in the previous chapter is applied based on the identification of the different NON- OBVIOUS TRENDS of the **original sequence**.
Despite the considerable amount of different bases, the **new sequences have functional "characteristics" common to the original sequence**.
This evidence can be appropriately used to obtain biologically relevant results from the pharmaceutical, biochemical, biotechnological, etc. industries.
Specifically, the **modifications of the RNA sequence "GTGAAATGGTCATGTGGCGG" of the *RdRp Gene (RdRp_SARSr-F)*** and **the methodology proposed in this document open up different perspectives to start targeted studies and innovative treatments of the Covid 19 pandemic.**

8.1 MODIFICATIONS OF THE RNA SEQUENCE "*GTGAAATGGTCATGTGTGGCGG*" OF THE RDRP GENE (RDRP_SARSR-F) AND GENERATION OF NEW SEQUENCES

The sequence "**GTGAAATGGTCATGTGTGGCGG**" [of the RdRp Gene (*RNA-dependent RNA polymerase*, RdRp_SARSr-F)] is part of "**ORF1ab**" (an open reading frame 1ab of SARS-CoV-2), one of the RNA sequences used as as a positive control in the nasophraryngeal swab, along with other sequences of SARS-CoV-2 genes.
Below are the tables with ONLY SOME the numerous modifications of this important sequence (see first paragraph 7.3 of Chapter 7).

ATTENTION:
Trend pairs whose application generates equal sequences (see paragraph 7.2 of Chapter 7) are highlighted in similar colors.

Original Sequence = GTGAAATGGTCATGTGTGGCGG

Modifications following the Trend n° 1 (T=0; C=1; G=2; A=3)

Original Sequence	G	T	G	A	A	A	T	G	G	T	C	A	T	G	T	G	T	G	G	C	G	G
Trend n° 1 (T=0; C=1; G=2; A=3)	2	0	2	3	3	3	0	2	2	0	1	3	0	2	0	2	0	2	2	1	2	2
Trend n° 1 (I i D d)	-	D	I	I	i	i	D	I	i	D	I	I	D	I	D	I	D	I	i	D	I	i
Trend n° 1/1	1	0	1	2	2	2	1	3	3	1	2	3	1	3	1	3	1	3	3	2	3	3
Sequence n° 1/1 (90,90%)	C	T	C	G	G	G	C	A	A	C	G	A	C	A	C	A	C	A	A	G	A	A

Original Sequence	G	T	G	A	A	A	T	G	G	T	C	A	T	G	T	G	T	G	G	C	G	G
Trend n° 1 (T=0; C=1; G=2; A=3)	2	0	2	3	3	3	0	2	2	0	1	3	0	2	0	2	0	2	2	1	2	2
Trend n° 1 (I i D d)	-	D	I	I	i	i	D	I	i	D	I	I	D	I	D	I	D	I	i	D	I	i
Trend n° 1/2	1	0	1	2	2	2	1	3	3	0	2	3	1	3	1	3	1	3	3	0	1	1
Sequence n° 1/2 (86,36%)	C	T	C	G	G	G	C	A	A	T	G	A	C	A	C	A	C	A	A	T	C	C

Original Sequence	G	T	G	A	A	A	T	G	G	T	C	A	T	G	T	G	T	G	G	C	G	G
Trend n° 1 (T=0; C=1; G=2; A=3)	2	0	2	3	3	3	0	2	2	0	1	3	0	2	0	2	0	2	2	1	2	2
Trend n° 1 (I i D d)	-	D	I	I	i	i	D	I	i	D	I	I	D	I	D	I	D	I	i	D	I	i
Trend n° 1/3	1	0	1	2	2	2	1	3	3	0	2	3	1	3	1	3	1	3	3	2	3	3
Sequence n° 1/3 (86,36%)	C	T	C	G	G	G	C	A	A	T	G	A	C	A	C	A	C	A	A	G	A	A

Modifications following the Trend n° 2 (T=3; C=2; G=1; A=0)

Original Sequence	G	T	G	A	A	A	T	G	G	T	C	A	T	G	T	G	T	G	G	C	G	G
Trend n° 2 (T=3; C=2; G=1; A=0)	1	3	1	0	0	0	3	1	1	3	2	0	3	1	3	1	3	1	1	2	1	1
Trend n° 2 (I i D d)	-	I	D	D	d	d	I	D	d	I	D	D	I	D	I	D	I	D	d	I	D	d
Trend n° 2/1	0	3	2	1	1	1	2	0	0	2	1	0	2	0	2	0	2	0	0	3	0	0
Sequence n° 2/1 (90,90%)	A	T	C	G	G	G	C	A	A	C	G	A	C	A	C	A	C	A	A	G	A	A

Original Sequence	**G**	**T**	**G**	**A**	**A**	**A**	**T**	**G**	**G**	**T**	**C**	**A**	**T**	**G**	**T**	**G**	**T**	**G**	**G**	**C**	**G**	**G**
Trend n° 2 (T=3; C=2; G=1; A=0)	**1**	**3**	**1**	**0**	**0**	**0**	**3**	**1**	**1**	**3**	**2**	**0**	**3**	**1**	**3**	**1**	**3**	**1**	**1**	**2**	**1**	**1**
Trend n° 2 (I i D d)	**-**	**I**	**D**	**D**	**d**	**d**	**I**	**D**	**d**	**I**	**D**	**D**	**I**	**D**	**I**	**D**	**I**	**D**	**d**	**I**	**D**	**d**
Trend n° 2/2	**0**	**3**	**2**	**1**	**1**	**1**	**3**	**0**	**0**	**3**	**1**	**0**	**2**	**0**	**2**	**0**	**2**	**0**	**0**	**3**	**0**	**0**
Sequence n° 2/2 (81,82%)	**A**	**T**	**C**	**G**	**G**	**G**	**T**	**C**	**C**	**T**	**G**	**A**	**C**	**A**	**C**	**A**	**C**	**A**	**A**	**G**	**A**	**A**

Modifications following the Trend n° 3 (T=2; C=0; G=3; A=1)

Original Sequence	**G**	**T**	**G**	**A**	**A**	**A**	**T**	**G**	**G**	**T**	**C**	**A**	**T**	**G**	**T**	**G**	**T**	**G**	**G**	**C**	**G**	**G**
Trend n° 3 (T=2; C=0; G=3; A=1)	**3**	**2**	**3**	**1**	**1**	**1**	**2**	**3**	**3**	**2**	**0**	**1**	**2**	**3**	**2**	**3**	**2**	**3**	**3**	**0**	**3**	**3**
Trend n° 3 (I i D d)	**-**	**D**	**I**	**D**	**d**	**d**	**I**	**I**	**i**	**D**	**D**	**I**	**I**	**I**	**D**	**I**	**D**	**I**	**i**	**D**	**I**	**i**
Trend n° 3/1	**2**	**1**	**2**	**0**	**0**	**0**	**1**	**2**	**2**	**1**	**0**	**1**	**2**	**3**	**1**	**2**	**1**	**2**	**2**	**1**	**2**	**2**
Sequence n° 3/1 (81,82%)	**T**	**A**	**T**	**C**	**C**	**C**	**A**	**T**	**T**	**A**	**C**	**A**	**T**	**G**	**A**	**T**	**A**	**T**	**T**	**A**	**T**	**T**

Original Sequence	**G**	**T**	**G**	**A**	**A**	**A**	**T**	**G**	**G**	**T**	**C**	**A**	**T**	**G**	**T**	**G**	**T**	**G**	**G**	**C**	**G**	**G**
Trend n° 3 (T=2; C=0; G=3; A=1)	**3**	**2**	**3**	**1**	**1**	**1**	**2**	**3**	**3**	**2**	**0**	**1**	**2**	**3**	**2**	**3**	**2**	**3**	**3**	**0**	**3**	**3**
Trend n° 3 (I i D d)	**-**	**D**	**I**	**D**	**d**	**d**	**I**	**I**	**i**	**D**	**D**	**I**	**I**	**I**	**D**	**I**	**D**	**I**	**i**	**D**	**I**	**i**
Trend n° 3/2	**2**	**1**	**2**	**0**	**0**	**0**	**1**	**2**	**2**	**1**	**0**	**1**	**2**	**3**	**0**	**2**	**0**	**1**	**1**	**0**	**1**	**1**
Sequence n° 3/2 (77,27%)	**T**	**A**	**T**	**C**	**C**	**C**	**A**	**T**	**T**	**A**	**C**	**A**	**T**	**G**	**C**	**T**	**C**	**A**	**A**	**C**	**A**	**A**

Original Sequence	**G**	**T**	**G**	**A**	**A**	**A**	**T**	**G**	**G**	**T**	**C**	**A**	**T**	**G**	**T**	**G**	**T**	**G**	**G**	**C**	**G**	**G**
Trend n° 3 (T=2; C=0; G=3; A=1)	**3**	**2**	**3**	**1**	**1**	**1**	**2**	**3**	**3**	**2**	**0**	**1**	**2**	**3**	**2**	**3**	**2**	**3**	**3**	**0**	**3**	**3**
Trend n° 3 (I i D d)	**-**	**D**	**I**	**D**	**d**	**d**	**I**	**I**	**i**	**D**	**D**	**I**	**I**	**I**	**D**	**I**	**D**	**I**	**i**	**D**	**I**	**i**
Trend n° 3/3	**1**	**0**	**2**	**0**	**0**	**0**	**1**	**2**	**2**	**1**	**0**	**1**	**2**	**3**	**0**	**2**	**1**	**2**	**2**	**0**	**1**	**1**
Sequence n° 3/3 (77,27%)	**A**	**C**	**T**	**C**	**C**	**C**	**A**	**T**	**T**	**A**	**C**	**A**	**T**	**G**	**C**	**T**	**A**	**T**	**T**	**C**	**A**	**A**

Original Sequence	**G**	**T**	**G**	**A**	**A**	**A**	**T**	**G**	**G**	**T**	**C**	**A**	**T**	**G**	**T**	**G**	**T**	**G**	**G**	**C**	**G**	**G**
Trend n° 3 (T=2; C=0; G=3; A=1)	**3**	**2**	**3**	**1**	**1**	**1**	**2**	**3**	**3**	**2**	**0**	**1**	**2**	**3**	**2**	**3**	**2**	**3**	**3**	**0**	**3**	**3**
Trend n° 3 (I i D d)	**-**	**D**	**I**	**D**	**d**	**d**	**I**	**I**	**i**	**D**	**D**	**I**	**I**	**I**	**D**	**I**	**D**	**I**	**i**	**D**	**I**	**i**
Trend n° 3/4	**1**	**0**	**1**	**0**	**0**	**0**	**1**	**2**	**2**	**1**	**0**	**1**	**2**	**3**	**0**	**3**	**0**	**1**	**1**	**0**	**1**	**1**
Sequence n° 3/4 (72,73%)	**A**	**C**	**A**	**C**	**C**	**C**	**A**	**T**	**T**	**A**	**C**	**A**	**T**	**G**	**C**	**G**	**C**	**A**	**A**	**C**	**A**	**A**

Original Sequence	**G**	**T**	**G**	**A**	**A**	**A**	**T**	**G**	**G**	**T**	**C**	**A**	**T**	**G**	**T**	**G**	**T**	**G**	**G**	**C**	**G**	**G**
Trend n° 3 (T=2; C=0; G=3; A=1)	**3**	**2**	**3**	**1**	**1**	**1**	**2**	**3**	**3**	**2**	**0**	**1**	**2**	**3**	**2**	**3**	**2**	**3**	**3**	**0**	**3**	**3**
Trend n° 3 (I i D d)	**-**	**D**	**I**	**D**	**d**	**d**	**I**	**I**	**i**	**D**	**D**	**I**	**I**	**I**	**D**	**I**	**D**	**I**	**i**	**D**	**I**	**i**
Trend n° 3/5	**2**	**0**	**2**	**0**	**0**	**0**	**1**	**2**	**2**	**1**	**0**	**1**	**2**	**3**	**1**	**3**	**1**	**2**	**2**	**0**	**1**	**1**
Sequence n° 3/5 (72,73%)	**T**	**C**	**T**	**C**	**C**	**C**	**A**	**T**	**T**	**A**	**C**	**A**	**T**	**G**	**A**	**G**	**A**	**T**	**T**	**C**	**A**	**A**

Original Sequence	**G**	**T**	**G**	**A**	**A**	**A**	**T**	**G**	**G**	**T**	**C**	**A**	**T**	**G**	**T**	**G**	**T**	**G**	**G**	**C**	**G**	**G**
Trend n° 3 (T=2; C=0; G=3; A=1)	**3**	**2**	**3**	**1**	**1**	**1**	**2**	**3**	**3**	**2**	**0**	**1**	**2**	**3**	**2**	**3**	**2**	**3**	**3**	**0**	**3**	**3**
Trend n° 3 (I i D d)	**-**	**D**	**I**	**D**	**d**	**d**	**I**	**I**	**i**	**D**	**D**	**I**	**I**	**I**	**D**	**I**	**D**	**I**	**i**	**D**	**I**	**i**
Trend n° 3/6	**3**	**2**	**3**	**2**	**2**	**2**	**1**	**2**	**2**	**1**	**0**	**1**	**2**	**3**	**0**	**2**	**0**	**3**	**3**	**0**	**1**	**1**
Sequence n° 3/6 (54,54%)	**G**	**T**	**G**	**T**	**T**	**T**	**A**	**T**	**T**	**A**	**C**	**A**	**T**	**G**	**C**	**T**	**C**	**G**	**G**	**C**	**A**	**A**

Modifications following the Trend n° 4 (T=1; C=3; G=0; A=2)

Original Sequence	**G**	**T**	**G**	**A**	**A**	**A**	**T**	**G**	**G**	**T**	**C**	**A**	**T**	**G**	**T**	**G**	**T**	**G**	**G**	**C**	**G**	**G**
Trend n° 4 (T=1; C=3; G=0; A=2)	**0**	**1**	**0**	**2**	**2**	**2**	**1**	**0**	**0**	**1**	**3**	**2**	**1**	**0**	**1**	**0**	**1**	**0**	**0**	**3**	**0**	**0**
Trend n° 4 (I i D d)	**-**	**I**	**D**	**I**	**i**	**i**	**D**	**D**	**d**	**I**	**I**	**D**	**D**	**D**	**I**	**D**	**I**	**D**	**d**	**I**	**D**	**d**
Trend n° 4/1	**1**	**2**	**1**	**3**	**3**	**3**	**2**	**1**	**1**	**2**	**3**	**2**	**1**	**0**	**2**	**1**	**2**	**1**	**1**	**2**	**1**	**1**
Sequence n° 4/1 (81,82%)	**T**	**A**	**T**	**C**	**C**	**C**	**A**	**T**	**T**	**A**	**C**	**A**	**T**	**G**	**A**	**T**	**A**	**T**	**T**	**A**	**T**	**T**

Modifications following the Trend n° 5 (T=3; C=1; G=2; A=0)

Original Sequence	**G**	**T**	**G**	**A**	**A**	**A**	**T**	**G**	**G**	**T**	**C**	**A**	**T**	**G**	**T**	**G**	**T**	**G**	**G**	**C**	**G**	**G**
Trend n° 5 (T=3; C=1; G=2; A=0)	**2**	**3**	**2**	**0**	**0**	**0**	**3**	**2**	**2**	**3**	**1**	**0**	**3**	**2**	**3**	**2**	**3**	**2**	**2**	**1**	**2**	**2**
Trend n° 5 (I i D d)	**-**	**I**	**D**	**D**	**d**	**d**	**I**	**D**	**d**	**I**	**D**	**D**	**I**	**D**	**I**	**D**	**I**	**D**	**d**	**D**	**I**	**i**
Trend n° 5/1	**1**	**3**	**2**	**1**	**1**	**1**	**2**	**1**	**1**	**2**	**1**	**0**	**1**	**0**	**1**	**0**	**2**	**1**	**1**	**0**	**1**	**1**
Sequence n° 5/1 (81,82%)	**C**	**T**	**G**	**C**	**C**	**C**	**G**	**C**	**C**	**G**	**C**	**A**	**C**	**A**	**C**	**A**	**G**	**C**	**C**	**A**	**C**	**C**

Original Sequence	**G**	**T**	**G**	**A**	**A**	**A**	**T**	**G**	**G**	**T**	**C**	**A**	**T**	**G**	**T**	**G**	**T**	**G**	**G**	**C**	**G**	**G**
Trend n° 5 (T=3; C=1; G=2; A=0)	**2**	**3**	**2**	**0**	**0**	**0**	**3**	**2**	**2**	**3**	**1**	**0**	**3**	**2**	**3**	**2**	**3**	**2**	**2**	**1**	**2**	**2**
Trend n° 5 (I i D d)	**-**	**I**	**D**	**D**	**d**	**d**	**I**	**D**	**d**	**I**	**D**	**D**	**I**	**D**	**I**	**D**	**I**	**D**	**d**	**D**	**I**	**i**
Trend n° 5/2	**1**	**2**	**1**	**0**	**0**	**0**	**2**	**1**	**1**	**3**	**2**	**1**	**2**	**1**	**2**	**1**	**2**	**1**	**1**	**0**	**1**	**1**
Sequence n° 5/2 (81,82%)	**C**	**G**	**C**	**A**	**A**	**A**	**G**	**C**	**C**	**T**	**G**	**C**	**G**	**C**	**G**	**C**	**G**	**C**	**C**	**A**	**C**	**C**

Modifications following the Trend n° 6 (T=2; C=3; G=0; A=1)

Original Sequence	**G**	**T**	**G**	**A**	**A**	**A**	**T**	**G**	**G**	**T**	**C**	**A**	**T**	**G**	**T**	**G**	**T**	**G**	**G**	**C**	**G**	**G**
Trend n° 6 (T=2; C=3; G=0; A=1)	**0**	**2**	**0**	**1**	**1**	**1**	**2**	**0**	**0**	**2**	**3**	**1**	**2**	**0**	**2**	**0**	**2**	**0**	**0**	**3**	**0**	**0**
Trend n° 6 (I i D d)	**-**	**I**	**D**	**I**	**i**	**i**	**I**	**D**	**d**	**I**	**I**	**D**	**I**	**D**	**I**	**D**	**I**	**D**	**d**	**I**	**D**	**d**
Trend n° 6/1	**1**	**3**	**1**	**2**	**2**	**2**	**3**	**1**	**1**	**2**	**3**	**0**	**1**	**0**	**3**	**1**	**3**	**1**	**1**	**2**	**1**	**1**
Sequence n° 6/1 (86,36%)	**A**	**C**	**A**	**T**	**T**	**T**	**C**	**A**	**A**	**T**	**C**	**G**	**A**	**G**	**C**	**A**	**C**	**A**	**A**	**T**	**A**	**A**

Original Sequence	**G**	**T**	**G**	**A**	**A**	**A**	**T**	**G**	**G**	**T**	**C**	**A**	**T**	**G**	**T**	**G**	**T**	**G**	**G**	**C**	**G**	**G**
Trend n° 6 (T=2; C=3; G=0; A=1)	**0**	**2**	**0**	**1**	**1**	**1**	**2**	**0**	**0**	**2**	**3**	**1**	**2**	**0**	**2**	**0**	**2**	**0**	**0**	**3**	**0**	**0**
Trend n° 6 (I i D d)	**-**	**I**	**D**	**I**	**i**	**i**	**I**	**D**	**d**	**I**	**I**	**D**	**I**	**D**	**I**	**D**	**I**	**D**	**d**	**I**	**D**	**d**
Trend n° 6/2	**0**	**3**	**1**	**2**	**2**	**2**	**3**	**0**	**0**	**1**	**2**	**0**	**3**	**1**	**3**	**1**	**3**	**1**	**1**	**2**	**1**	**1**
Sequence n° 6/2 (86,36%)	**G**	**C**	**A**	**T**	**T**	**T**	**C**	**G**	**G**	**A**	**T**	**G**	**C**	**A**	**C**	**A**	**C**	**A**	**A**	**T**	**A**	**A**

Modifications following the Trend n° 7 (T=2; C=1; G=0; A=3)

Original Sequence	**G**	**T**	**G**	**A**	**A**	**A**	**T**	**G**	**G**	**T**	**C**	**A**	**T**	**G**	**T**	**G**	**T**	**G**	**G**	**C**	**G**	**G**
Trend n° 7 (T=2; C=1; G=0; A=3)	**0**	**2**	**0**	**3**	**3**	**3**	**2**	**0**	**0**	**2**	**1**	**3**	**2**	**0**	**2**	**0**	**2**	**0**	**0**	**1**	**0**	**0**
Trend n° 7 (I i D d)	**-**	**I**	**D**	**I**	**i**	**i**	**D**	**D**	**d**	**I**	**D**	**I**	**D**	**D**	**I**	**D**	**I**	**D**	**d**	**I**	**D**	**d**
Trend n° 7/1	**1**	**3**	**1**	**3**	**3**	**3**	**1**	**0**	**0**	**1**	**0**	**2**	**1**	**0**	**3**	**1**	**3**	**1**	**1**	**2**	**1**	**1**
Sequence n° 7/1 (72,73%)	**C**	**A**	**C**	**A**	**A**	**A**	**C**	**G**	**G**	**C**	**G**	**T**	**C**	**G**	**A**	**C**	**A**	**C**	**C**	**T**	**C**	**C**

Original Sequence	G	T	G	A	A	A	T	G	G	T	C	A	T	G	T	G	T	G	G	C	G	G
Trend n° 7 (T=2; C=1; G=0; A=3)	0	2	0	3	3	3	2	0	0	2	1	3	2	0	2	0	2	0	0	1	0	0
Trend n° 7 (I i D d)	-	I	D	I	i	i	D	D	d	I	D	I	D	D	I	D	I	D	d	I	D	d
Trend n° 7/2	1	2	1	2	2	2	1	0	0	3	0	3	1	0	1	0	3	2	2	3	2	2
Sequence n° 7/2 (72,73%)	C	T	C	T	T	T	C	G	G	A	G	A	C	G	C	G	A	T	T	A	T	T

Original Sequence	G	T	G	A	A	A	T	G	G	T	C	A	T	G	T	G	T	G	G	C	G	G
Trend n° 7 (T=2; C=1; G=0; A=3)	0	2	0	3	3	3	2	0	0	2	1	3	2	0	2	0	2	0	0	1	0	0
Trend n° 7 (I i D d)	-	I	D	I	i	i	D	D	d	I	D	I	D	D	I	D	I	D	d	I	D	d
Trend n° 7/3	0	3	1	2	2	2	1	0	0	3	0	2	1	0	3	1	3	1	1	2	1	1
Sequence n° 7/3 (81,82%)	G	A	C	T	T	T	C	G	G	A	G	T	C	G	A	C	A	C	C	T	C	C

Modifications following the Trend n° 8 (T=1; C=0; G=3; A=2)

Original Sequence	G	T	G	A	A	A	T	G	G	T	C	A	T	G	T	G	T	G	G	C	G	G
Trend n° 8 (T=1; C=0; G=3; A=2)	3	1	3	2	2	2	1	3	3	1	0	2	1	3	1	3	1	3	3	0	3	3
Trend n° 8 (I i D d)	-	D	I	D	d	d	D	I	i	D	D	I	D	I	D	I	D	I	i	D	I	i
Trend n° 8/1	2	0	2	1	1	1	0	3	3	2	1	3	0	2	0	2	0	2	2	1	2	2
Sequence n° 8/1 (90,90%)	A	C	A	T	T	T	C	G	G	A	T	G	C	A	C	A	C	A	A	T	A	A

Modifications following the Trend n° 9 (T=2; C=0; G=1; A=3)

Original Sequence	G	T	G	A	A	A	T	G	G	T	C	A	T	G	T	G	T	G	G	C	G	G
Trend n° 9 (T=2; C=0; G=1; A=3)	1	2	1	3	3	3	2	1	1	2	0	3	2	1	2	1	2	1	1	0	1	1
Trend n° 9 (I i D d)	-	I	D	I	i	i	D	D	d	I	D	I	D	D	I	D	I	D	d	D	I	i
Trend n° 9/1	0	1	0	2	2	2	1	0	0	3	1	2	1	0	3	2	3	2	2	1	2	2
Sequence n° 9/1 (100%)	C	G	C	T	T	T	G	C	C	A	G	T	G	C	A	T	A	T	T	G	T	T

Original Sequence	G	T	G	A	A	A	T	G	G	T	C	A	T	G	T	G	T	G	G	C	G	G
Trend n° 9 (T=2; C=0; G=1; A=3)	1	2	1	3	3	3	2	1	1	2	0	3	2	1	2	1	2	1	1	0	1	1
Trend n° 9 (I i D d)	-	I	D	I	i	i	D	D	d	I	D	I	D	D	I	D	I	D	d	D	I	i
Trend n° 9/2	0	1	0	2	2	2	1	0	0	3	1	2	1	0	1	0	3	2	2	1	2	2
Sequence n° 9/2 (100%)	C	G	C	T	T	T	G	C	C	A	G	T	G	C	G	C	A	T	T	G	T	T

Modifications following the Trend n° 10 (T=2; C=1; G=3; A=0)

Original Sequence	G	T	G	A	A	A	T	G	G	T	C	A	T	G	T	G	T	G	G	C	G	G
Trend n° 10 (T=2; C=1; G=3; A=0)	3	2	3	0	0	0	2	3	3	2	1	0	2	3	2	3	2	3	3	1	3	3
Trend n° 10 (I i D d)	-	D	I	D	d	d	I	I	i	D	D	D	I	I	D	I	D	I	i	D	I	i
Trend n° 10/1	2	1	2	1	1	1	2	3	3	2	1	0	1	2	1	2	1	2	2	0	2	2
Sequence n° 10/1 (72,73%)	T	C	T	C	C	C	T	G	G	T	C	A	C	T	C	T	C	T	T	A	T	T

Modifications following the Trend n° 11 (T=3; C=0; G=1; A=2)

Original Sequence	**G**	**T**	**G**	**A**	**A**	**A**	**T**	**G**	**G**	**T**	**C**	**A**	**T**	**G**	**T**	**G**	**T**	**G**	**G**	**C**	**G**	**G**
Trend n° 11 (T=3; C=0; G=1; A=2)	**1**	**3**	**1**	**2**	**2**	**2**	**3**	**1**	**1**	**3**	**0**	**2**	**3**	**1**	**3**	**1**	**3**	**1**	**1**	**0**	**1**	**1**
Trend n° 11 (I i D d)	**-**	**I**	**D**	**I**	**i**	**i**	**I**	**D**	**d**	**I**	**D**	**I**	**I**	**D**	**I**	**D**	**I**	**D**	**d**	**D**	**I**	**i**
Trend n° 11/1	**0**	**2**	**0**	**1**	**1**	**1**	**2**	**0**	**0**	**2**	**0**	**1**	**2**	**0**	**2**	**0**	**3**	**2**	**2**	**1**	**2**	**2**
Sequence n° 11/1 (90,90%)	**C**	**A**	**C**	**G**	**G**	**G**	**A**	**C**	**C**	**A**	**C**	**G**	**A**	**C**	**A**	**C**	**T**	**A**	**A**	**G**	**A**	**A**

Modifications following the Trend n° 13 (T=3; C=0; G=2; A=1)

Original Sequence	**G**	**T**	**G**	**A**	**A**	**A**	**T**	**G**	**G**	**T**	**C**	**A**	**T**	**G**	**T**	**G**	**T**	**G**	**G**	**C**	**G**	**G**
Trend n° 13 (T=3; C=0; G=2; A=1)	**2**	**3**	**2**	**1**	**1**	**1**	**3**	**2**	**2**	**3**	**0**	**1**	**3**	**2**	**3**	**2**	**3**	**2**	**2**	**0**	**2**	**2**
Trend n° 13 (I i D d)	**-**	**I**	**D**	**D**	**d**	**d**	**I**	**D**	**d**	**I**	**D**	**I**	**I**	**D**	**I**	**D**	**I**	**D**	**d**	**D**	**I**	**i**
Trend n° 13/1	**1**	**2**	**1**	**0**	**0**	**0**	**2**	**1**	**1**	**2**	**1**	**2**	**3**	**1**	**2**	**1**	**2**	**1**	**1**	**0**	**3**	**3**
Sequence n° 13/1 (90,90%)	**A**	**G**	**A**	**C**	**C**	**C**	**G**	**A**	**A**	**G**	**A**	**G**	**T**	**A**	**G**	**A**	**G**	**A**	**A**	**C**	**T**	**T**

Modifications following the Trend n° 14 (T=1; C=3; G=2; A=0)

Original Sequence	**G**	**T**	**G**	**A**	**A**	**A**	**T**	**G**	**G**	**T**	**C**	**A**	**T**	**G**	**T**	**G**	**T**	**G**	**G**	**C**	**G**	**G**
Trend n° 14 (T=1; C=3; G=2; A=0)	**2**	**1**	**2**	**0**	**0**	**0**	**1**	**2**	**2**	**1**	**3**	**0**	**1**	**2**	**1**	**2**	**1**	**2**	**2**	**3**	**2**	**2**
Trend n° 14 (I i D d)	**-**	**D**	**I**	**D**	**d**	**d**	**I**	**I**	**i**	**D**	**I**	**D**	**I**	**I**	**D**	**I**	**D**	**I**	**i**	**I**	**D**	**d**
Trend n° 14/1	**1**	**0**	**3**	**1**	**1**	**1**	**2**	**3**	**3**	**0**	**2**	**1**	**2**	**3**	**2**	**3**	**0**	**1**	**1**	**2**	**1**	**1**
Sequence n° 14/1 (100%)	**T**	**A**	**C**	**T**	**T**	**T**	**G**	**C**	**C**	**A**	**G**	**T**	**G**	**C**	**G**	**C**	**A**	**T**	**T**	**G**	**T**	**T**

Modifications following the Trend n° 15 (T=3; C=2; G=0; A=1)

Original Sequence	**G**	**T**	**G**	**A**	**A**	**A**	**T**	**G**	**G**	**T**	**C**	**A**	**T**	**G**	**T**	**G**	**T**	**G**	**G**	**C**	**G**	**G**
Trend n° 15 (T=3; C=2; G=0 A=1)	**0**	**3**	**0**	**1**	**1**	**1**	**3**	**0**	**0**	**3**	**2**	**1**	**3**	**0**	**3**	**0**	**3**	**0**	**0**	**2**	**0**	**0**
Trend n° 15 (I i D d)	**-**	**I**	**D**	**I**	**i**	**i**	**I**	**D**	**d**	**I**	**D**	**D**	**I**	**D**	**I**	**D**	**I**	**D**	**d**	**I**	**D**	**d**
Trend n° 15/1	**0**	**2**	**1**	**2**	**2**	**2**	**3**	**1**	**1**	**2**	**1**	**0**	**2**	**1**	**2**	**1**	**2**	**1**	**1**	**2**	**1**	**1**
Sequence n° 15/1 (90,90%)	**G**	**C**	**A**	**C**	**C**	**C**	**T**	**A**	**A**	**C**	**A**	**G**	**C**	**A**	**C**	**A**	**C**	**A**	**A**	**T**	**A**	**A**

Modifications following the Trend n° 17 (T=2; C=3; G=1; A=0)

Original Sequence	**G**	**T**	**G**	**A**	**A**	**A**	**T**	**G**	**G**	**T**	**C**	**A**	**T**	**G**	**T**	**G**	**T**	**G**	**G**	**C**	**G**	**G**
Trend n° 17 (T=2; C=3; G=1; A=0)	**1**	**2**	**1**	**0**	**0**	**0**	**2**	**1**	**1**	**2**	**3**	**0**	**2**	**1**	**2**	**1**	**2**	**1**	**1**	**3**	**1**	**1**
Trend n° 17 (I i D d)	**-**	**I**	**D**	**D**	**d**	**d**	**I**	**D**	**d**	**I**	**I**	**D**	**I**	**D**	**I**	**D**	**I**	**D**	**d**	**I**	**D**	**d**
Trend n° 17/1	**0**	**3**	**2**	**1**	**1**	**1**	**3**	**0**	**0**	**1**	**2**	**1**	**3**	**2**	**3**	**2**	**3**	**0**	**0**	**2**	**0**	**0**
Sequence n° 17/1 (100%)	**A**	**C**	**T**	**G**	**G**	**G**	**C**	**A**	**A**	**G**	**T**	**G**	**C**	**T**	**C**	**T**	**C**	**A**	**A**	**T**	**A**	**A**

Modifications following the Trend n° 18 (T=0; C=1; G=3; A=2)

Original Sequence	**G**	**T**	**G**	**A**	**A**	**A**	**T**	**G**	**G**	**T**	**C**	**A**	**T**	**G**	**T**	**G**	**T**	**G**	**G**	**C**	**G**	**G**
Trend n° 18 (T=0; C=1; G=3; A=2)	**3**	**0**	**3**	**2**	**2**	**2**	**0**	**3**	**3**	**0**	**1**	**2**	**0**	**3**	**0**	**3**	**0**	**3**	**3**	**1**	**3**	**3**
Trend n° 18 (I i D d)	**-**	**D**	**I**	**D**	**d**	**d**	**D**	**I**	**i**	**D**	**I**	**I**	**D**	**I**	**D**	**I**	**D**	**I**	**i**	**D**	**I**	**i**
Trend n° 18/1	**2**	**1**	**2**	**1**	**1**	**1**	**0**	**2**	**2**	**1**	**2**	**3**	**1**	**2**	**1**	**2**	**1**	**2**	**2**	**0**	**2**	**2**
Sequence n° 18/1 (95,45%)	**A**	**C**	**A**	**C**	**C**	**C**	**T**	**A**	**A**	**C**	**A**	**G**	**C**	**A**	**C**	**A**	**C**	**A**	**A**	**T**	**A**	**A**

Original Sequence	**G**	**T**	**G**	**A**	**A**	**A**	**T**	**G**	**G**	**T**	**C**	**A**	**T**	**G**	**T**	**G**	**T**	**G**	**G**	**C**	**G**	**G**
Trend n° 18 (T=0; C=1; G=3; A=2)	**3**	**0**	**3**	**2**	**2**	**2**	**0**	**3**	**3**	**0**	**1**	**2**	**0**	**3**	**0**	**3**	**0**	**3**	**3**	**1**	**3**	**3**
Trend n° 18 (I i D d)	**-**	**D**	**I**	**D**	**d**	**d**	**D**	**I**	**i**	**D**	**I**	**I**	**D**	**I**	**D**	**I**	**D**	**I**	**i**	**D**	**I**	**i**
Trend n° 18/2	**2**	**0**	**2**	**1**	**1**	**1**	**0**	**2**	**2**	**1**	**2**	**3**	**1**	**2**	**1**	**2**	**1**	**2**	**2**	**0**	**2**	**2**
Sequence n° 18/2 (90,90%)	**A**	**T**	**A**	**C**	**C**	**C**	**T**	**A**	**A**	**C**	**A**	**G**	**C**	**A**	**C**	**A**	**C**	**A**	**A**	**T**	**A**	**A**

Modifications following the Trend n° 19 (T=1; C=2; G=3; A=0)

Original Sequence	**G**	**T**	**G**	**A**	**A**	**A**	**T**	**G**	**G**	**T**	**C**	**A**	**T**	**G**	**T**	**G**	**T**	**G**	**G**	**C**	**G**	**G**
Trend n° 19 (T=1; C=2; G=3; A=0)	**3**	**1**	**3**	**0**	**0**	**0**	**1**	**3**	**3**	**1**	**2**	**0**	**1**	**3**	**1**	**3**	**1**	**3**	**3**	**2**	**3**	**3**
Trend n° 19 (I i D d)	**-**	**D**	**I**	**D**	**D**	**d**	**I**	**I**	**i**	**D**	**I**	**D**	**I**	**I**	**D**	**I**	**D**	**I**	**i**	**D**	**I**	**i**
Trend n° 19/1	**2**	**0**	**2**	**1**	**1**	**1**	**2**	**3**	**3**	**2**	**3**	**1**	**2**	**3**	**0**	**2**	**0**	**2**	**2**	**1**	**2**	**2**
Sequence n° 19/1 (90,90%)	**C**	**A**	**C**	**T**	**T**	**T**	**C**	**G**	**G**	**C**	**G**	**T**	**C**	**G**	**A**	**C**	**A**	**C**	**C**	**T**	**C**	**C**

Modifications following the Trend n° 24 (T=1; C=0; G=2; A=3)

Original Sequence	**G**	**T**	**G**	**A**	**A**	**A**	**T**	**G**	**G**	**T**	**C**	**A**	**T**	**G**	**T**	**G**	**T**	**G**	**G**	**C**	**G**	**G**
Trend n° 24 (T=1; C=0; G=2; A=3)	**2**	**1**	**2**	**3**	**3**	**3**	**1**	**2**	**2**	**1**	**0**	**3**	**1**	**2**	**1**	**2**	**1**	**2**	**2**	**0**	**2**	**2**
Trend n° 24 (I i D d)	**-**	**D**	**I**	**I**	**i**	**i**	**D**	**I**	**i**	**D**	**D**	**I**	**D**	**I**	**D**	**I**	**D**	**I**	**i**	**D**	**I**	**i**
Trend n° 24/1	**3**	**0**	**1**	**2**	**2**	**2**	**0**	**3**	**3**	**2**	**1**	**2**	**0**	**1**	**0**	**1**	**0**	**3**	**3**	**1**	**3**	**3**
Sequence n° 24/1 (100%)	**A**	**C**	**T**	**G**	**G**	**G**	**C**	**A**	**A**	**G**	**T**	**G**	**C**	**T**	**C**	**T**	**C**	**A**	**A**	**T**	**A**	**A**

Original Sequence	**G**	**T**	**G**	**A**	**A**	**A**	**T**	**G**	**G**	**T**	**C**	**A**	**T**	**G**	**T**	**G**	**T**	**G**	**G**	**C**	**G**	**G**
Trend n° 24 (T=1; C=0; G=2; A=3)	**2**	**1**	**2**	**3**	**3**	**3**	**1**	**2**	**2**	**1**	**0**	**3**	**1**	**2**	**1**	**2**	**1**	**2**	**2**	**0**	**2**	**2**
Trend n° 24 (I i D d)	**-**	**D**	**I**	**I**	**i**	**i**	**D**	**I**	**i**	**D**	**D**	**I**	**D**	**I**	**D**	**I**	**D**	**I**	**i**	**D**	**I**	**i**
Trend n° 24/2	**1**	**0**	**1**	**2**	**2**	**2**	**0**	**3**	**3**	**2**	**1**	**2**	**0**	**1**	**0**	**1**	**0**	**3**	**3**	**1**	**3**	**3**
Sequence n° 24/2 (100%)	**T**	**C**	**T**	**G**	**G**	**G**	**C**	**A**	**A**	**G**	**T**	**G**	**C**	**T**	**C**	**T**	**C**	**A**	**A**	**T**	**A**	**A**

CHAPTER IX

BLAST RESEARCH AND DISCUSSION OF THE RESULTS OBTAINED OF THE SIGNIFICANT ALIGNMENTS OF SEQUENCE N° 1/3

GOALS OF CHAPTER IX

This chapter contains the results of the *significant alignments* of BLAST research carried out on **Sequence n° 1/3**. This sequence is one of many generated by the **Original Sequence** [the sequence "**GTGAAATGGTCATGTGTGGCGG**" of the RdRp Gene (*RNA-dependent RNA polymerase, RdRp_SARSr-F*)] and from which it differs by **86,36 %** of the bases (see the previous Chapter for the *modification procedure*).
The results of the significant alignments of the BLAST research (**Hitlist size 100**) covered several **common "organisms"** (distinguishable in different colours in the following pages) to the **"organisms"** identified by BLAST research carried out on sequences to which **a base of the original sequence has been replaced** (see Chapters V and VI).
Extending the BLAST research (**Hitlist size 1000**), many **common "organisms"** have been identified, two sequences of ***Pipistrellus pipistrellus*** and a sequence of the ***HIV-1 virus***.
This last sequence was certainly not among those expected, apparently having nothing in "common" with the **original sequence**, other than the fact that both sequences refer to **viruses**, but of a different nature.
HIV infection is caused by one of two similar ***retroviruses*** (***HIV-1*** and ***HIV-2***) that destroy *CD4+ T lymphocytes* and *inhibit the cell-mediated immune response*, increasing the risk of certain infections and tumors. Primary infection can produce a *non-specific febrile syndrome*. The risk of subsequent manifestations, related to the state of immunodeficiency, is proportional to the level of depletion of the circulating CD4+ lymphocyte (https://www.msdmanuals.com/it-it/professionale/malattie-infettive/virus-dell-immunodeficienza-umana-hiv/infezione-da-virus-dell-immunodeficienza-umana-hiv).
Because of the peculiarity of the ***HIV-1 Virus***, its significant alignment has been the object of our in-depth study and, here, it was proposed a **HYPOTHETICAL prediction of the constitutive codons of this segment of Sequence n° 1/3 and of those adjacent to it** (if such *codons* become coding for proteins).
Below there is the ***prediction of the constitutive codons*** and the ***complete sequence of their corresponding amino acids***.
BLAST researches carried out on these **codons** have yielded very interesting results. The roles played by the "**zinc finger protein**" and by the families of **hydrolases** (in particular, the **serine hydrolase**) seem to be of considerable importance.
As it can be appreciated in paragraph 9.7, the results of the BLAST research on **Sequence n° 1/3 (HIV-1) Modified** have identified a significant number of "organisms" and, among these, **three species** of **bats** [**Rhinolophus** (*ferrumequinum*) and **Pipistrellus** (*pipistrellus* e *kuhlii*)], **two species** of **pangolins** [**Manis javanica** (*malayan pangolin*) and **Manis pentadactyla** (*chinese pangolin*)], and various "organisms" involved in mainly **infectious**, **febrile** and **pulmonary diseases** (in particular, **Klebsiella sp.**, **Legionella pneumophila**, and **Simian immunodeficiency virus**). Also, unpredictable alignments with some sequences that refer to **olfactory receptor 6Y1-like**, in **Manis pentadactyla** and **Manis javanica**, a type of *olfactory receptor also common to humans*. A significant codons alignment refers to the **viral neuraminidase** of ***Influenza A Virus***.
Finally, the identification of **two sequences** is very important: **one** that refers to the ***N-terminal domain of the DNA-directed RNA polymerase*** and the **other** that concerns the ***glycosyltransferases family 2*** of the ***Bacillus anthracis***. This **bacterium**, which will also be identified from other sequences obtained from modifications of the **original sequence**, due to its **lethality** and its massive involvement in **lung damage**, will be the subject of extensive discussion in the following chapters.

SEQUENCE N° 1/3

(T=0; C=1; G=2; A=3)

9.1 SEQUENCE N° 1/3 (T=0; C=1; G=2; A=3) – (Hitlist size: 100)

Sequenza Originaria	**G**	**T**	**G**	**A**	**A**	**A**	**T**	**G**	**G**	**T**	**C**	**A**	**T**	**G**	**T**	**G**	**T**	**G**	**G**	**C**	**G**	**G**
Trend n° 1 (T=0; C=1; G=2; A=3)	**2**	**0**	**2**	**3**	**3**	**3**	**0**	**2**	**2**	**0**	**1**	**3**	**0**	**2**	**0**	**2**	**0**	**2**	**2**	**1**	**2**	**2**
Trend n° 1 (I i D d)	**-**	**D**	**I**	**I**	**i**	**i**	**D**	**I**	**i**	**D**	**I**	**I**	**D**	**I**	**D**	**I**	**D**	**I**	**i**	**D**	**I**	**i**
Trend n° 1/3	**1**	**0**	**1**	**2**	**2**	**2**	**1**	**3**	**3**	**0**	**2**	**3**	**1**	**3**	**1**	**3**	**1**	**3**	**3**	**2**	**3**	**3**
Sequenza n° 1/3 (86,36%)	**C**	**T**	**C**	**G**	**G**	**G**	**C**	**A**	**A**	**T**	**G**	**A**	**C**	**A**	**C**	**A**	**C**	**A**	**A**	**G**	**A**	**A**

Query: **CTCGGGCAATGACACACAAGAA**

Search Parameters	
Program	blastn
Word size	7
Expect value	1000
Hitlist size	100
Match/Mismatch scores	1,-3
Gapcosts	5,2
Filter string	F
Genetic Code	1

100 sequences selected:

Accession Seq	Description	Comomm Name	Max Score	Total Score	Query C.	E value	Per. Id.	Acc. Len
XM_033973516.1	PREDICTED: Periophthalmus magnuspinnatus SLU7 homolog, splicing factor (slu7), mRNA	Periophthalmus magnuspinnatus	38.2	38.2	86%	2.3	100.00%	2097
XM_005827469.1	Guillardia theta CCMP2712 hypothetical protein (GUITHDRAFT_113332) mRNA, complete cds	Guillardia theta CCMP2712	38.2	38.2	86%	2.3	100.00%	795
LR778318.1	Raphanus sativus genome assembly, chromosome: 9	radish	36.2	36.2	81%	8.9	100.00%	28894300
LR778313.1	Raphanus sativus genome assembly, chromosome: 4	radish	36.2	66.4	95%	8.9	100.00%	44034337
LR778311.1	Raphanus sativus genome assembly, chromosome: 2	radish	36.2	237	81%	8.9	100.00%	53721154
LR778310.1	Raphanus sativus genome assembly, chromosome: 1	radish	36.2	174	81%	8.9	100.00%	58898396
LR778269.1	Coregonus sp. 'balchen' genome assembly, chromosome: 17	Coregonus sp. 'balchen'	36.2	36.2	81%	8.9	100.00%	54216998
XM_027380184.1	PREDICTED: Penaeus vannamei calmodulin-regulated spectrin-associated protein 2-like (LOC113827292), mRNA	Pacific white shrimp	36.2	36.2	81%	8.9	100.00%	3562
XM_018611851.1	PREDICTED: Raphanus sativus protein phosphatase 2C 3-like (LOC108839004), transcript variant X2, mRNA	radish	36.2	36.2	81%	8.9	100.00%	957
XM_037937904.1	PREDICTED: Penaeus monodon uncharacterized LOC119589286 (LOC119589286), mRNA	black tiger shrimp	36.2	36.2	81%	8.9	100.00%	11816
AC127413.3	Mus musculus BAC clone RP23-236A4 from 15, complete sequence	house mouse	36.2	36.2	81%	8.9	100.00%	194448
LR778315.1	Raphanus sativus genome assembly, chromosome: 6	radish	36.2	36.2	81%	8.9	100.00%	35522300
LR778314.1	Raphanus sativus genome assembly, chromosome: 5	radish	36.2	36.2	81%	8.9	100.00%	43617977
LT222054.1	Fusarium graminearum genome assembly, chromosome: II	Fusarium graminearum	36.2	36.2	81%	8.9	100.00%	9025752
LR606190.1	Aquila chrysaetos chrysaetos genome assembly, chromosome: 10	Aquila chrysaetos chrysaetos	36.2	96.6	86%	8.9	100.00%	43949808
LR584434.1	Salmo trutta genome assembly, chromosome: 34	river trout	36.2	36.2	81%	8.9	100.00%	42900288
LR131992.1	Gouania willdenowi genome assembly, chromosome: 16	blunt-snouted clingfish	36.2	66.4	81%	8.9	100.00%	40234402

XM_009860016.3	PREDICTED: Ciona intestinalis vang-like protein 2 (LOC100187060), transcript variant X2, mRNA	vase tunicate	36.2	36.2	81%	8.9		100.00%	2338
XM_002121917.4	PREDICTED: Ciona intestinalis vang-like protein 2 (LOC100187060), transcript variant X1, mRNA	vase tunicate	36.2	36.2	81%	8.9		100.00%	2369
CP012813.1	Saccharomycopsis fibuligera x Saccharomycopsis cf. fibuligera strain KJJ81 chromosome A5, complete sequence	Saccharomycopsis fibuligera x Saccharomycopsis cf. fibuligera	36.2	36.2	81%	8.9		100.00%	2119959
CP015982.1	Saccharomycopsis fibuligera strain ATCC 36309 chromosome 5, complete sequence	Saccharomycopsis fibuligera	36.2	36.2	81%	8.9		100.00%	2406875
CP012827.1	Saccharomycopsis fibuligera strain KPH12 chromosome 5, complete sequence	Saccharomycopsis fibuligera	36.2	36.2	81%	8.9		100.00%	2100688
HG970333.1	Fusarium graminearum chromosome 2, complete genome	Fusarium graminearum	36.2	36.2	81%	8.9		100.00%	8997558
CP064753.1	Fusarium pseudograminearum strain Class2-1C chromosome 1	Fusarium pseudograminearum	36.2	36.2	81%	8.9		100.00%	9153420
XM_035854878.1	PREDICTED: Folsomia candida uncharacterized LOC110863674 (LOC110863674), transcript variant X2, mRNA	Folsomia candida	36.2	36.2	81%	8.9		100.00%	1996
XM_022113035.2	PREDICTED: Folsomia candida uncharacterized LOC110863674 (LOC110863674), transcript variant X1, mRNA	Folsomia candida	36.2	36.2	81%	8.9		100.00%	2620
XM_011324171.1	Fusarium graminearum PH-1 hypothetical protein partial mRNA	Fusarium graminearum PH-1	36.2	36.2	81%	8.9		100.00%	1197
AK113083.1	Ciona intestinalis cDNA, clone:ciad005e21, full insert sequence	vase tunicate	36.2	36.2	81%	8.9		100.00%	2377
LR812108.1	Erithacus rubecula genome assembly, chromosome: 6	European robin	34.2	34.2	77%		35	100.00%	60684490
LR812132.1	Erithacus rubecula genome assembly, chromosome: 28	European robin	34.2	34.2	77%		35	100.00%	5344592
LR736851.1	Pecten maximus genome assembly, chromosome: 14	Pecten maximus	34.2	64.4	77%		35	100.00%	38909756
LR633949.1	Gadus morhua genome assembly, chromosome: 7	Atlantic cod	34.2	64.4	77%		35	100.00%	34137969
XM_029518756.1	PREDICTED: Echeneis naucrates fructose-bisphosphate aldolase C-B (LOC115053889), mRNA	live sharksucker	34.2	34.2	77%		35	100.00%	2268
LR584231.2	Takifugu rubripes genome assembly, chromosome: 1	torafugu	34.2	34.2	77%		35	100.00%	29232231
LR594596.1	Syngnathus acus genome assembly, chromosome: 11	greater pipefish	34.2	64.4	81%		35	100.00%	8880460
LR594595.1	Syngnathus acus genome assembly, chromosome: 10	greater pipefish	34.2	94.6	81%		35	100.00%	28444102
XM_028725994.1	PREDICTED: Podarcis muralis CREB/ATF bZIP transcription factor (CREBZF), mRNA	Common wall lizard	34.2	34.2	77%		35	100.00%	2516
LR584409.1	**Salmo trutta genome assembly, chromosome: 9**	river trout	34.2	34.2	77%		35	100.00%	49359203
LR584404.1	Rhinatrema bivittatum genome assembly, chromosome: 18	two-lined caecilian	34.2	64.4	81%		35	100.00%	61353747
LR584066.1	Scleropages formosus genome assembly, chromosome: 1	Asian bonytongue	34.2	215	86%		35	100.00%	54115380
LR584055.1	Echeneis naucrates genome assembly, chromosome: 14	live sharksucker	34.2	64.4	95%		35	100.00%	22459569
LR131995.1	**Gouania willdenowi genome assembly, chromosome: 10**	blunt-snouted clingfish	34.2	34.2	77%		35	100.00%	42970088
CP024767.1	Pseudomonas arsenicoxydans strain ACM1 chromosome	Pseudomonas arsenicoxydans	34.2	34.2	77%		35	100.00%	6620935
CP034496.1	Eukaryotic synthetic construct chromosome 18	eukaryotic synthetic construct	34.2	34.2	77%		35	100.00%	84899006
LR134442.1	Propionibacterium australiense strain NCTC13651 genome assembly, chromosome: 1	Propionibacterium australiense	34.2	34.2	77%		35	100.00%	3044331
XM_023378987.1	PREDICTED: Centruroides sculpturatus beta-2 adrenergic receptor-like (LOC111634253), mRNA	bark scorpion	34.2	34.2	77%		35	100.00%	1695
KY906251.1	Claviceps gigantea strain SP1 DmaW, EasG, EasF, EasE, EasD, EasC, CloA, EasA, Lps2, truncated Lps1, and EasH genes, complete cds	Claviceps gigantea	34.2	34.2	77%		35	100.00%	40254
LT606948.1	Polynucleobacter necessarius	Polynucleobacter	34.2	34.2	77%		35	100.00%	1655757

	isolate PPGSP3 genome assembly, chromosome: I	necessarius						
XR_002010978.1	PREDICTED: Cyprinus carpio uncharacterized LOC109048656 (LOC109048656), transcript variant X2, ncRNA	common carp	34.2	34.2	95%	35	95.24%	360
XR_002010977.1	PREDICTED: Cyprinus carpio uncharacterized LOC109048656 (LOC109048656), transcript variant X1, ncRNA	common carp	34.2	34.2	95%	35	95.24%	424
XR_002016362.1	PREDICTED: Cyprinus carpio uncharacterized LOC109082096 (LOC109082096), ncRNA	common carp	34.2	34.2	95%	35	95.24%	596
XR_002015984.1	PREDICTED: Cyprinus carpio uncharacterized LOC109079651 (LOC109079651), ncRNA	common carp	34.2	34.2	95%	35	95.24%	626
CP017815.1	Sclerotinia sclerotiorum chromosome 2, complete sequence	Sclerotinia sclerotiorum 1980 UF-70	34.2	34.2	77%	35	100.00%	3683506
XM_018231673.1	PREDICTED: Xenopus laevis alpha-1,3-mannosyl-glycoprotein 4-beta-N-acetylglucosaminyltransferase B-like (LOC108699587), mRNA	African clawed frog	34.2	34.2	77%	35	100.00%	3316
XR_461531.2	PREDICTED: Monodelphis domestica uncharacterized LOC103103228 (LOC103103228), ncRNA	gray short-tailed opossum	34.2	34.2	77%	35	100.00%	760
XM_014937915.1	PREDICTED: Calidris pugnax FK506 binding protein 14 (FKBP14), mRNA	ruff	34.2	34.2	77%	35	100.00%	1807
FO906018.1	Leptosphaeria maculans lepidii ibcn84_scaffold00006 complete sequence	Leptosphaeria maculans 'lepidii' group	34.2	34.2	77%	35	100.00%	1544579
LN598375.2	Cyprinus carpio genome assembly common carp genome, scaffold 000000475	common carp	34.2	68.4	95%	35	95.24%	1926052
LN590701.1	Cyprinus carpio genome assembly common carp genome, scaffold: LG1, chromosome: 1	common carp	34.2	64.4	100%	35	95.24%	15297981
XM_037263922.1	PREDICTED: Syngnathus acus RUNX1 partner transcriptional co-repressor 1 (runx1t1), transcript variant X5, mRNA	greater pipefish	34.2	34.2	77%	35	100.00%	10264
AP023478.1	Homo sapiens DNA, chromosome 18, nearly complete genome	human	34.2	34.2	77%	35	100.00%	77846715
XM_023272289.2	PREDICTED: Amphiprion ocellaris transmembrane protein 182-like (LOC111569869), mRNA	clown anemonefish	34.2	34.2	77%	35	100.00%	2086
LR877200.1	Onychomys torridus genome assembly, chromosome: 13	southern grasshopper mouse	34.2	159	77%	35	100.00%	78371027
LR862393.1	Arvicola amphibius genome assembly, chromosome: 13	Eurasian water vole	34.2	94.6	86%	35	100.00%	75712621
XM_034973366.1	PREDICTED: Aphantopus hyperantus pre-mRNA-splicing factor Slu7 (LOC117986526), mRNA	ringlet	34.2	34.2	77%	35	100.00%	1744
CP008956.1	Herbaspirillum rubrisubalbicans Os34, complete genome	Herbaspirillum rubrisubalbicans Os34	34.2	34.2	77%	35	100.00%	6122717
CP050582.1	Canis lupus familiaris breed Labrador retriever chromosome 27a	dog	34.2	34.2	77%	35	100.00%	46050380
CP050640.1	Canis lupus familiaris breed Labrador retriever chromosome 27b	dog	34.2	34.2	77%	35	100.00%	46054595
AP022707.1	Plectropomus leopardus DNA, chromosome 8, nearly complete sequence	leopard coralgrouper	34.2	34.2	77%	35	100.00%	35874999
LR761658.1	Aphantopus hyperantus genome assembly, chromosome: 11	ringlet	34.2	34.2	77%	35	100.00%	15241335
XM_032439116.1	PREDICTED: Drosophila virilis octopamine receptor Oamb (LOC6629914), transcript variant X2, mRNA	Drosophila virilis	34.2	34.2	77%	35	100.00%	4221
XM_032439114.1	PREDICTED: Drosophila virilis octopamine receptor Oamb (LOC6629914), transcript variant X1, mRNA	Drosophila virilis	34.2	34.2	77%	35	100.00%	5158
XM_031711235.1	PREDICTED: Actinia tenebrosa serine/threonine-protein kinase Haspin homolog ALK1-like	Australian red waratah sea anemone	34.2	34.2	77%	35	100.00%	4119

	(LOC116302041), mRNA							
LR606204.1	Aquila chrysaetos chrysaetos genome assembly, chromosome: 24	Aquila chrysaetos chrysaetos	34.2	34.2	77%	35	100.00%	20993271
XM_028460065.1	**PREDICTED: Gouania willdenowi SLU7 homolog, splicing factor (slu7), mRNA**	blunt-snouted clingfish	34.2	34.2	77%	35	100.00%	2000
XR_003423601.1	PREDICTED: Acinonyx jubatus uncharacterized LOC113602220 (LOC113602220), transcript variant X2, ncRNA	cheetah	34.2	34.2	77%	35	100.00%	1382
XM_026144372.1	PREDICTED: Astatotilapia calliptera POU class 2 homeobox 1 (pou2f1), transcript variant X5, mRNA	eastern happy	34.2	34.2	77%	35	100.00%	9015
CP032600.1	Lateolabrax maculatus linkage group 3 sequence	spotted sea bass	34.2	34.2	77%	35	100.00%	29073605
XM_025865289.1	PREDICTED: Callorhinus ursinus olfactory receptor-like protein OLF4 (LOC112818023), mRNA	northern fur seal	34.2	34.2	77%	35	100.00%	933
XM_024828719.1	Aspergillus novofumigatus IBT 16806 WD domain protein (P174DRAFT_447931), mRNA	Aspergillus novofumigatus IBT 16806	34.2	34.2	77%	35	100.00%	2649
CP027092.1	Bos mutus isolate yakQH1 chromosome 24	wild yak	34.2	34.2	77%	35	100.00%	61837517
XR_002737306.1	PREDICTED: Felis catus uncharacterized LOC109493369 (LOC109493369), transcript variant X1, ncRNA	domestic cat	34.2	34.2	77%	35	100.00%	1389
CP022464.2	Enterocloster bolteae strain ATCC BAA-613 chromosome, complete genome	Enterocloster bolteae	34.2	34.2	77%	35	100.00%	6570176
XM_016051987.2	PREDICTED: Parasteatoda tepidariorum pre-mRNA-splicing factor SLU7-like (LOC107439400), transcript variant X3, mRNA	common house spider	34.2	34.2	77%	35	100.00%	2977
XM_016051986.2	PREDICTED: Parasteatoda tepidariorum pre-mRNA-splicing factor SLU7-like (LOC107439400), transcript variant X2, mRNA	common house spider	34.2	34.2	77%	35	100.00%	2977
XM_016051985.2	PREDICTED: Parasteatoda tepidariorum pre-mRNA-splicing factor SLU7-like (LOC107439400), transcript variant X1, mRNA	common house spider	34.2	34.2	77%	35	100.00%	2747
XM_008668489.2	PREDICTED: Zea mays uncharacterized LOC103645435 (LOC103645435), mRNA	Zea mays	34.2	34.2	77%	35	100.00%	1182
XM_019870024.1	PREDICTED: Hippocampus comes SLU7 homolog, splicing factor (slu7), mRNA	tiger tail seahorse	34.2	34.2	77%	35	100.00%	2118
XM_016622770.1	PREDICTED: Nicotiana tabacum uncharacterized LOC107799630 (LOC107799630), mRNA	common tobacco	34.2	34.2	77%	35	100.00%	951
XM_016051988.1	PREDICTED: Parasteatoda tepidariorum pre-mRNA-splicing factor SLU7-like (LOC107439400), transcript variant X4, mRNA	common house spider	34.2	34.2	77%	35	100.00%	2720
XM_014149233.1	PREDICTED: Salmo salar Receptor-type tyrosine-protein phosphatase-like N (ptprn), transcript variant X5, mRNA	Atlantic salmon	34.2	34.2	77%	35	100.00%	3767
XM_014149231.1	PREDICTED: Salmo salar Receptor-type tyrosine-protein phosphatase-like N (ptprn), transcript variant X3, mRNA	Atlantic salmon	34.2	34.2	77%	35	100.00%	3855
LN595020.1	Cyprinus carpio genome assembly common carp genome, scaffold 000011327	common carp	34.2	34.2	95%	35	95.24%	171567
LN591285.1	Cyprinus carpio genome assembly common carp genome, scaffold 000001455	common carp	34.2	68.4	95%	35	95.24%	657936
LN590715.1	Cyprinus carpio genome assembly common carp genome, scaffold: LG46, chromosome: 46	common carp	34.2	68.4	95%	35	95.24%	7828959
XM_037261829.1	PREDICTED: Syngnathus acus SLU7 homolog, splicing factor (slu7), mRNA	greater pipefish	34.2	34.2	77%	35	100.00%	2126
XM_035668914.1	PREDICTED: Morone saxatilis SLU7 homolog, splicing factor (slu7), mRNA	striped sea-bass	34.2	34.2	77%	35	100.00%	2257
LR828303.1	Ananas comosus genome assembly, chromosome: 23	pineapple	34.2	34.2	77%	35	100.00%	7979763

LR812531.1	Danio kyathit genome assembly, chromosome: 13	Danio kyathit	34.2	64.4	86%	35	100.00%	63723151
LR812514.1	Danio aesculapii genome assembly, chromosome: 18	Danio aesculapii	34.2	94.6	86%	35	100.00%	56436977

ALIGNMENT VIEW FLAT QUERY-ANCHORED WITH LETTERS FOR IDENTITIES

Query range 1: 1 to 22

Query	1	CTCGGGCAATGACACACAAGAA	22
XM_033973516.1	561	GGGCAATGACACACAAGAA	579
XM_005827469.1	555	GGGCAATGACACACAAGAA	573
LR778318.1	20488114	GGCAATGACACACAAGAA	20488097
LR778313.1	37627953	GGCAATGACACACAAGAA	37627936
LR778313.1	10890035	TCGGGCAATGACACA	10890049
LR778311.1	43569344	GGCAATGACACACAAGAA	43569327
LR778311.1	45938608	GGCAATGACACACAAGAA	45938591
LR778311.1	45939112	GGCAATGACACACAAGAA	45939095
LR778311.1	51271958	GGCAATGACACACAAGAA	51271941
LR778311.1	46217694	GCAATGACACACAAGA	46217709
LR778311.1	3082211	CAATGACACACAAGA	3082225
LR778311.1	27542258	GCAATGACACACAAG	27542272
LR778310.1	23211306	GGCAATGACACACAAGAA	23211289
LR778310.1	41776867	GGCAATGACACACAAGAA	41776850
LR778310.1	54685832	GGCAATGACACACAAGAA	54685815
LR778310.1	39860408	GGCAATGACACACAAGAA	39860425
LR778310.1	12039277	CAATGACACACAAGA	12039291
LR778269.1	36586821	GGCAATGACACACAAGAA	36586804
XM_027380184.1	978	GGCAATGACACACAAGAA	961
XM_018611851.1	227	GGCAATGACACACAAGAA	210
XM_037937904.1	6452	GGCAATGACACACAAGAA	6435
AC127413.3	115955	GGGCAATGACACACAAGA	115938
LR778315.1	9667185	GGCAATGACACACAAGAA	9667202
LR778314.1	38969598	GGCAATGACACACAAGAA	38969615
LT222054.1	4700671	GGCAATGACACACAAGAA	4700688
LR606190.1	1396954	GGGCAATGACACACAAGA	1396971
LR606190.1	38591960	GGGCAATGACACACA	38591946
LR606190.1	14729696	AATGACACACAAGAA	14729710
LR584434.1	1993267	GGCAATGACACACAAGAA	1993284
LR131992.1	30835046	TCGGGCAATGACACACAA	30835063
LR131992.1	2784243	GGCAATGACACACAA	2784257
XM_009860016.3	1090	CGGGCAATGACACACAAG	1107
XM_002121917.4	1121	CGGGCAATGACACACAAG	1138
CP012813.1	344930	GGCAATGACACACAAGAA	344947
CP015982.1	345332	GGCAATGACACACAAGAA	345349
CP012827.1	326328	GGCAATGACACACAAGAA	326345
HG970333.1	4700478	GGCAATGACACACAAGAA	4700495
CP064753.1	4902498	GGCAATGACACACAAGAA	4902515
XM_035854878.1	1101	GGCAATGACACACAAGAA	1118
XM_022113035.2	1096	GGCAATGACACACAAGAA	1113

XM_011324171.1	999	GGCAATGACACACAAGAA	1016
AK113083.1	1119	CGGGCAATGACACACAAG	1136
LR812108.1	29479735	GGGCAATGACACACAAG	29479719
LR812132.1	2930334	TCGGGCAATGACACACA	2930318
LR736851.1	7357322	GGCAATGACACACAAGA	7357306
LR736851.1	27850302	GCAATGACACACAAG	27850316
LR633949.1	19930737	GGGCAATGACACACAAG	19930721
LR633949.1	26898888	GCAATGACACACAAG	26898874
XM_029518756.1	2053	GCAATGACACACAAGAA	2037
LR584231.2	65340	GGCAATGACACACAAGA	65324
LR594596.1	1284376	GCAATGACACACAAGAA	1284360
LR594596.1	867331	GGCAATGACACACAA	867345
LR594595.1	10662806	GCAATGACACACAAGAA	10662790
LR594595.1	393033	CAATGACACACAAGA	393047
LR594595.1	23467433	GGCAATGACACACAA	23467447
XM_028725994.1	2200	GGGCAATGACACACAAG	2184
LR584409.1	7773759	GGGCAATGACACACAAG	7773743
LR584404.1	50363248	GGCAATGACACACAAGA	50363232
LR584404.1	23932806	AATGACACACAAGAA	23932820
LR584066.1	4025644	GCAATGACACACAAGAA	4025628
LR584066.1	1721156	AATGACACACAAGAA	1721142
LR584066.1	4808488	GGGCAATGACACACA	4808474
LR584066.1	19109764	GGGCAATGACACACA	19109750
LR584066.1	29792876	GGGCAATGACACACA	29792862
LR584066.1	46635111	GGGCAATGACACACA	46635097
LR584066.1	13615803	GGGCAATGACACACA	13615817
LR584055.1	11087997	GCAATGACACACAAGAA	11087981
LR584055.1	21657622	TCGGGCAATGACACA	21657636
LR131995.1	19577432	GCAATGACACACAAGAA	19577416
CP024767.1	4158911	CTCGGGCAATGACACAC	4158895
CP034496.1	71419010	GGGCAATGACACACAAG	71418994
LR134442.1	471516	GCAATGACACACAAGAA	471500
XM_023378987.1	496	GGGCAATGACACACAAG	480
KY906251.1	38015	CTCGGGCAATGACACAC	37999
LT606948.1	467718	GGGCAATGACACACAAG	467702
XR_002010978.1	116	CTCGGGCAATGACGCACAAGA	96
XR_002010977.1	116	CTCGGGCAATGACGCACAAGA	96
XR_002016362.1	118	CTCGGGCAATGACGCACAAGA	98
XR_002015984.1	116	CTCGGGCAATGACGCACAAGA	96
CP017815.1	2944133	GGGCAATGACACACAAG	2944117
XM_018231673.1	36	GGGCAATGACACACAAG	20
XR_461531.2	701	GGGCAATGACACACAAG	685
XM_014937915.1	1231	GCAATGACACACAAGAA	1215
FO906018.1	1140563	CGGGCAATGACACACAA	1140547
LN598375.2	649314	CTCGGGCAATGACGCACAAGA	649294
LN598375.2	655726	CTCGGGCAATGACGCACAAGA	655706
LN590701.1	3634732	CTCGGGCAATGACGCACAAGA	3634712
LN590701.1	14912533	AATGACACACAAGAA	14912547
XM_037263922.1	251	GCAATGACACACAAGAA	235
AP023478.1	71303764	GGGCAATGACACACAAG	71303748
XM_023272289.2	2019	CGGGCAATGACACACAA	2003
LR877200.1	16623045	GCAATGACACACAAGAA	16623029
LR877200.1	48091612	GCAATGACACACAAGAA	48091596
LR877200.1	21008003	GCAATGACACACAAG	21008017
LR877200.1	44106519	AATGACACACAAGAA	44106533
LR877200.1	60749189	CAATGACACACAAGA	60749203
LR862393.1	51550451	GCAATGACACACAAGAA	51550435
LR862393.1	10046671	GGGCAATGACACACA	10046657
LR862393.1	52001369	AATGACACACAAGAA	52001383
XM_034973366.1	398	GCAATGACACACAAGAA	414

CP008956.1	2013233	GCAATGACACACAAGAA	2013249
CP050582.1	4860768	CTCGGGCAATGACACAC	4860784
CP050640.1	4858108	CTCGGGCAATGACACAC	4858124
AP022707.1	6289195	GCAATGACACACAAGAA	6289211
LR761658.1	1638750	GCAATGACACACAAGAA	1638766
XM_032439116.1	354	CTCGGGCAATGACACAC	370
XM_032439114.1	333	CTCGGGCAATGACACAC	349
XM_031711235.1	1995	GCAATGACACACAAGAA	2011
LR606204.1	18859071	GGGCAATGACACACAAG	18859087
XM_028460065.1	501	GCAATGACACACAAGAA	517
XR_003423601.1	609	GCAATGACACACAAGAA	625
XM_026144372.1	1675	CGGGCAATGACACACAA	1691
CP032600.1	8126049	GGGCAATGACACACAAG	8126065
XM_025865289.1	8	CGGGCAATGACACACAA	24
XM_024828719.1	650	GGCAATGACACACAAGA	666
CP027092.1	45563484	CGGGCAATGACACACAA	45563500
XR_002737306.1	610	GCAATGACACACAAGAA	626
CP022464.2	5016426	CGGGCAATGACACACAA	5016442
XM_016051987.2	477	GCAATGACACACAAGAA	493
XM_016051986.2	477	GCAATGACACACAAGAA	493
XM_016051985.2	538	GCAATGACACACAAGAA	554
XM_008668489.2	735	GGCAATGACACACAAGA	751
XM_019870024.1	502	GCAATGACACACAAGAA	518
XM_016622770.1	73	GGCAATGACACACAAGA	89
XM_016051988.1	538	GCAATGACACACAAGAA	554
XM_014149233.1	1259	GGGCAATGACACACAAG	1275
XM_014149231.1	1260	GGGCAATGACACACAAG	1276
LN595020.1	5226	CTCGGGCAATGACGCACAAGA	5246
LN591285.1	601630	CTCGGGCAATGACGCACAAGA	601650
LN591285.1	602415	CTCGGGCAATGACGCACAAGA	602435
LN590715.1	3440524	CTCGGGCAATGACGCACAAGA	3440544
LN590715.1	3447044	CTCGGGCAATGACGCACAAGA	3447064
XM_037261829.1	486	GCAATGACACACAAGAA	502
XM_035668914.1	519	GCAATGACACACAAGAA	535
LR828303.1	653605	CGGGCAATGACACACAA	653621
LR812531.1	54775021	GCAATGACACACAAGAA	54775037
LR812531.1	33111773	GGGCAATGACACACA	33111787
LR812514.1	54032613	GGCAATGACACACAAGA	54032629
LR812514.1	50154682	GGGCAATGACACACA	50154668
LR812514.1	34011470	AATGACACACAAGAA	34011484

9.2 SEQUENCE N° 1/3 (T=0; C=1; G=2; A=3) (Hitlist size 1000)

Query: CTCGGGCAATGACACACAAGAA

Search Parameters	
Program	blastn
Word size	7
Expect value	1000
Hitlist size	**1000**
Match/Mismatch scores	1,-3
Gapcosts	5,2
Filter string	F
Genetic Code	1

998 sequences selected:

Accession Seq	Description	Comomm Name	Max Score	Total Score	Query C.	E value	Per. Id.	Acc. Len
.....	.....	.....	.....	.....	.....	.....	.....	...
LR778269.1	Coregonus sp. 'balchen' genome assembly, chromosome: 17	Coregonus sp. 'balchen'	36.2	36.2	81%	9.2	100.00%	54216998
.....	.....	.....	.....	.....	.....	.....	.....	...
LR584434.1	Salmo trutta genome assembly, chromosome: 34	Salmo trutta	36.2	36.2	81%	9.2	100.00%	42900288
LR131992.1	Gouania willdenowi genome assembly, chromosome: 16	Gouania willdenowi	36.2	66.4	81%	9.2	100.00%	40234402
.....	.....	.....	.....	.....	.....	.....	.....	...
LR584409.1	Salmo trutta genome assembly, chromosome: 9	Salmo trutta	34.2	34.2	77%	37	100.00%	49359203
.....	.....	.....	.....	.....	.....	.....	.....	...
LR131995.1	Gouania willdenowi genome assembly, chromosome: 10	Gouania willdenowi	34.2	34.2	77%	37	100.00%	42970088
.....	.....	.....	.....	.....	.....	.....	.....	...
XM_028460065.1	PREDICTED: Gouania willdenowi SLU7 homolog, splicing factor (slu7), mRNA	Gouania willdenowi	34.2	34.2	77%	37	100.00%	2000
.....	.....	.....	.....	.....	.....	.....	.....	...
OD568846.1	4_Tbi_b3v08	Timema bartmani	34.2	34.2	77%	37	100.00%	57173
.....	.....	.....	.....	.....	.....	.....	.....	...
FJ388909.1	**HIV-1 isolate CY064 from Cyprus nonfunctional gag protein (gag) gene, complete sequence; pol protein (pol) gene, partial cds; vif protein (vif), vpr protein (vpr), tat protein (tat), rev protein (rev), and vpu protein (vpu) genes, complete cds; nonfunctional envelope glycoprotein (env) gene, complete sequence; and nef protein (nef) gene, partial cds**	Human immunodeficiency virus 1	34.2	34.2	77%	37	100.00%	8211
.....	.....	.....	.....	.....	.....	.....	.....	...
LR778274.1	Coregonus sp. 'balchen' genome	Coregonus sp. 'balchen'	32.2	64.4	77%	144	100.00%	56862223

	assembly, chromosome: 22							
LR778272.1	Coregonus sp. 'balchen' genome assembly, chromosome: 20	Coregonus sp. 'balchen'	32.2	62.4	77%	144	100.00%	54335267
LR778266.1	Coregonus sp. 'balchen' genome assembly, chromosome: 14	Coregonus sp. 'balchen'	32.2	32.2	72%	144	100.00%	47256133
.....	.....	.....	.....	.....	.....	.....	.....	...
XR_003985082.1	PREDICTED: Sparus aurata uncharacterized LOC115587570 (LOC115587570), transcript variant X2, ncRNA	Sparus aurata	32.2	32.2	72%	144	100.00%	715
.....	.....	.....	.....	.....	.....	.....	.....	...
LR697121.1	Chanos chanos genome assembly, chromosome: 16	Chanos chanos	32.2	126	77%	144	100.00%	18553434
LR697113.1	Chanos chanos genome assembly, chromosome: 8	Chanos chanos	32.2	155	81%	144	100.00%	48343650
LR697106.1	Chanos chanos genome assembly, chromosome: 1	Chanos chanos	32.2	187	81%	144	100.00%	63121898
.....	.....	.....	.....	.....	.....	.....	.....	...
LR537129.1	Sparus aurata genome assembly, chromosome: 9	Sparus aurata	32.2	32.2	72%	144	100.00%	37012328
.....	.....	.....	.....	.....	.....	.....	.....	...
XM_029711715.1	PREDICTED: Salmo trutta serine/arginine repetitive matrix protein 1-like (LOC115161037), transcript variant X2, mRNA	Salmo trutta	32.2	32.2	72%	144	100.00%	5051
XM_029711714.1	PREDICTED: Salmo trutta serine/arginine repetitive matrix protein 1-like (LOC115161037), transcript variant X1, mRNA	Salmo trutta	32.2	32.2	72%	144	100.00%	5054
.....	.....	.....	.....	.....	.....	.....	.....	...
LR597571.1	Myripristis murdjan genome assembly, chromosome: 22	Myripristis murdjan	32.2	32.2	72%	144	100.00%	31899175
LR597569.1	Myripristis murdjan genome assembly, chromosome: 20	Myripristis murdjan	32.2	32.2	72%	144	100.00%	29743470
.....	.....	.....	.....	.....	.....	.....	.....	...
LR584421.1	Salmo trutta genome assembly, chromosome: 23	Salmo trutta	32.2	32.2	72%	144	100.00%	51488243
LR584412.1	Salmo trutta genome assembly, chromosome: 24	Salmo trutta	32.2	32.2	72%	144	100.00%	50333583
LR584410.1	Salmo trutta genome assembly, chromosome: 1	Salmo trutta	32.2	62.4	81%	144	100.00%	81542925
.....	.....	.....	.....	.....	.....	.....	.....	...
CP018158.1	Oryza sativa Indica Group cultivar Shuhui498 chromosome 2 sequence	Oryza sativa Indica Group	32.2	32.2	72%	144	100.00%	37764328
.....	.....	.....	.....	.....	.....	.....	.....	...
AP014958.1	Oryza sativa Japonica Group DNA, chromosome 2, cultivar: Nipponbare, complete sequence	Oryza sativa Japonica Group	32.2	32.2	72%	144	100.00%	35937250
CP012610.1	Oryza sativa Indica Group cultivar RP Bio-226 chromosome 2 sequence	Oryza sativa Indica Group	32.2	32.2	72%	144	100.00%	36385228
.....	.....	.....	.....	.....	.....	.....	.....	...
HG975443.1	Solanum pennellii chromosome ch04, complete genome	Solanum pennellii	32.2	32.2	72%	144	100.00%	77197300
.....	.....	.....	.....	.....	.....	.....	.....	...
CP056053.1	Oryza sativa Indica Group cultivar Zhenshan 97 chromosome 2	Oryza sativa Indica Group	32.2	32.2	72%	144	100.00%	37267338
CP054677.1	Oryza sativa Indica Group cultivar Minghui 63 chromosome 2	Oryza sativa Indica Group	32.2	32.2	72%	144	100.00%	37301368
.....	.....	.....	.....	.....	.....	.....	.....	...
OC319781.1	3_Tce_b3v08	Timema cristinae	32.2	32.2	72%	144	100.00%	79776
.....	.....	.....	.....	.....	.....	.....	.....	...
OD063034.1	1_Tps_b3v08	Timema poppensis	32.2	32.2	72%	144	100.00%	1644
OC724875.1	3_Tce_b3v08	Timema cristinae	32.2	32.2	72%	144	100.00%	286
OC027313.1	2_Tsi_b3v08	Timema shepardi	32.2	32.2	72%	144	100.00%	2793
.....	.....	.....	.....	.....	.....	.....	.....	...
LR877236.1	Acomys russatus	Acomys russatus	32.2	122	95%	144	100.00%	53031086

	genome assembly, chromosome: 25								
LR877232.1	Acomys russatus genome assembly, chromosome: 21	Acomys russatus	32.2	64.4	81%	144	100.00%	61132863	
LR877225.1	Acomys russatus genome assembly, chromosome: 14	Acomys russatus	32.2	92.7	90%	144	100.00%	62749011	
.....	.....	.....	.....	.....	.....	.....	.....	...	
LR862372.1	Pipistrellus pipistrellus genome assembly, chromosome: 16	Pipistrellus pipistrellus	32.2	96.6	77%	144	100.00%	46231460	
LR862362.1	Pipistrellus pipistrellus genome assembly, chromosome: 6	Pipistrellus pipistrellus	32.2	62.4	72%	144	100.00%	85793147	
.....	.....	.....	.....	.....	.....	.....	.....	...	
AP004066.2	Oryza sativa Japonica Group genomic DNA, chromosome 2, BAC clone:OJ1572_F02	Oryza sativa Japonica Group	32.2	32.2	72%	144	100.00%	93626	
.....	.....	.....	.....	.....	.....	.....	.....	...	
LR778285.1	Coregonus sp. 'balchen' genome assembly, chromosome: 33	Coregonus sp. 'balchen'	32.2	92.7	86%	144	100.00%	44662967	
LR778262.1	Coregonus sp. 'balchen' genome assembly, chromosome: 10	Coregonus sp. 'balchen'	32.2	32.2	72%	144	100.00%	60468309	
LR778253.1	Coregonus sp. 'balchen' genome assembly, chromosome: 1	Coregonus sp. 'balchen'	32.2	32.2	72%	144	100.00%	93459789	
.....	.....	.....	.....	.....	.....	.....	.....	...	
LR722985.1	Thalassophryne amazonica genome assembly, chromosome: 20	Thalassophryne amazonica	32.2	124	95%	144	100.00%	70111961	
LR722979.1	Thalassophryne amazonica genome assembly, chromosome: 14	Thalassophryne amazonica	32.2	32.2	72%	144	100.00%	99266177	
XM_030771950.1	PREDICTED: Chanos chanos glycerophosphocholine phosphodiesterase 1 (gpcpd1), mRNA	Chanos chanos	32.2	32.2	72%	144	100.00%	3535	
.....	.....	.....	.....	.....	.....	.....	.....	...	
XM_030436014.1	PREDICTED: Sparus aurata sigma non-opioid intracellular receptor 1 (sigmar1), mRNA	Sparus aurata	32.2	32.2	72%	144	100.00%	2861	
.....	.....	.....	.....	.....	.....	.....	.....	...	
LR697118.1	Chanos chanos genome assembly, chromosome: 13	Chanos chanos	32.2	32.2	72%	144	100.00%	23837493	
LR697115.1	Chanos chanos genome assembly, chromosome: 10	Chanos chanos	32.2	157	77%	144	100.00%	40347274	
LR697112.1	Chanos chanos genome assembly, chromosome: 7	Chanos chanos	32.2	32.2	72%	144	100.00%	49187548	
LR697109.1	Chanos chanos genome assembly, chromosome: 4	Chanos chanos	32.2	153	77%	144	100.00%	53909382	
.....	.....	.....	.....	.....	.....	.....	.....	...	
LR537132.1	Sparus aurata genome assembly, chromosome: 12	Sparus aurata	32.2	62.4	86%	144	100.00%	30193437	
.....	.....	.....	.....	.....	.....	.....	.....	...	
LR597562.1	Myripristis murdjan genome assembly, chromosome: 13	Myripristis murdjan	32.2	32.2	72%	144	100.00%	42404719	
LR597553.1	Myripristis murdjan genome assembly, chromosome: 4	Myripristis murdjan	32.2	32.2	72%	144	100.00%	36530440	
.....	.....	.....	.....	.....	.....	.....	.....	...	
LR132004.1	Gouania willdenowi genome assembly, chromosome: 12	Gouania willdenowi	32.2	92.7	72%	144	100.00%	33478699	
LR131988.1	Gouania willdenowi genome assembly, chromosome: 22	Gouania willdenowi	32.2	64.4	86%	144	100.00%	34676616	
.....	.....	.....	.....	.....	.....	.....	.....	...	
CP018162.1	Oryza sativa Indica Group cultivar Shuhui498 chromosome 6 sequence	Oryza sativa Indica Group	32.2	32.2	72%	144	100.00%	32465040	
.....	.....	.....	.....	.....	.....	.....	.....	...	

XM_019790575.1	PREDICTED: Branchiostoma belcheri pre-mRNA-splicing factor SLU7-like (LOC109486685), mRNA	Branchiostoma belcheri	32.2	32.2	72%	144	100.00%	1869
.....	.....	.....	.....	.....	.....	.....	.....	...
XM_008640479.2	PREDICTED: Corvus brachyrhynchos lysosomal associated membrane protein 2 (LAMP2), transcript variant X2, mRNA	Corvus brachyrhynchos	32.2	32.2	72%	144	100.00%	3902
XM_017743113.1	PREDICTED: Corvus brachyrhynchos lysosomal associated membrane protein 2 (LAMP2), transcript variant X1, mRNA	Corvus brachyrhynchos	32.2	32.2	72%	144	100.00%	3750
XM_017502118.1	PREDICTED: Cebus imitator SLU7 homolog, splicing factor (SLU7), transcript variant X2, mRNA	Cebus imitator	32.2	32.2	72%	144	100.00%	2095
.....	.....	.....	.....	.....	.....	.....	.....	...
AP014962.1	Oryza sativa Japonica Group DNA, chromosome 6, cultivar: Nipponbare, complete sequence	Oryza sativa Japonica Group	32.2	32.2	72%	144	100.00%	31248787
CP012614.1	Oryza sativa Indica Group cultivar RP Bio-226 chromosome 6 sequence	Oryza sativa Indica Group	32.2	32.2	72%	144	100.00%	30903862
.....	.....	.....	.....	.....	.....	.....	.....	...
HG975449.1	Solanum pennellii chromosome ch10, complete genome	Solanum pennellii	32.2	62.4	81%	144	100.00%	82529941
.....	.....	.....	.....	.....	.....	.....	.....	...
CP056057.1	Oryza sativa Indica Group cultivar Zhenshan 97 chromosome 6	Oryza sativa Indica Group	32.2	32.2	72%	144	100.00%	32119910
CP054681.1	Oryza sativa Indica Group cultivar Minghui 63 chromosome 6	Oryza sativa Indica Group	32.2	32.2	72%	144	100.00%	31921180
OD789924.1	4_Tbi_b3v08	Timema bartmani	32.2	32.2	72%	144	100.00%	315
OC326021.1	3_Tce_b3v08	Timema cristinae	32.2	32.2	72%	144	100.00%	8757
.....	.....	.....	.....	.....	.....	.....	.....	...
OA603930.1	1_Tdi_b3v08	Timema douglasi	32.2	32.2	72%	144	100.00%	1642
XM_017502117.2	PREDICTED: Cebus imitator SLU7 homolog, splicing factor (SLU7), transcript variant X1, mRNA	Cebus imitator	32.2	32.2	72%	144	100.00%	2045
.....	.....	.....	.....	.....	.....	.....	.....	...
LR877217.1	Acomys russatus genome assembly, chromosome: 6	Acomys russatus	32.2	92.7	86%	144	100.00%	89442464
.....	.....	.....	.....	.....	.....	.....	.....	...
XM_035891949.1	PREDICTED: Egretta garzetta lysosomal associated membrane protein 2 (LAMP2), transcript variant X3, mRNA	Egretta garzetta	32.2	32.2	72%	144	100.00%	3562
XM_035891948.1	PREDICTED: Egretta garzetta lysosomal associated membrane protein 2 (LAMP2), transcript variant X2, mRNA	Egretta garzetta	32.2	32.2	72%	144	100.00%	3639
XM_009636403.2	PREDICTED: Egretta garzetta lysosomal associated membrane protein 2 (LAMP2), transcript variant X1, mRNA	Egretta garzetta	32.2	32.2	72%	144	100.00%	3718
.....	.....	.....	.....	.....	.....	.....	.....	...
AP004741.3	Oryza sativa Japonica Group genomic DNA, chromosome 6, BAC clone:OSJNBb0024N18	Oryza sativa Japonica Group	32.2	32.2	72%	144	100.00%	128739
AP003525.2	Oryza sativa Japonica Group genomic DNA, chromosome 6, PAC clone:P0537F07	Oryza sativa Japonica Group	32.2	32.2	72%	144	100.00%	147724
.....	.....	.....	.....	.....	.....	.....	.....	...
CP056062.1	Oryza sativa Indica Group cultivar Zhenshan 97	Oryza sativa Indica Group	30.2	30.2	68%	570	100.00%	32624309

	chromosome 11							
.....	.....	.....	.....	.....	.....	.....	.....	...
OD565795.1	4_Tbi_b3v08	Timema bartmani	30.2	30.2	68%	570	100.00%	154946
.....	.....	.....	.....	.....	.....	.....	.....	...
OD003075.1	1_Tps_b3v08	Timema poppensis	30.2	30.2	68%	570	100.00%	71863
OC491475.1	3_Tce_b3v08	Timema cristinae	30.2	30.2	68%	570	100.00%	435
OD571540.1	4_Tbi_b3v08	Timema bartmani	30.2	30.2	68%	570	100.00%	26132
OC335513.1	3_Tce_b3v08	Timema cristinae	30.2	30.2	68%	570	100.00%	3456
OD081554.1	1_Tps_b3v08	Timema poppensis	30.2	30.2	68%	570	100.00%	1154
OC466630.1	3_Tce_b3v08	Timema cristinae	30.2	30.2	68%	570	100.00%	472
OC331821.1	3_Tce_b3v08	Timema cristinae	30.2	30.2	68%	570	100.00%	4211
.....	.....	.....	.....	.....	.....	.....	.....	...
OC320933.1	3_Tce_b3v08	Timema cristinae	30.2	30.2	68%	570	100.00%	53721
OC320403.1	3_Tce_b3v08	Timema cristinae	30.2	30.2	68%	570	100.00%	63999
OC318841.1	3_Tce_b3v08	Timema cristinae	30.2	30.2	68%	570	100.00%	111344
.....	.....	.....	.....	.....	.....	.....	.....	...
OC317760.1	3_Tce_b3v08	Timema cristinae	30.2	30.2	68%	570	100.00%	181983
OC316989.1	3_Tce_b3v08	Timema cristinae	30.2	30.2	68%	570	100.00%	317938
OC316749.1	3_Tce_b3v08	Timema cristinae	30.2	30.2	68%	570	100.00%	423377
.....	.....	.....	.....	.....	.....	.....	.....	...
OD573488.1	4_Tbi_b3v08	Timema bartmani	30.2	30.2	68%	570	100.00%	13991
.....	.....	.....	.....	.....	.....	.....	.....	...
OD565227.1	4_Tbi_b3v08	Timema bartmani	30.2	30.2	68%	570	100.00%	203179
OD564842.1	4_Tbi_b3v08	Timema bartmani	30.2	30.2	68%	570	100.00%	268306
OD023040.1	1_Tps_b3v08	Timema poppensis	30.2	30.2	68%	570	100.00%	4464
OD004973.1	1_Tps_b3v08	Timema poppensis	30.2	30.2	68%	570	100.00%	47068
.....	.....	.....	.....	.....	.....	.....	.....	...
OC013708.1	2_Tsi_b3v08	Timema shepardi	30.2	30.2	68%	570	100.00%	7225
OC008934.1	2_Tsi_b3v08	Timema shepardi	30.2	30.2	68%	570	100.00%	21003
OC008543.1	2_Tsi_b3v08	Timema shepardi	30.2	30.2	68%	570	100.00%	23489
OC001153.1	2_Tsi_b3v08	Timema shepardi	30.2	30.2	68%	570	100.00%	186648
OC000460.1	2_Tsi_b3v08	Timema shepardi	30.2	30.2	68%	570	100.00%	287612
OC000021.1	2_Tsi_b3v08	Timema shepardi	30.2	30.2	68%	570	100.00%	691222
OC000017.1	2_Tsi_b3v08	Timema shepardi	30.2	30.2	68%	570	100.00%	703576
OA567862.1	1_Tdi_b3v08	Timema douglasi	30.2	30.2	68%	570	100.00%	79571
.....	.....	.....	.....	.....	.....	.....	.....	...
OA635940.1	1_Tdi_b3v08	Timema douglasi	30.2	30.2	68%	570	100.00%	739
.....	.....	.....	.....	.....	.....	.....	.....	...
OA582280.1	1_Tdi_b3v08	Timema douglasi	30.2	30.2	68%	570	100.00%	4466
.....	.....	.....	.....	.....	.....	.....	.....	...
CP056056.1	Oryza sativa Indica Group cultivar Zhenshan 97 chromosome 5	Oryza sativa Indica Group	30.2	90.7	72%	570	100.00%	30786153
CP056052.1	Oryza sativa Indica Group cultivar Zhenshan 97 chromosome 1	Oryza sativa Indica Group	30.2	30.2	68%	570	100.00%	44754788
CP054680.1	Oryza sativa Indica Group cultivar Minghui 63 chromosome 5	Oryza sativa Indica Group	30.2	90.7	72%	570	100.00%	31307418
CP054676.1	Oryza sativa Indica Group cultivar Minghui 63 chromosome 1	Oryza sativa Indica Group	30.2	30.2	68%	570	100.00%	45027022
.....	.....	.....	.....	.....	.....	.....	.....	...
OD000794.1	1_Tps_b3v08	Timema poppensis	30.2	30.2	68%	570	100.00%	156903
OD572797.1	4_Tbi_b3v08	Timema bartmani	30.2	30.2	68%	570	100.00%	17394
OD569022.1	4_Tbi_b3v08	Timema bartmani	30.2	30.2	68%	570	100.00%	54319
OD568644.1	4_Tbi_b3v08	Timema bartmani	30.2	30.2	68%	570	100.00%	60441
OC468217.1	3_Tce_b3v08	Timema cristinae	30.2	30.2	68%	570	100.00%	469
OD287289.1	1_Tps_b3v08	Timema poppensis	30.2	30.2	68%	570	100.00%	409

OC573845.1	3_Tce_b3v08	Timema cristinae	30.2	30.2	68%	570	100.00%	358
OC318855.1	3_Tce_b3v08	Timema cristinae	30.2	30.2	68%	570	100.00%	110493
OC318171.1	3_Tce_b3v08	Timema cristinae	30.2	30.2	68%	570	100.00%	146552
OC317201.1	3_Tce_b3v08	Timema cristinae	30.2	30.2	68%	570	100.00%	259221
.....	.....	.....	.....	.....	.....	.....	.....	...

Pipistrellus pipistrellus genome assembly, chromosome: 16

GenBank: LR862372.1

FASTA Graphics

Go to:

LOCUS LR862372 16 bp DNA linear MAM **03-AUG-2020**

DEFINITION Pipistrellus pipistrellus genome assembly, chromosome: 16.

ACCESSION LR862372 REGION: 18132203..18132218

VERSION LR862372.1

DBLINK BioProject: PRJEB39564
BioSample: SAMEA994724

KEYWORDS .

SOURCE Pipistrellus pipistrellus (common pipistrelle)

ORGANISM Pipistrellus pipistrellus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Laurasiatheria; Chiroptera; Microchiroptera;
Vespertilionidae; Pipistrellus.

REFERENCE 1

CONSRTM Wellcome Sanger Institute Data Sharing

TITLE Direct Submission

JOURNAL **Submitted (27-JUL-2020) SC, Wellcome Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, United Kingdom**

COMMENT The assembly mPipPip1.1 is based on 62x PacBio data, 34x 10X
Genomics Chromium data, and 58x Arima Hi-C data generated at the
Wellcome Sanger Institute. The Hi-C data was derived from a
separate sample, mPipPip2. The assembly process included the
following sequence of steps: initial PacBio assembly generation
with Falcon-unzip, retained haplotig separation with purge_dups,
10X based scaffolding with scaff10x, Hi-C based scaffolding with
SALSA2, Arrow polishing, two rounds of FreeBayes polishing. The
mitochondrial assembly was produced at The Rockefeller University
using mitoVGP. Finally, the assembly was analysed and manually
improved using gEVAL. Chromosome-scale scaffolds confirmed by the
Hi-C data have been named in order of size.

FEATURES Location/Qualifiers

source 1..16

/organism="Pipistrellus pipistrellus"

/mol type="genomic DNA"

/db_xref="taxon:59474"

/chromosome="16"

ORIGIN

1 **ttcttgtgtg tcattg**

//

ATTENTION:

The sequence "**ttcttgtgtg tcattg**" shown above is the *reverse and mirror sequence* of the sequence "**CAATGACACACAAGAA**" (partial sequence of **Sequence n° 1/3**, "**CTCGGGCAATGACACACAAGAA**"):

TTCTTGTGTGTCATTG

| | | | | | | | | | | | | | | |

AAGAACACACAGTAAC

Pipistrellus pipistrellus genome assembly, chromosome: 16
Sequence ID: LR862372.1 Length: 46231460 Number of Matches: 3
Range 1: 18132203 to 18132218 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
32.2 bits(16)	147	16/16(100%)	0/16(0%)	Plus/Minus

```
Query  7         CAATGACACACAAGAA  22
                 ||||||||||||||||
Sbjct  18132218  CAATGACACACAAGAA  18132203
```

Range 2: 5948763 to 5948778 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
32.2 bits(16)	147	16/16(100%)	0/16(0%)	Plus/Plus

```
Query  6         GCAATGACACACAAGA  21
                 ||||||||||||||||
Sbjct  5948763   GCAATGACACACAAGA  5948778
```

Range 3: 18392280 to 18392295 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
32.2 bits(16)	147	16/16(100%)	0/16(0%)	Plus/Plus

```
Query  7         CAATGACACACAAGAA  22
                 ||||||||||||||||
Sbjct  18392280  CAATGACACACAAGAA  18392295
```

Pipistrellus pipistrellus genome assembly, chromosome: 6

GenBank: LR862362.1

FASTA Graphics

LOCUS LR862362 16 bp DNA linear MAM **03-AUG-2020**

DEFINITION Pipistrellus pipistrellus genome assembly, chromosome: 6.

ACCESSION LR862362 REGION: 71493976..71493991

VERSION LR862362.1

DBLINK BioProject: PRJEB39564

BioSample: SAMEA994724

KEYWORDS .

SOURCE Pipistrellus pipistrellus (common pipistrelle)

ORGANISM Pipistrellus pipistrellus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Laurasiatheria; Chiroptera; Microchiroptera; Vespertilionidae; Pipistrellus.

REFERENCE 1

CONSRTM Wellcome Sanger Institute Data Sharing

TITLE Direct Submission

JOURNAL **Submitted (27-JUL-2020) SC, Wellcome Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, United Kingdom**

COMMENT The assembly mPipPip1.1 is based on 62x PacBio data, 34x 10X Genomics Chromium data, and 58x Arima Hi-C data generated at the Wellcome Sanger Institute. The Hi-C data was derived from a separate sample, mPipPip2. The assembly process included the following sequence of steps: initial PacBio assembly generation with Falcon-unzip, retained haplotig separation with purge dups, 10X based scaffolding with scaff10x, Hi-C based scaffolding with SALSA2, Arrow polishing, two rounds of FreeBayes polishing. The mitochondrial assembly was produced at The Rockefeller University using mitoVGP. Finally, the assembly was analysed and manually improved using gEVAL. Chromosome-scale scaffolds confirmed by the Hi-C data have been named in order of size.

FEATURES Location/Qualifiers

source 1..16

/organism="Pipistrellus pipistrellus"

/mol_type="genomic DNA"

/db_xref="taxon:59474"

/chromosome="6"

ORIGIN

1 **ttgtgtgtca ttgccc**

//

Pipistrellus pipistrellus genome assembly, chromosome: 6
Sequence ID: LR862362.1 Length: 85793147 Number of Matches: 2
Range 1: 71493976 to 71493991 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
32.2 bits(16)	147	16/16(100%)	0/16(0%)	Plus/Minus

```
Query  4         GGGCAATGACACACAA  19
                 ||||||||||||||||
Sbjct  71493991  GGGCAATGACACACAA  71493976
```

Range 2: 16647082 to 16647096 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
30.2 bits(15)	579	15/15(100%)	0/15(0%)	Plus/Minus

```
Query  4         GGGCAATGACACACA  18
                 |||||||||||||||
Sbjct  16647096  GGGCAATGACACACA  16647082
```

9.3 THE SEGMENT OF SEQUENCE N° 1/3 IDENTIFIED AS CORRISPONDING TO THE HIV-1 VIRUS

In the following pages, the result of the BLAST search related to the *significant alignment* that identified a sequence of the ***HIV-1 Virus*** is presented and discussed in great depth.

IV-1 isolate CY064 from Cyprus nonfunctional gag protein (gag) gene, complete sequence; pol protein (pol) gene, partial cds; vif protein (vif), vpr protein (vpr), tat protein (tat), rev protein (rev), and vpu protein (vpu) genes, complete cds; nonfunctional e...

GenBank: FJ388909.1

FASTA Graphics PopSet

Go to:

```
LOCUS       FJ388909                 17 bp    DNA     linear   VRL 24-JUL-2016
DEFINITION  HIV-1 isolate CY064 from Cyprus nonfunctional gag protein (gag)
            gene, complete sequence; pol protein (pol) gene, partial cds; vif
            protein (vif), vpr protein (vpr), tat protein (tat), rev protein
            (rev), and vpu protein (vpu) genes, complete cds; nonfunctional
            envelope glycoprotein (env) gene, complete sequence; and nef
            protein (nef) gene, partial cds.
ACCESSION   FJ388909 REGION: 6661..6677
VERSION     FJ388909.1
KEYWORDS    .
SOURCE      Human immunodeficiency virus 1 (HIV-1)
  ORGANISM  Human immunodeficiency virus 1
            Viruses; Riboviria; Pararnavirae; Artverviricota; Revtraviricetes;
            Ortervirales; Retroviridae; Orthoretrovirinae; Lentivirus.
REFERENCE   1  (bases 1 to 17)
  AUTHORS   Kousiappa,I., Van de Vijver,D.A.M.C. and Kosrtikis,L.G.
  TITLE     Near Full-Length Genetic Analysis of HIV Sequences Derived from
            Cyprus: Evidence of a Highly Polyphyletic and Evolving Infection
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 17)
  AUTHORS   Kousiappa,I., Van de Vijver,D.A.M.C. and Kosrtikis,L.G.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-OCT-2008) Biological Sciences, University of Cyprus,
            75 Kallipoleos Avenue, 20537, Nicosia 1678, Cyprus
FEATURES             Location/Qualifiers
     source          1..17
                     /organism="Human immunodeficiency virus 1"
                     /proviral
                     /mol_type="genomic DNA"
                     /isolate="CY064"
                     /host="Homo sapiens"
```

/db_xref="taxon:11676"
/country="Cyprus"
/collection_date="23-Jun-2005"
gene <1..>17
/gene="tat"
gene <1..>17
/gene="rev"
gene <1..>17
/gene="env"
misc_feature <1..>17
/gene="env"
/note="nonfunctional envelope glycoprotein due to mutation"

ORIGIN

1 ggcaatgaca cacaaga

//

HIV-1 isolate CY064 from Cyprus nonfunctional gag protein (gag) gene, complete sequence; pol protein (pol) gene, partial cds; vif protein (vif), vpr protein (vpr), tat protein (tat), rev protein (rev), and vpu protein (vpu) genes, complete cds; nonfunctional e...

GenBank: FJ388909.1

FASTA Graphics PopSet

Go to:

LOCUS FJ388909 2613 bp DNA linear VRL 24-JUL-2016

DEFINITION HIV-1 isolate CY064 from Cyprus nonfunctional gag protein (gag) gene, complete sequence; pol protein (pol) gene, partial cds; vif protein (vif), vpr protein (vpr), tat protein (tat), rev protein (rev), and vpu protein (vpu) genes, complete cds; nonfunctional envelope glycoprotein (env) gene, complete sequence; and nef protein (nef) gene, partial cds.

ACCESSION FJ388909 REGION: 5446..8058

VERSION FJ388909.1

KEYWORDS .

SOURCE Human immunodeficiency virus 1 (HIV-1)

ORGANISM Human immunodeficiency virus 1

Viruses; Riboviria; Pararnavirae; Artverviricota; Revtraviricetes; Ortervirales; Retroviridae; Orthoretrovirinae; Lentivirus.

REFERENCE 1 (bases 1 to 2613)

AUTHORS Kousiappa,I., Van de Vijver,D.A.M.C. and Kosrtikis,L.G.

TITLE Near Full-Length Genetic Analysis of HIV Sequences Derived from Cyprus: Evidence of a Highly Polyphyletic and Evolving Infection

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 2613)
AUTHORS Kousiappa,I., Van de Vijver,D.A.M.C. and Kosrtikis,L.G.
TITLE Direct Submission
JOURNAL **Submitted (15-OCT-2008) Biological Sciences, University of Cyprus,75 Kallipoleos Avenue, 20537, Nicosia 1678, Cyprus**

FEATURES Location/Qualifiers
source 1..2613
/organism="Human immunodeficiency virus 1"
/proviral
/mol_type="genomic DNA"
/isolate="CY064"
/host="Homo sapiens"
/db_xref="taxon:11676"
/country="Cyprus"
/collection_date="23-Jun-2005"
gene <1..2266
/gene="tat"
CDS <2176..2266
/gene="tat"
/codon start=1
/product="tat protein"
/protein id="ACR55221.1"
/translation="MDPVDPNLEPWNHPGSRPTTPCSKCYCKACCFHCLVCFQNKGLG
ISYGRRKRRRRRGTPQSCEDHQNPVPKQPTPQFKGISTDPEESKKKVESKAETDRFD"
gene <1..2471
/gene="rev"
CDS <2176..2471
/gene="rev"
/codon_start=1
/product="rev protein"
/protein_id="ACR55222.1"
/translation="MAGGSGDGEEELLRAVRIIKILYQSNPPPSSRGSRQTRKNRRRR
WRARQRQIDSISERVLSACLGRPPEPVPLQLPPLERLHLDCSEDCGTSGTQQSQGVET
GVGRPQVSGESSDILGEGDKK"
gene <1..83
/gene="vpu"
CDS <1..83
/gene="vpu"
/codon_start=3
/product="vpu protein"
/protein id="ACR55223.1"
/translation="MSPLEICAIAGLVVALILAIVVWTIVGIEIKQVLKQRKIDRLIE
RIRERAEDSGN**ESDGDAEDLSLLVEMGHDNLGYDNNL**"
gene 1..2613

/gene="env"

misc_feature 1..2613

/gene="env"

/note="nonfunctional envelope glycoprotein due to mutation"

ORIGIN

```
      1 atgagagtga tggggacgca gaggacttgt ccttgcttgt ggagatgggg catgataatc
     61 ttgggtatga taataatttg   tagtgttaca ggaaacttgt gggttactgt ctactatggg
    121 gtacctgtgt ggaaagatgc agacaccacc ttattttgtg catcagatgc taaagcatat
    181 gacacagaag tgcataatgt ctgggctaca catgcctgtg tgcccacaga ccccaaccca
    241 caagagataa atttggagaa tgtgacagaa gagtttaaca tgtggaaaaa taacatggta
    301 gagcagatgc atacagatat aatcagtcta tgggaccaaa gtctaaagcc atgtgtacag
    361 ttaacccctc tctgcgttac gttaaattgt agagatgatg acgacatcaa tgtcgccaat
    421 agcacagaca ccactaacga agtcaaaaac ggatctgcca atatcacaga tgaaaaaaac
    481 tgctctttca ctatgaccac tgacctaata gataagctaa ttgaagtatt ttctgttttt
    541 tatagacttg atgttgtaca tattgataaa aatagtaata acagtgggta tagattaata
    601 aattgaaata ccacaaccgt tacacaggct ggtccaaagg tatcctttga tccaattccc
    661 atacattatt gggcccctgc tggttttgcg attctaaagt gtattgagaa ggacttcaat
    721 ggaacaggaa catgcaagaa tgtcagctca atacactgta cacatggaat caagccagta
    781 gtatcaactc aactgctgtt aaatggcagt ctggcagaag gagaggtaaa gattagatct
    841 gaaaatctca caaacaatgc caaaaccata atagtacagt ttgttgagcc tgtacacatt
    901 acttgtatca gaccttacaa caatacaaga aaaggtatac atataggacc agggcaagca
    961 ttctatgcaa tgagtaacat aataggagat ataagacaag cacattgtaa cgtcagtaaa
   1021 acaaaatgga atgcaacgtt atcccaggta gccaaacaat tagggaataa gtttaacaac
   1081 agaacaataa tctttaatag ctcctcagga ggggatatag aacttacaag acatagtttt
   1141 aactgtggag gagagttttt ctattgtaat acatcaggcc tgtttaatag cacttggaat
   1201 aatgacacct ggaatggcaa tgacacacaa gagtcaaatg aaacctcgtc aaatggcacc
   1261 ataattccac catgcaaaat aaaacaaatt gtaaatatgt ggcagagagt acaacaagca
   1321 atgtatcccc ctctcgtcga aggagtaata aggtgtgaat ccatcattac aggactacta
   1381 tcaacaagag atggtgggat taacagcagg acaaatgaaa ccttcagacc tgggggagga
   1441 gatatgaggg acaattggag aagtgaacta tataaatata aggtagtaaa aattgaacca
   1501 ctaggagtag cacccaccac ggcaaagaga agagtggtgg agagagaaaa aagagcagct
   1561 ggactgggag ttatgttcct tgggttctta ggagcagccg gaagcactat gggcgcggcg
   1621 tcaataacgc tgacggtaca ggccagacag ttattatctg gtatagtgca acagcaaagc
   1681 aatttgctga aggctataga ggctcaacag cacctgttga gactcacggt ctggggcatt
   1741 aaacagctcc aggcaagagt cctggctgtg gaaagatacc taaaggatca acagctccta
   1801 ggaatttggg gctgctctgg aaaactcatc tgcaccacta ctgtgccctg gaactctagt
   1861 tggagtaata gatctcagga ggaaatatgg aataacatga cctggctgca atgggagaaa
   1921 gaaattggca attacacagg tataatatat agactacttg aagaatcgca gaaccagcag
   1981 gaaaagaatg aacaagactt attggcattg gacaagtggg caagtctgtg gaattggttt
   2041 agcatatcaa actggctatg gtatattaaa atatttataa tgatagtagg aggcttaata
   2101 ggattaagaa tagtttttac tgtgctttct ataataaata gagttaggca gggatactca
   2161 cccttgtcgt ttcagaccca ccccccagtt caaggggatc tcgacagacc cggaagaatc
   2221 gaagaagaag gtggagagca aggcagagac agatcgattc gattagtgag cgggttctta
```

```
2281 gcgcttgcct gggacgacct ccggagcctg tgcctcttca gctaccaccg cttgagagac
2341 ttcatcttga ttgcagcgag gactgtggaa cttctgggac acagcagtct caaggggttg
2401 agactggggt gggaaggcct caagtatctg gggaatcttc tgatatattg ggggaaggag
2461 ataaaaaata gtgctgttaa tttggttgat actatagcaa tagcagtagc tggctggaca
2521 gatagggtta tagaaatagg acaaagagtt ggtagagcta ttctcaacat acctagaaga
2581 attagacagg gcttagaaag ggctttgcta taa  //
```

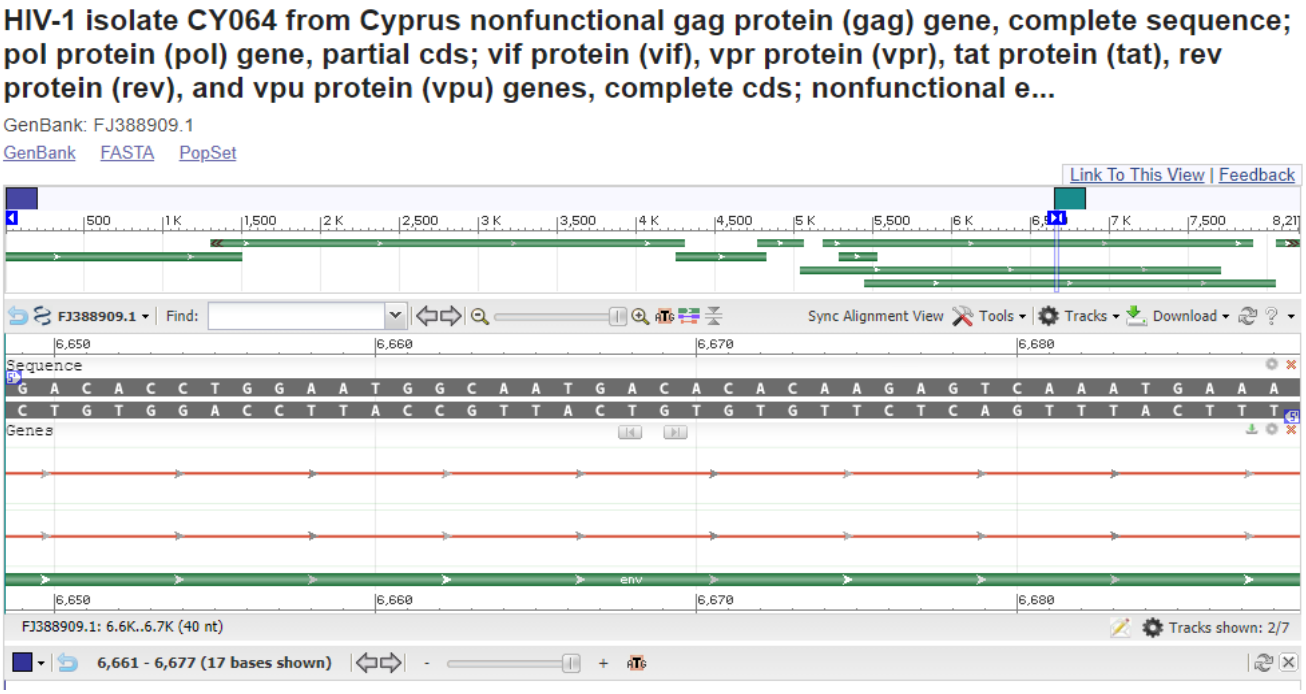
HIV-1 isolate CY064 from Cyprus nonfunctional gag protein (gag) gene, complete sequence; pol protein (pol) gene, partial cds; vif protein (vif), vpr protein (vpr), tat protein (tat), rev protein (rev), and vpu protein (vpu) genes, complete cds; nonfunctional envelope glycoprotein (env) gene, complete sequence; and nef protein (nef) gene, partial cds

Sequence ID: FJ388909.1 **Length: 8211 Number of Matches: 1**

Range 1: 6661 to 6677 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
34.2 bits(17)	35	17/17(100%)	0/17(0%)	Plus/Plus

```
Query  5     GGCAATGACACACAAGA  21
             |||||||||||||||||
Sbjct  6661  GGCAATGACACACAAGA  6677
```



As can be seen on page 179, the red colored letters (within the larger sequence of letters highlighted in yellow) are part ("**GGCAATGACACACAAGA**") of **Sequence n° 1/3** ("**CTCGGGCAATGACACACAAGAA**"). This part of the sequence has been identified as corresponding to the ***HIV-1 Virus*** and refers to a **"misc_feature" region** (i.e. a region of biological interest which cannot be described by any other feature key, i.e. a new or rare feature) defined as "**nonfunctional envelope glycoprotein due to mutation**" of the **Env gene** [a *viral gene* that codes for the protein that forms the ***viral envelope*** [(https://en.wikipedia.org/wiki/Env_(gene)) and (https://en.wikipedia.org/wiki/Viral_envelope)].

A **viral envelope** is the outermost layer of many types of viruses. It protects the genetic material in their life-cycle when traveling between ***host cells***. The envelopes are typically derived from *portions of the host cell membranes*, but include some *viral glycoproteins*. **They may help viruses avoid the** host immune system. **Glycoproteins on the surface of the envelope** serve to identify and bind to *receptor sites* on the host's membrane. The *viral envelope* then fuses with the host's membrane, allowing the *capsid* (another *protein layer*, between the *envelope* and the *genome*) and *viral genome* to **enter and infect the host**. The cell from which a virus buds often dies or is weakened, and sheds more **viral particles** for an extended period (https://en.wikipedia.org/wiki/Viral_envelope).

The **lipid bilayer** *envelope* of these viruses is relatively sensitive to *desiccation*, *heat*, *soap* and *detergents*, therefore these viruses are easier to sterilize than *non-enveloped viruses*, have limited survival outside host environments, and typically must transfer directly from host to host. Enveloped viruses possess great adaptability and can change in a short time in order to evade the immune system. Enveloped viruses can cause persistent infections (https://en.wikipedia.org/wiki/Viral_envelope).

Among **enveloped viruses** (https://en.wikipedia.org/wiki/Viral_envelope), among others, we include *DNA Virus* (Herpesvirus), *Retrovirus* (such as *HIV*) and *Coronavirus* [*RNA Virus* (https://en.wikipedia.org/wiki/Coronavirus)]. **The enveloped viruses have great adaptability, they can change in a short time to evade the immune system and cause persistent infection** (https://en.wikipedia.org/wiki/Viral_envelope), those that occur because the body is unable to clear the organism after the initial infection (https://en.wikipedia.org/wiki/Infection#Persistent_infection).

The mature product of the ***Env gene*** is the **viral spike protein** (whose notoriety has reached the general public because it has been mentioned often since the beginning of the ***SARS-CoV-2 pandemic***), which has **two main parts**: *the surface protein* (SU) and the *transmembrane protein* (TM). The tropism (https://en.wikipedia.org/wiki/Tissue_tropism) of the virus (consisting of the cells and tissues of a host that support the growth of a particular virus or bacterium) is determined by the **SU protein** domain because it is responsible for the *receptor-binding function of the virus*. **The SU domain therefore determines the specificity of the virus for a single receptor molecule** (https://en.wikipedia.org/wiki/Env_(gene)).

Observing the **sequence of bases** identified by the BLAST search as corresponding to the ***HIV-1 Virus***, we note that the initial part (see the blue colored bases on page 179) of this **sequence** refers to the **final part** of the ***Vpu gene*** (**CDS <1..83 / gene = "vpu" / codon_start = 3 / product = "vpu protein"**) which encodes the ***Vpu protein***.

Vpu is an *accessory protein* that in HIV is encoded by the ***vpu gene***. **Vpu** stands for "*Viral Protein U*" (https://en.wikipedia.org/wiki/Vpu_protein). The Vpu protein acts in the degradation of CD4 in the endoplasmic reticulum (I remind you that HIV destroy *CD4+ T lymphocytes* and *inhibit the cell-mediated immune response*) and in the enhancement of **virion** (complete viral particles consisting of *nucleic acid* surrounded by a protective layer of proteins called *capsid*) release from the **plasma membrane of infected cells**. *Vpu* induces the degradation of the CD4 viral receptor and therefore participates in the general downregulation of CD4 expression during the course of HIV infection. Vpu-mediated CD4 degradation is thought to prevent CD4-Env binding in the endoplasmic reticulum in order to facilitate proper Env assembly into virions. **It is found in the membranes of infected cells, but not the virus particles themselves** (https://en.wikipedia.org/wiki/Vpu_protein).

It is very important to point out that there are **structural similarities between *Vpu* and another small viral protein, *M2*, encoded by Influenza A Virus**. *Vpu* also permeabilizes **membranes of bacteria** and, therefore, it is considered a member of the ***Viroporins*** family, small and usually hydrophobic *multifunctional viral proteins* that modify cellular membranes, thereby facilitating virus release from infected cells (https://en.wikipedia.org/wiki/Viroporin).
The ***Viroporins*** are found in a variety of viral genomes but are particularly common in **RNA viruses**. Many viruses that cause human disease express *viroporins*. These viruses include hepatitis C virus, **HIV-1**, **influenza A virus**, poliovirus, respiratory syncytial virus, and **SARS-CoV** (https://en.wikipedia.org/wiki/Viroporin). Most viroporins are not essential for **viral replication**, but their absence significantly reduces the efficiency of viral propagation (viroporins can also be considered *virulence factors*). In some cases the *membrane permeabilization effects of viroporins* activate the *inflammasome*, a *protein complex* associated with activation of **innate immunity** which, when **overactive**, can cause **disease symptoms**.
Le ***Viroporine*** often considered to be appealing drug targets for development of antiviral drugs, but the clinical usage is relatively rare and controversial (https://en.wikipedia.org/wiki/Viroporin).

9.4 PREDICTION OF THE CONSTITUTIVE CODONS OF THE SEGMENT OF THE SEQUENCE N ° 1/3 AND DISCUSSION ON THEIR HYPOTHETICAL PRODUCTS

Scrolling down up to **line 1201** of the sequence shown on page x we encounter the red colored letters (**GGCAATGACACACAAGA**, shown below) which are part of **Sequence n° 1/3** (**CTCGGGCAATGACACACAAGAA**).

```
1201          gaatggcaa tgacacacaa gagtcaaatg aaacctcgtc
        aaatggcacc
1261    ataa
```

In the section "Objectives of chapter IX" we mentioned that the proposal of a **HYPOTHETICAL prediction of the constitutive codons of this segment of Sequence n° 1/3 and of those adjacent to it** (if such *codons* become coding for proteins).
Below, there is the ***prediction of the constitutive codons*** and the ***complete sequence of their corresponding amino acids***.

```
START
atg gca atg aca cac aag agt caa atg aaa cct cgt caa
 M   A   M   T   H   K   S   Q   M   K   P   R   Q

atg gca cca taa
 M   A   P  STOP
```

In the following pages ***different BLAST searches*** performed on different **amino acid sequences** (formed by segments of the *constitutive codons*) are shown and the ***main results obtained*** are discussed.

9.5 SEQUENCE OF AMINO ACIDS "M A M T H"

This paragraph shows some results of the BLAST search carried out on the sequence of amino acids "**M A M T H**".

Query: "M A M T H" (atg gca atg aca cac)

Search Parameters	
Program	blastp
Word size	2
Expect value	200000
Hitlist size	**1000**
Gapcosts	9,1
Matrix	PAM30
Filter string	F
Genetic Code	1
Window Size	40
Threshold	11
Composition-based stats	0

1000 sequences selected:

Accession Seq	Description	Comomm Name	Max Score	Total Score	Query C.	E value	Per. Id.	Acc. Len
.....	.....	.....	.....	.....	.....	.....	.....	.....
EFE8358987.1	DUF1317 family protein [Escherichia coli]	Escherichia coli	21.8	21.8	100%	10826	100.00%	16
.....	.....	.....	.....	.....	.....	.....	.....	.....
AKQ22649.1	malonyl CoA-acyl carrier protein transacylase [Scytonema sp. PCC 10023]	Scytonema sp. PCC 10023	21.8	21.8	100%	11626	100.00%	8106
.....	.....	.....	.....	.....	.....	.....	.....	.....
WP_155743632.1	SDR family NAD(P)-dependent oxidoreductase [Scytonema sp. UIC 10036]	Scytonema sp. UIC 10036	21.8	21.8	100%	11626	100.00%	8011
AVI26425.1	polyketide synthase [Candidatus Entotheonella serta]	Candidatus Entotheonella serta	21.8	21.8	100%	11626	100.00%	7991
.....	.....	.....	.....	.....	.....	.....	.....	.....
XP_017719480.1	PREDICTED: WD repeat-containing protein 66 isoform X1 [Rhinopithecus bieti]	Rhinopithecus bieti	21.8	21.8	100%	11679	100.00%	1160
.....	.....	.....	.....	.....	.....	.....	.....	.....
XP_030795051.1	cilia- and flagella-associated protein 251 [Rhinopithecus roxellana]	Rhinopithecus roxellana	21.8	21.8	100%	11679	100.00%	1149
.....	.....	.....	.....	.....	.....	.....	.....	.....
XP_030776188.1	zinc finger protein 658 [Rhinopithecus roxellana]	Rhinopithecus roxellana	21.8	21.8	100%	11685	100.00%	1059
.....	.....	.....	.....	.....	.....	.....	.....	.....

DUF1317 family protein [Escherichia coli]
Sequence ID: EFE8358987.1 Length: 16 Number of Matches: 1
Range 1: 1 to 5 GenPept Graphics

Score	Expect	Identities	Positives	Gaps
21.8 bits(44)	10826	5/5(100%)	5/5(100%)	0/5(0%)

```
Query  1  MAMTH  5
          MAMTH
Sbjct  1  MAMTH  5
```

Protein of unknown function (DUF1317)

This family consists of several **hypothetical bacterial and phage proteins** of around 60 residues in length. **The function of this family is unknown**
(https://www.ncbi.nlm.nih.gov/Structure/cdd/PF07026).

malonyl CoA-acyl carrier protein transacylase [Scytonema sp. PCC 10023]
Sequence ID: AKQ22649.1 Length: 8106 Number of Matches: 1
Range 1: 7540 to 7544 GenPept Graphics

Score	Expect	Identities	Positives	Gaps
21.8 bits(44)	11626	5/5(100%)	5/5(100%)	0/5(0%)

```
Query  1     MAMTH  5
             MAMTH
Sbjct  7540  MAMTH  7544
```

malonyl CoA-acyl carrier protein transacylase

A **[acyl-carrier-protein] S-malonyltransferase** is an enzyme that belongs to the family of **transferases** [class of enzymes that enact the transfer of specific functional groups from one molecule (called the donor) to another (called the acceptor). See also what is written on pages 233 and 253 on ***glicosyltransferase*** and on page 241 on ***immunoglobulin E-set***, specifically those *acyltransferases* transferring groups other than aminoacyl groups. This enzyme participates in fatty **acid biosynthesis** (https://en.wikipedia.org/wiki/(acyl-carrier-protein)_S-malonyltransferase).
The protein has an **α/β type** architecture, but the fold is unique. The active site inferred from the location of the catalytic **Ser92** (*serine 92*) contains a typical nucleophilic elbow as observed in **α/β hydrolase**. **Serine 92** is *hydrogen* bonded *to His 201* in a fashion similar to various **serine hyrdolases**. The enzymes that participates in **fatty acid biosynthesis** are attractive **targets for antibacterials**. sono interessanti **bersagli per gli antibatterici**. **FabD** (***Acyl-Carrier-Protein S-Malonyltransferase***) is a reasonable target; however, no ***FabD inhibitors*** have been reported in the literature and review articles on this topic. The simple structure and acidity of *malonate* seem to permit few approaches to synthesizing derivatives (acting as **potential inhibitors**) that *retain the character of the molecule* (https://en.wikipedia.org/wiki/(acyl-carrier-protein)_S-malonyltransferase).

SDR family NAD(P)-dependent oxidoreductase [Scytonema sp. UIC 10036]
Sequence ID: WP_155743632.1 Length: 8011 Number of Matches: 1

See 1 more title(s) See all Identical Proteins(IPG)

Related Information
Identical Proteins-Identical proteins to WP_155743632.1
Range 1: 7471 to 7475 GenPept Graphics

Score	Expect	Identities	Positives	Gaps
21.8 bits(44)	11626	5/5(100%)	5/5(100%)	0/5(0%)

```
Query  1     MAMTH  5
             MAMTH
Sbjct  7471  MAMTH  7475
```

SDR family NAD(P)-dependent oxidoreductase

The **SDRs** (***Short-Chain Dehydrogenase/Reductase and Related Enzymes***) represent one of the *oldest protein families* and contribute to essential functions and interactions of all forms of life (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2896744/, http://pfam.xfam.org/family/PF00106.25). Most of these enzymes in this family are known to be ***NAD-*** or ***NADP-dependent oxidoreductases***. An ***oxidoreductase*** is an enzyme that catalyzes the *transfer of electrons* from one molecule, the *reductant*, also called the electron donor, to another, the oxidant, also called the *electron acceptor*. This group of enzymes usually utilizes **NADP+** (*Nicotinamide adenine dinucleotide phosphate*) or **NAD+** as cofactors. ***Transmembrane oxidoreductases*** create electron transport chains in bacteria, chloroplasts and mitochondria, including respiratory complexes I, II and III. Some others can associate with *biological membranes* as *peripheral membrane proteins* or be anchored to the membranes through a single transmembrane helix (https://en.wikipedia.org/wiki/Oxidoreductase). Various properties of ***NAD(P)H-dependent oxidoreductases*** have been targeted by **protein engineering** and/or **directed evolution approaches** to make them more suitable for applications. These include *enhancing the kinetic properties of enzymes*, *increasing their stability* and *modifying the specificity of enzymes for their substrate or cofactor* (https://www.sciencedirect.com/science/article/pii/S157096391730273X?via%3Dihub).

polyketide synthase [Candidatus Entotheonella serta]
Sequence ID: AVI26425.1 Length: 7991 Number of Matches: 1
Range 1: 7424 to 7428 GenPept Graphics

Score	Expect	Identities	Positives	Gaps
21.8 bits(44)	11626	5/5(100%)	5/5(100%)	0/5(0%)

```
Query  1     MAMTH  5
             MAMTH
Sbjct  7424  MAMTH  7428
```

Polyketide synthases (PKSs)

Polyketide synthases (PKSs) are a family of *multi-domain enzymes or enzyme complexes* that produce ***polyketides***, a large class of **secondary metabolites**, in bacteria, fungi, plants, and a few animal lineages. **The biosyntheses of polyketides share striking similarities with fatty acid biosynthesis** (https://en.wikipedia.org/wiki/Polyketide_synthase). ***Polyketide synthases*** are an important source of *naturally* occurring *small molecules* used for **chemotherapy**. Many of the commonly used **antibiotics**, such as **tetracycline** and **macrolides**, are produced by ***polyketide synthases***.

[**Tetracycline** is used to treat a number of infections, his includes acne, cholera, brucellosis, plague, malaria, syphilis and certain rare or **exotic infections**, including **anthrax**. The mechanism of action for the ***antibacterial effect of tetracyclines*** relies on *disrupting protein translation in bacteria, thereby damaging the ability of microbes to grow and repair*; however, protein translation is also disrupted in eukaryotic mitochondria leading to effects that may confound experimental results (https://en.wikipedia.org/wiki/Tetracycline). In ***genetic engineering***, **tetracycline** is used in **transcriptional activation**. It has been used as an engineered "**control switch**" able to develop a **retrovirus** that induced a particular type of *leukemia* in mice, and could then switch the cancer "**on**" and "**off**" through tetracycline *administration* (https://en.wikipedia.org/wiki/Tetracycline)].

[The **macrolides** belong to the ***polyketide*** class of natural products. The *first macrolide* discovered was **erythromycin**. Some macrolides have ***antibiotic*** or ***antifungal activity*** and are used as pharmaceutical drugs (mostly to treat ***infections*** caused by ***Gram-positive bacteria***), including azithromycin, **ZITROMAX** (https://www.drugs.com/zithromax.html), and clarithromycin, stemmed from chemically modifying erythromycin. **Macrolides** are **bacteriostatic** in that they **suppress or inhibit bacterial growth rather than killing bacteria completely**. The mechanism of action of macrolides is **inhibition of bacterial protein biosynthesis**, and they are thought to do this by *preventing peptidyltransferase* from adding the growing *peptide* attached to tRNA to the next *amino acid*, as well as inhibiting *bacterial ribosomal translation*. Another potential mechanism is *premature dissociation of the peptidyl-tRNA from the ribosome*. Macrolides are actively concentrated *within leukocytes*, and thus are *transported into the site of infection* (https://en.wikipedia.org/wiki/Macrolide)].

Other industrially important ***polyketides*** are *sirolimus* (immunosuppressant), *lovastatin* (anticholesterol drug), and *epothilone B* (anticancer drug) (https://en.wikipedia.org/wiki/Polyketide_synthase).

PREDICTED: WD repeat-containing protein 66 isoform X1 [Rhinopithecus bieti]
Sequence ID: XP_017719480.1 Length: 1160 Number of Matches: 1

See 1 more title(s) See all Identical Proteins(IPG)

Related Information
Gene-associated gene details
Genome Data Viewer-aligned genomic context
Identical Proteins-Identical proteins to XP_017719480.1
Range 1: 354 to 358 GenPept Graphics

Score	Expect	Identities	Positives	Gaps
21.8 bits(44)	11679	5/5(100%)	5/5(100%)	0/5(0%)

```
Query  1    MAMTH  5
            MAMTH
Sbjct  354  MAMTH  358
```

WD repeat-containing protein 66

WD repeat-containing protein 66 (WDR66) o ***Cilia and flagella associated protein 251 (CFAP251)*** (https://www.ncbi.nlm.nih.gov/gene/108523392) is present in **sperm** and *functions in flagellar motility and* ***axoneme*** *structure* (https://www.omim.org/entry/618146) [An **axoneme** is the microtubule-based cytoskeletal structure that forms the core of a cilium or flagellum (*https://en.wikipedia.org/wiki/Axoneme*)]. *Knockdown of Wdr66* in *T. brucei* led to **decreased cell proliferation** compared with wildtype. *Knockdown of Wdr66* affected *flagellar motility* and resulted in *disorganization of the axoneme*. It was demonstrated that the *C-terminal domain* is required for protein function and that *WDR66* was required for **formation of the mitochondrial sheath around the axoneme during flagellogenesis** (https://www.omim.org/entry/618146).

cilia- and flagella-associated protein 251 [Rhinopithecus roxellana]
Sequence ID: XP_030795051.1 Length: 1149 Number of Matches: 1
Related Information
Gene-associated gene details
Genome Data Viewer-aligned genomic context
Range 1: 352 to 356 GenPept Graphics

Score	Expect	Identities	Positives	Gaps
21.8 bits(44)	11679	5/5(100%)	5/5(100%)	0/5(0%)

```
Query  1    MAMTH  5
            MAMTH
Sbjct  352  MAMTH  356
```

See what is written above.

zinc finger protein 658 [Rhinopithecus roxellana]
Sequence ID: XP_030776188.1 Length: 1059 Number of Matches: 1

See 2 more title(s) See all Identical Proteins(IPG)

Related Information
Gene-associated gene details
Genome Data Viewer-aligned genomic context
Identical Proteins-Identical proteins to XP_030776188.1
Range 1: 292 to 296 GenPept Graphics

Score	Expect	Identities	Positives	Gaps
21.8 bits(44)	11685	5/5(100%)	5/5(100%)	0/5(0%)

```
Query  1    MAMTH  5
            MAMTH
Sbjct  292  MAMTH  296
```

Zinc Finger Protein 658

ZNF658 ***(Zinc Finger Protein 658)*** is a protein coding gene. Among its related pathways are ***gene expression*** and ***Herpes simplex virus 1 infection***. Gene Ontology (GO) annotations related to this gene include ***nucleic acid binding and DNA-binding transcription factor activity***. An important paralog of this gene is ZNF33B. **ZNF658** *mediates transcriptional repression in response to zinc*. ***Represses several genes***, including SLC30A5, SLC30A10 and CBWD1, by binding to the zinc transcriptional regulatory element (ZTRE) (5'-C[AC]C[TAG]CC[TC]-N(0-50)-[GA]G[ATC]G[TG]G-3') found in the promoter region. May play a role in the ***control of ribosome biogenesis***, regulating predominantly ***rRNA levels***, as well as those of several ***ribosomal proteins***, thus coordinating this highly zinc-demanding process with the available zinc supply (https://www.genecards.org/cgi-bin/carddisp.pl?gene=ZNF658). On the **Zinc Finger Protein**, please also read what is reported on page 232.

From the results shown in the previous pages it was observed that the **M A M T H** sequence **did not show** *significant alignments* with sequences of ***Bats***, ***Pangolins*** or other "*organisms*" considered important for the goals of this study.

9.6 THE MODIFICATION OF THE SEQUENCE OF AMINO ACIDS "M A M T H" IN THE SEQUENCE OF AMINO ACIDS "M V M T H"

Observing the **M A M T H** sequence I noticed the presence of **two amino acids** of **Methionine (M)** whose positions in the sequence reminded me of a short segment of the sequences (already shown in the previous chapters and which I propose again below) of the "**Bat SARS-like coronavirus"** and the "**Severe acute respiratory syndrome coronavirus 2**".

Bat SARS-like coronavirus isolate bat-SL-CoVZC45, complete genome
Sequence ID: MG772933.1 **Length: 29802 Number of Matches: 1**
Range 1: 15418 to 15439 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
44.1 bits(22)	0.036	22/22(100%)	0/22(0%)	Plus/Plus

```
CDS: Putative 1       1          E  M  V  M  C  G  G
Query                 1      GTGAGATGGTCATGTGTGGCGG  22
                             ||||||||||||||||||||||
Sbjct                 15418  GTGAGATGGTCATGTGTGGCGG  15439
CDS:non-structural p  5052   S  E  M  V  M  C  G  G
```

Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/CA-CZB-11679/2020 ORF1ab polyprotein (ORF1ab), ORF1a polyprotein (ORF1ab), surface glycoprotein (S), ORF3a protein (ORF3a), envelope protein (E), membrane glycoprotein (M), ORF6 protein (ORF6), ORF7a protein (ORF7a), ORF7b (ORF7b), ORF8 protein (ORF8), nucleocapsid phosphoprotein (N), and ORF10 protein (ORF10) genes, complete cds
Sequence ID: MW276628.1 Length: 29741 Number of Matches: 1
Range 1: 15377 to 15398 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
40.1 bits(20)	0.59	21/22(95%)	0/22(0%)	Plus/Plus

```
CDS: Putative 1       1          E  M  V  M  C  G  G
Query                 1      GTGAGATGGTCATGTGTGGCGG  22
                             ||||  ||||||||||||||||
Sbjct                 15377  GTGARATGGTCATGTGTGGCGG  15398
CDS:ORF1ab polyprote  5056   S  E  M  V  M  C  G  G
ATTENTION:  R = A o G
```

Intrigued by this "coincidence", I did an in-depth BLAST search on the *5 amino acid* sequence "**M V M C G**" of the "**Bat SARS-like coronavirus"** and "**Severe acute respiratory syndrome coronavirus 2**". Leaving aside the sequences already mentioned, all the other *significant alignments* found refer to the following sequences of the ***Bacterium of the Pseudomonas species***:

non-ribosomal peptide synthetase [Pseudomonas sp. MWU13-2860]
Sequence ID: RBH56499.1 Length: 10329 Number of Matches: 1
Range 1: 9435 to 9439 GenPept Graphics

Score	Expect	Identities	Positives	Gaps
22.3 bits(45)	8256	5/5(100%)	5/5(100%)	0/5(0%)

```
Query  1     MVMCG  5
             MVMCG
Sbjct  9435  MVMCG  9439
```

Non-ribosomal peptide synthetase

Nonribosomal peptides (**NRP**) are a class of *peptide secondary metabolites*, usually produced by microorganisms like *bacteria* and *fungi*. **Nonribosomal peptides** are also found in higher organisms, such as *nudibranchs*, but are thought to be made by **bacteria** inside these organisms. While there exist a wide range of peptides that are not synthesized by *ribosomes*, the term ***nonribosomal peptide*** typically refers to a very specific set of these (https://en.wikipedia.org/wiki/Nonribosomal_peptide). **Nonribosomal peptides** are synthesized by **nonribosomal peptide synthetases**, which, unlike the ribosomes, are **independent of messenger RNA**. Each nonribosomal peptide synthetase can synthesize only one type of peptide. On occasion, dehydration is performed on ***serines***, resulting in ***dehydroalanine***. **Nonribosomal peptides** are a very diverse family of natural products with an extremely broad range of ***biological activities and pharmacological properties***. They are often toxins, siderophores, or pigments. Nonribosomal peptide antibiotics, cytostatics, and immunosuppressants are in commercial use (https://en.wikipedia.org/wiki/Nonribosomal_peptide).

non-ribosomal peptide synthase/polyketide synthase [Pseudomonas asplenii]
Sequence ID: WP_121140405.1Length: 10302Number of Matches: 1

See 4 more title(s) See all Identical Proteins(IPG)

Related Information
Identical Proteins-Identical proteins to WP_121140405.1
Range 1: 9408 to 9412 GenPept Graphics

Score	Expect	Identities	Positives	Gaps
22.3 bits(45)	8256	5/5(100%)	5/5(100%)	0/5(0%)

```
Query  1     MVMCG  5
             MVMCG
Sbjct  9408  MVMCG  9412
```

Polyketide synthases (**PKSs**)

See what's already written on page 187.

non-ribosomal peptide synthase/polyketide synthase [Pseudomonas fuscovaginae]
Sequence ID: WP_081354542.1 Length: 10286 Number of Matches: 1

See 1 more title(s) See all Identical Proteins(IPG)

Related Information
Identical Proteins-Identical proteins to WP_081354542.1
Range 1: 9390 to 9394 GenPept Graphics

Score	Expect	Identities	Positives	Gaps
22.3 bits(45)	8256	5/5(100%)	5/5(100%)	0/5(0%)

```
Query  1     MVMCG  5
             MVMCG
Sbjct  9390  MVMCG  9394
```

non-ribosomal peptide synthase/polyketide synthase [Pseudomonas asplenii]
Sequence ID: WP_090208474.1Length: 10286Number of Matches: 1

See 1 more title(s) See all Identical Proteins(IPG)

Related Information
Identical Proteins-Identical proteins to WP_090208474.1
Range 1: 9390 to 9394 GenPept Graphics

Score	Expect	Identities	Positives	Gaps
22.3 bits(45)	8256	5/5(100%)	5/5(100%)	0/5(0%)

```
Query  1     MVMCG  5
             MVMCG
Sbjct  9390  MVMCG  9394
```

JesC [Pseudomonas sp. QS1027]
Sequence ID: QMS47800.1 Length: 10276 Number of Matches: 1
Range 1: 9382 to 9386 GenPept Graphics

Score	Expect	Identities	Positives	Gaps
22.3 bits(45)	8256	5/5(100%)	5/5(100%)	0/5(0%)

```
Query  1     MVMCG  5
             MVMCG
Sbjct  9382  MVMCG  9386
```

Bacterial non-ribosomal peptide synthetase synthesizing jessenipeptin

Jessenipeptina is a nonribosomal cyclic lipopeptide, and mupirocin, a known polyketide antibiotic. Both compounds are quorum-sensing regulated and display potent synergistic inhibitory activity against clinically relevant methicillin-resistant *Staphylococcus aureus* (MRSA) (https://www.pnas.org/content/115/15/3758).

Regarding the characteristics of the ***Pseudomonas Bacterium*** and its significant implications in ***lung infections***, I refer to the specialized literature and to an original and very articulated study of mine [DNA or RNA Sequence Analysis and Modification through T_T_E_S_ (Chapter I° - First Part) and page 302 in DNA or RNA Sequence Analysis and Modification through T_T_E_S_ (Chapter I° - Second Part)].

At this point, to discover any links between the "**M V M C G**" sequence and the "**M A M T H**" sequence, I replaced the amino acid **Alanine** (codon **GCA** in the segment "**ATGGCAATGACACAC**" of the **Sequence n° 1/3**) with the amino acid **Valine** (replacing the base in the **GCA** codon "**C**" with the base "**T**"), obtaining the sequence "**M V M T H**" and the sequence "**ATGGTAATGACACAC**".

Two results from the BLAST search on the "**M V M T H**" amyoid sequence are shown below and on the next page.

immunoglobulin heavy chain junction region [Homo sapiens]
Sequence ID: MOP80811.1Length: 16Number of Matches: 1

See 2 more title(s) See all Identical Proteins(IPG)

Related Information
Identical Proteins-Identical proteins to MOP80811.1
Range 1: 7 to 11 GenPept Graphics

Score	Expect	Identities	Positives	Gaps
22.3 bits(45)	7457	5/5(100%)	5/5(100%)	0/5(0%)

```
Query  1   MVMTH  5
           MVMTH
Sbjct  7   MVMTH  11
```

Immunoglobulin heavy chain junction region

The **immunoglobulin heavy chain** (**IgH**) is the large *polypeptide subunit* of an **antibody** (*immunoglobulin*). In human genome, the *IgH gene* loci are on *chromosome 14*. A typical **antibody** is composed of *two immunoglobulin (Ig) heavy chains* and *two Ig light chains*. Several different types of heavy chain exist that define the class or isotype of an antibody. These heavy chain types vary between different animals. All ***heavy chains*** contain a series of *immunoglobulin domains*, usually with one *variable domain* (V_H) that is important for binding antigen and several constant domains (C_H1, C_H2, ecc.). **Production of a viable heavy chain is a key step in B cell maturation**. (https://en.wikipedia.org/wiki/Immunoglobulin_heavy_chain#:~:text=Each%20heavy%20chain%20has%20two,hinge%20region%20for%20added%20flexibility).

Below there is the *sequence of the bases* (and *amino acids*) corresponding to the **junction region of the hemoglobin heavy chain**:

ATG GTC ATG ACC CAC

M V M T H

Although the amino acids remain the same (remember that different codons can encode the same amino acid), the *two letters* in bold yellow in the sequence shown above are **different** from the corresponding ones in the segment "**ATG GTA ATG ACA CAC**" of **Sequence n° 1/3**.

```
CDS             <306..>353
                /function="JUNCTION"
                /inference="COORDINATES: protein motif:IgBLAST:1.7.0"
                /codon_start=1
                /product="immunoglobulin heavy chain junction region"
                /protein_id="MOP80811.1"
                /translation="CARHAAMVMTHPPDYW"
```

SDR family NAD(P)-dependent oxidoreductase [Chryseobacterium sp. c4a]
Sequence ID: WP_160138415.1 Length: 4641 Number of Matches: 1
Range 1: 2526 to 2530 GenPept Graphics

Score	Expect	Identities	Positives	Gaps
22.3 bits(45)	8209	5/5(100%)	5/5(100%)	0/5(0%)

```
Query  1     MVMTH  5
             MVMTH
Sbjct  2526  MVMTH  2530
```

SDR family NAD (P)-dependent oxidoreductase was taken into consideration because it was already identified by the BLAST search done previously on the **M A M T H** sequence. Unfortunately, of this significant alignment I could not find the file in which it is possible to check the bases.

To do a cross-check, I performed a BLAST search on the "**ATGGTAATGACACAC**" base sequence. The results are shown below and on the next page.

HIV-1 isolate KRC414204 from South Korea envelope glycoprotein (env) gene, partial cds
Sequence ID: KC690757.1 Length: 1155 Number of Matches: 1
Range 1: 836 to 850 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
30.2 bits(15)	606	15/15(100%)	0/15(0%)	Plus/Plus

```
CDS: Putative 1      1       G  N  D  T
Query                1    ATGGTAATGACACAC  15
                          ||||||||||||||
Sbjct                836  ATGGTAATGACACAC  850
CDS:envelope glycopr 279  N  G  N  D  T
```

HIV-1 isolate 17 from Zambia envelope glycoprotein (env) gene, complete cds
Sequence ID: GU329450.1 Length: 3074 Number of Matches: 1
Range 1: 1439 to 1453 GenBank Graphics

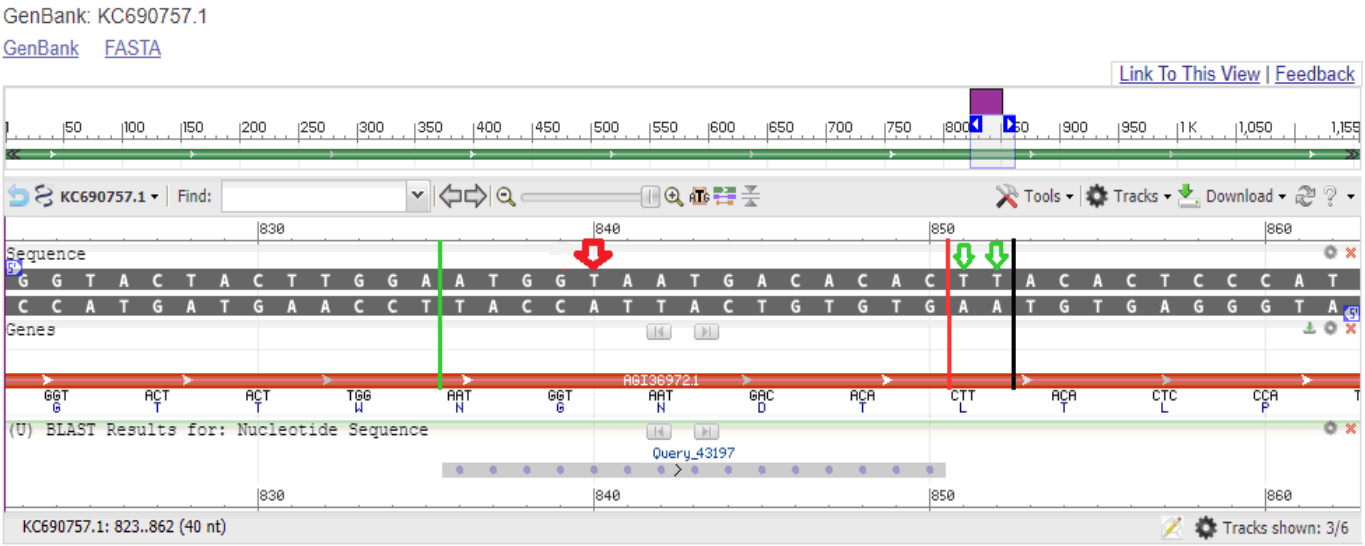
Score	Expect	Identities	Gaps	Strand
30.2 bits(15)	606	15/15(100%)	0/15(0%)	Plus/Plus

```
CDS: Putative 1       1        G  N  D  T
Query                 1    ATGGTAATGACACAC  15
                           ||||||||||||||||
Sbjct                 1439 ATGGTAATGACACAC  1453
CDS:envelope glycopr  392  N  G  N  D  T
```

As can be seen from the two results of the BLAST research reported above, the base sequence "**ATGGTAATGACACAC**" refers to two sequences of **HIV-1**, but the amino acids are different from those expected because the *coding correspondence of the codons is different*. Despite the ***non-correspondence of the codons***, a structure similar to **Query M V M T H** can be observed, that is **three** amino acids of which ***two*** *lateral equal* and *one different central* (**N G N**).
Being clearly not satisfied with these results and driven by the need to clarify the role of this ***repeated configuration*** (**three** amino acids, ***two*** *of which are lateral* and ***one*** *central different*, a configuration that will be found in other sequences shown in the following chapters) in the short sequence "**ATGGTAATGACACAC**", I further deepened the study in search of information that could give me clearer answers and allow me to continue or definitively abandon this investigation.

Looking at the graph presented below of the sequence "**HIV-1 isolate KRC414204 from South Korea envelope glycoprotein (env) gene, partial cds"** (see the previous page) I noticed that the sequence continued with the "**TT**" bases (see the green arrows in the graph).

HIV-1 isolate KRC414204 from South Korea envelope glycoprotein (env) gene, partial cds



I did a new BLAST search by adding two trailing "**T**" **bases** to the "**ATGGTAATGACACAC**" sequence to make the sequence longer and have more opportunities for potentially interesting significant alignments. In addition to the predictable **HIV-1 isolate KRC414204 from South Korea envelope glycoprotein (env) gene** and **HIV-1 isolate 17 from Zambia envelope glycoprotein (env) gene**, the research reveals ***other interesting results***: **three** significant alignments of the "**ATGGTAATGACACACTT**" sequence refer to **three sequences** (**one** of **chromosome 13** and **two** of **chromosome 7**) of **two *Bats***.

HIV-1 isolate KRC414204 from South Korea envelope glycoprotein (env) gene, partial cds
Sequence ID: KC690757.1 Length: 1155 Number of Matches: 1
Range 1: 836 to 852 GenBank Graphics Next Match Previous Match

Score	Expect	Identities	Gaps	Strand
34.2 bits(17)	39	17/17(100%)	0/17(0%)	Plus/Plus

```
CDS: Putative 1      1        G  N  D  T  L
Query                1    ATGGTAATGACACACTT  17
                          ||||||||||||||||
Sbjct                836  ATGGTAATGACACACTT  852
CDS:envelope glycopr 279  N  G  N  D  T  L
```

HIV-1 isolate 17 from Zambia envelope glycoprotein (env) gene, complete cds
Sequence ID: GU329450.1 Length: 3074 Number of Matches: 1
Range 1: 1439 to 1454 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
32.2 bits(16)	153	16/16(100%)	0/16(0%)	Plus/Plus

```
CDS: Putative 1      1        G  N  D  T  L
Query                1    ATGGTAATGACACACT  16
                          ||||||||||||||||
Sbjct                1439 ATGGTAATGACACACT  1454
CDS:envelope glycopr 392  N  G  N  D  T  L
```

Pipistrellus pipistrellus genome assembly, chromosome: 13
Sequence ID: LR862369.1 Length: 53281546 Number of Matches: 1
Range 1: 21690317 to 21690332 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
32.2 bits(16)	153	16/16(100%)	0/16(0%)	Plus/Minus

```
Query  1         ATGGTAATGACACACT  16
                 ||||||||||||||||
Sbjct  21690332  ATGGTAATGACACACT  21690317
```

Pipistrellus pipistrellus genome assembly, chromosome: 7
Sequence ID: LR862363.1 Length: 84548104 Number of Matches: 2
Range 1: 14470523 to 14470538 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
32.2 bits(16)	153	16/16(100%)	0/16(0%)	Plus/Minus

```
Query  2         TGGTAATGACACACTT  17
                 ||||||||||||||||
Sbjct  14470538  TGGTAATGACACACTT  14470523
```

Range 2: 69199040 to 69199054 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
30.2 bits(15)	606	15/15(100%)	0/15(0%)	Plus/Plus

```
Query  4         GTAATGACACACTTT  18
                 |||||||||||||||
Sbjct  69199040  GTAATGACACACTTT  69199054
```

"**Audentes fortuna iuvat**" ("*Destiny favors those who dare*") is the famous ***Latin phrase*** commonly translated into English "**Fortune favors the bold**". Never was a motto so appropriate!
Repeating the BLAST search on the sequence "**ATG GTA ATG ACA CAC TT**" to confirm the *significant alignments identified*, I make a very trivial error in transcribing the letters of the sequence and write: "**ATG GTA ATG ACA CAT T**", omitting a "**C**" before the two "**T**".
It is therefore not a real (involuntary) modification of the sequence capable of affecting its "**Non-Obvious Trends**" (if not marginally in its final part), but a **reduction of its length** before the addition of the two bases of **Thymine** (**T**). An actual modification of the sequence had already occurred with the *replacement* of the letter "**C**" with the letter "**T**" in the **GCA** codon, to the sequence "**G GCA ATG ACA CAC AAG A**" (the sequence identified as corresponding to the ***HIV-1 Virus*** and referred to the "**misc_feature**" ***region***).

In addition, the *omission* of the "**C**" may not significantly affect the formation of the *last codon* because the "**CAC**" and "**CAT**" codons both encode the amino acid **Histidine** (**H**). Of course, the latter circumstance will **only** occur if the coding starts from one of the two potential **ATG** codons present in the sequence or from other previous ones that respect this order of codon formation.
The results of the BLAST research presented in the next paragraph represent some of the most important revelations of this very long and laborious document.

9.7 “AUDENTES FORTUNA IUVAT”: THE SEQUENCE “ATGGTAATGACACATT”

QUERY: “ATGGTAATGACACATT”.

Search Parameters	
Program	blastn
Word size	7
Expect value	1000
Hitlist size	**5000**
Match/Mismatch scores	1,-3
Gapcosts	5,2
Filter string	F
Genetic Code	

5000 sequences selected:

Accession Seq	Description	Comomm Name	Max Score	Total Score	Query C.	E value	Per. Id.	Acc. Len
.....	.....	.....	.....	.....	.....	.....	.....	.....
CP056483.1	**Klebsiella sp. RHBSTW-00464 chromosome, complete genome**	**Klebsiella sp. RHBSTW-00464**	32.2	32.2	100%	153	100.00%	6331260
CP055481.1	**Klebsiella sp. RHBSTW-00484 chromosome, complete genome**	Klebsiella sp. RHBSTW-00484	32.2	32.2	100%	153	100.00%	6330682
.....	.....	.....	.....	.....	.....	.....	.....	.....
XM_033096151.1	**PREDICTED: Rhinolophus ferrumequinum F-box and WD repeat domain containing 11 (FBXW11), transcript variant X9, mRNA**	**Rhinolophus ferrumequinum**	30.2	30.2	93%	606	100.00%	4093
XM_033096150.1	**PREDICTED: Rhinolophus ferrumequinum F-box and WD repeat domain containing 11 (FBXW11), transcript variant X8, mRNA**	Rhinolophus ferrumequinum	30.2	30.2	93%	606	100.00%	4334
XM_033096148.1	**PREDICTED: Rhinolophus ferrumequinum F-box and WD repeat domain containing 11 (FBXW11), transcript variant X7, mRNA**	Rhinolophus ferrumequinum	30.2	30.2	93%	606	100.00%	4295
XM_033096147.1	**PREDICTED: Rhinolophus ferrumequinum F-box and WD repeat domain containing 11 (FBXW11), transcript variant X6, mRNA**	Rhinolophus ferrumequinum	30.2	30.2	93%	606	100.00%	4369
XM_033096146.1	**PREDICTED: Rhinolophus ferrumequinum F-box and WD repeat domain containing 11 (FBXW11), transcript variant X5, mRNA**	Rhinolophus ferrumequinum	30.2	30.2	93%	606	100.00%	4397
XM_033096145.1	**PREDICTED: Rhinolophus**	Rhinolophus	30.2	30.2	93%	606	100.00%	4197

	ferrumequinum F-box and WD repeat domain containing 11 (FBXW11), transcript variant X4, mRNA	ferrumequinum						
XM_033096144.1	**PREDICTED: Rhinolophus ferrumequinum F-box and WD repeat domain containing 11 (FBXW11), transcript variant X3, mRNA**	Rhinolophus ferrumequinum	30.2	30.2	93%	606	100.00%	1981
XM_033096143.1	**PREDICTED: Rhinolophus ferrumequinum F-box and WD repeat domain containing 11 (FBXW11), transcript variant X2, mRNA**	Rhinolophus ferrumequinum	30.2	30.2	93%	606	100.00%	2189
XM_033096142.1	**PREDICTED: Rhinolophus ferrumequinum F-box and WD repeat domain containing 11 (FBXW11), transcript variant X1, mRNA**	Rhinolophus ferrumequinum	30.2	30.2	93%	606	100.00%	4260
XM_033130381.1	**PREDICTED: Rhinolophus ferrumequinum beta-transducin repeat containing E3 ubiquitin protein ligase (BTRC), transcript variant X13, mRNA**	Rhinolophus ferrumequinum	30.2	30.2	93%	606	100.00%	6365
XM_033130380.1	**PREDICTED: Rhinolophus ferrumequinum beta-transducin repeat containing E3 ubiquitin protein ligase (BTRC), transcript variant X12, mRNA**	Rhinolophus ferrumequinum	30.2	30.2	93%	606	100.00%	6352
XM_033130379.1	**PREDICTED: Rhinolophus ferrumequinum beta-transducin repeat containing E3 ubiquitin protein ligase (BTRC), transcript variant X11, mRNA**	Rhinolophus ferrumequinum	30.2	30.2	93%	606	100.00%	6378
XM_033130378.1	**PREDICTED: Rhinolophus ferrumequinum beta-transducin repeat containing E3 ubiquitin protein ligase (BTRC), transcript variant X10, mRNA**	Rhinolophus ferrumequinum	30.2	30.2	93%	606	100.00%	6526
XM_033130377.1	**PREDICTED: Rhinolophus ferrumequinum beta-transducin repeat containing E3 ubiquitin protein ligase (BTRC), transcript variant X9, mRNA**	Rhinolophus ferrumequinum	30.2	30.2	93%	606	100.00%	6539
XM_033130375.1	**PREDICTED: Rhinolophus ferrumequinum beta-transducin repeat containing E3 ubiquitin protein ligase (BTRC), transcript variant X8, mRNA**	Rhinolophus ferrumequinum	30.2	30.2	93%	606	100.00%	6430
XM_033130374.1	**PREDICTED: Rhinolophus ferrumequinum beta-transducin repeat containing E3 ubiquitin protein ligase (BTRC), transcript variant X7, mRNA**	Rhinolophus ferrumequinum	30.2	30.2	93%	606	100.00%	6443
XM_033130373.1	**PREDICTED: Rhinolophus ferrumequinum beta-transducin repeat containing**	Rhinolophus ferrumequinum	30.2	30.2	93%	606	100.00%	6368

XM_033130372.1	**E3 ubiquitin protein ligase (BTRC), transcript variant X6, mRNA PREDICTED: Rhinolophus ferrumequinum beta-transducin repeat containing E3 ubiquitin protein ligase (BTRC), transcript variant X5, mRNA**	Rhinolophus ferrumequinum	30.2	30.2	93%	606	100.00%	6446
XM_033130371.1	**PREDICTED: Rhinolophus ferrumequinum beta-transducin repeat containing E3 ubiquitin protein ligase (BTRC), transcript variant X4, mRNA**	Rhinolophus ferrumequinum	30.2	30.2	93%	606	100.00%	6476
XM_033130370.1	**PREDICTED: Rhinolophus ferrumequinum beta-transducin repeat containing E3 ubiquitin protein ligase (BTRC), transcript variant X3, mRNA**	Rhinolophus ferrumequinum	30.2	30.2	93%	606	100.00%	6386
XM_033130369.1	**PREDICTED: Rhinolophus ferrumequinum beta-transducin repeat containing E3 ubiquitin protein ligase (BTRC), transcript variant X2, mRNA**	Rhinolophus ferrumequinum	30.2	30.2	93%	606	100.00%	6436
XM_033130368.1	**PREDICTED: Rhinolophus ferrumequinum beta-transducin repeat containing E3 ubiquitin protein ligase (BTRC), transcript variant X1, mRNA**	Rhinolophus ferrumequinum	30.2	30.2	93%	606	100.00%	2322
.....	.....	.....	.....	.....	.....	.....	.....	.....
CP048618.1	**Legionella pneumophila strain ERS1305867 chromosome, complete genome**	**Legionella pneumophila**	30.2	30.2	93%	606	100.00%	3405663
.....	.....	.....	.....	.....	.....	.....	.....	.....
LR133933.1	**Legionella pneumophila strain NCTC12180 genome assembly, chromosome: 1**	Legionella pneumophila	30.2	30.2	93%	606	100.00%	3369585
.....	.....	.....	.....	.....	.....	.....	.....	.....
LR134380.1	**Legionella pneumophila subsp. pascullei strain NCTC12273 genome assembly, chromosome: 1**	**Legionella pneumophila subsp. pascullei**	30.2	30.2	93%	606	100.00%	3458917
.....	.....	.....	.....	.....	.....	.....	.....	.....
LS483412.1	**Legionella pneumophila subsp. pascullei strain NCTC12272 genome assembly, chromosome: 1**	Legionella pneumophila subsp. pascullei	30.2	30.2	93%	606	100.00%	3458914
.....	.....	.....	.....	.....	.....	.....	.....	.....
CP021286.1	**Legionella pneumophila subsp. pneumophila strain Albuquerque 1 (D-7474) chromosome, complete genome**	**Legionella pneumophila subsp. pneumophila**	30.2	30.2	93%	606	100.00%	3508676
CP021281.1	**Legionella pneumophila subsp. pneumophila strain Flint 2 (D-7477) chromosome, complete genome**	Legionella pneumophila subsp. pneumophila	30.2	30.2	93%	606	100.00%	3588028
CP021279.1	**Legionella pneumophila subsp. fraseri strain F-4198 chromosome, complete genome**	**Legionella pneumophila subsp. fraseri**	30.2	30.2	93%	606	100.00%	3461540

CP021274.1	**Legionella pneumophila strain D-4040 chromosome, complete genome**	Legionella pneumophila	30.2	30.2	93%		606	100.00%	3503573
CP021272.1	**Legionella pneumophila strain D-5265 chromosome, complete genome**	Legionella pneumophila	30.2	30.2	93%		606	100.00%	3498323
CP021270.1	**Legionella pneumophila strain D-7787 chromosome, complete genome**	Legionella pneumophila	30.2	30.2	93%		606	100.00%	3363033
CP021268.1	**Legionella pneumophila subsp. pneumophila strain Birmingham 1 (D-7470) chromosome, complete genome**	Legionella pneumophila subsp. pneumophila	30.2	30.2	93%		606	100.00%	3436658
CP021263.1	**Legionella pneumophila subsp. fraseri strain D-3137 chromosome, complete genome**	Legionella pneumophila subsp. fraseri	30.2	30.2	93%		606	100.00%	3474810
CP021262.1	**Legionella pneumophila subsp. pascullei strain U8W (D-7160) chromosome, complete genome**	Legionella pneumophila subsp. pascullei	30.2	30.2	93%		606	100.00%	3458918
CP021261.1	**Legionella pneumophila strain NY23 (D-7705) chromosome, complete genome**	Legionella pneumophila	30.2	30.2	93%		606	100.00%	3546234
CP021260.1	**Legionella pneumophila strain NY24 (D-7706) chromosome, complete genome**	Legionella pneumophila	30.2	30.2	93%		606	100.00%	3581477
CP021259.1	**Legionella pneumophila strain D-7708 chromosome, complete genome**	Legionella pneumophila	30.2	30.2	93%		606	100.00%	3412234
CP021258.1	**Legionella pneumophila subsp. fraseri strain D-5744 chromosome, complete genome**	Legionella pneumophila subsp. fraseri	30.2	60.5	100%		606	100.00%	3487066
CP021257.1	**Legionella pneumophila subsp. fraseri strain Lansing 3 chromosome, complete genome**	Legionella pneumophila subsp. fraseri	30.2	60.5	100%		606	100.00%	3611000
CP021256.1	**Legionella pneumophila strain D-4954 chromosome, complete genome**	Legionella pneumophila	30.2	30.2	93%		606	100.00%	3661362
.....	.....	.....	.....	.....	.....	.....		.....	.....
LT906452.1	**Legionella pneumophila strain NCTC11985 genome assembly, chromosome: 1**	Legionella pneumophila	30.2	30.2	93%		606	100.00%	3514598
.....	.....	.....	.....	.....	.....	.....		.....	.....
CP017458.1	**Legionella pneumophila subsp. fraseri strain Dallas 1E, complete genome**	Legionella pneumophila subsp. fraseri	30.2	60.5	100%		606	100.00%	3500943
CP017457.1	**Legionella pneumophila subsp. fraseri strain Detroit-1, complete genome**	Legionella pneumophila subsp. fraseri	30.2	30.2	93%		606	100.00%	3515377
LT632615.1	**Legionella pneumophila strain ST23 genome assembly, chromosome: 1**	Legionella pneumophila	30.2	30.2	93%		606	100.00%	3509586
LT632614.1	**Legionella pneumophila strain ST62 genome assembly, chromosome: 1**	Legionella pneumophila	30.2	30.2	93%		606	100.00%	3530817
.....	.....	.....	.....	.....	.....	.....		.....	.....
CP016030.2	**Legionella pneumophila strain OLDA, complete genome**	Legionella pneumophila	30.2	30.2	93%		606	100.00%	3486107
CP016029.2	**Legionella pneumophila strain Pontiac, complete genome**	Legionella pneumophila	30.2	30.2	93%		606	100.00%	3545003

.....	.....	.....	.....	.....	.....	.....		.....	.....
CP011105.1	**Legionella pneumophila strain L10-023, complete genome**	Legionella pneumophila	30.2	30.2	93%		606	100.00%	3641882
CP012019.1	**Legionella pneumophila subsp. pneumophila strain Toronto-2005, complete genome**	Legionella pneumophila subsp. pneumophila	30.2	30.2	93%		606	100.00%	3573898
CP014256.1	**Legionella pneumophila subsp. pascullei strain D-7158, complete genome**	Legionella pneumophila subsp. pascullei	30.2	30.2	93%		606	100.00%	3458918
.....	.....	.....	.....	.....	.....	.....		.....	.....
KR350687.1	**Legionella pneumophila subsp. pneumophila ATCC 33215 O-antigen gene cluster, complete sequence**	**Legionella pneumophila subsp. pneumophila ATCC 33215**	30.2	30.2	93%		606	100.00%	19262
KR350684.1	**Legionella pneumophila strain NCTC12181 O-antigen gene cluster, complete sequence**	Legionella pneumophila	30.2	30.2	93%		606	100.00%	18701
.....	.....	.....	.....	.....	.....	.....		.....	.....
CP003730.1	**Legionella pneumophila subsp. pneumophila str. Thunder Bay, complete genome**	**Legionella pneumophila subsp. pneumophila str. Thunder Bay**	30.2	30.2	93%		606	100.00%	3455167
.....	.....	.....	.....	.....	.....	.....		.....	.....
XM_017678656.2	**PREDICTED: Manis javanica beta-transducin repeat containing E3 ubiquitin protein ligase (BTRC), transcript variant X7, mRNA**	**Manis javanica**	30.2	30.2	93%		606	100.00%	6347
XM_017678647.2	**PREDICTED: Manis javanica beta-transducin repeat containing E3 ubiquitin protein ligase (BTRC), transcript variant X6, mRNA**	Manis javanica	30.2	30.2	93%		606	100.00%	6425
XM_017678640.2	**PREDICTED: Manis javanica beta-transducin repeat containing E3 ubiquitin protein ligase (BTRC), transcript variant X5, mRNA**	Manis javanica	30.2	30.2	93%		606	100.00%	6455
XM_017678629.2	**PREDICTED: Manis javanica beta-transducin repeat containing E3 ubiquitin protein ligase (BTRC), transcript variant X4, mRNA**	Manis javanica	30.2	30.2	93%		606	100.00%	6408
XM_017678621.2	**PREDICTED: Manis javanica beta-transducin repeat containing E3 ubiquitin protein ligase (BTRC), transcript variant X3, mRNA**	Manis javanica	30.2	30.2	93%		606	100.00%	6682
XR_001856019.2	**PREDICTED: Manis javanica beta-transducin repeat containing E3 ubiquitin protein ligase (BTRC), transcript variant X2, misc_RNA**	Manis javanica	30.2	30.2	93%		606	100.00%	5913

XM_017678606.2	**PREDICTED: Manis javanica beta-transducin repeat containing E3 ubiquitin protein ligase (BTRC), transcript variant X1, mRNA**	Manis javanica	30.2	30.2	93%	606	100.00%	6533
…….	…….	…….	…….	…..	……	……	……	…….
XM_036898782.1	**PREDICTED: Manis pentadactyla beta-transducin repeat containing E3 ubiquitin protein ligase (BTRC), transcript variant X7, mRNA**	**Manis pentadactyla**	30.2	30.2	93%	606	100.00%	4933
XM_036898780.1	**PREDICTED: Manis pentadactyla beta-transducin repeat containing E3 ubiquitin protein ligase (BTRC), transcript variant X6, mRNA**	Manis pentadactyla	30.2	30.2	93%	606	100.00%	5011
XM_036898779.1	**PREDICTED: Manis pentadactyla beta-transducin repeat containing E3 ubiquitin protein ligase (BTRC), transcript variant X5, mRNA**	Manis pentadactyla	30.2	30.2	93%	606	100.00%	5037
XM_036898778.1	**PREDICTED: Manis pentadactyla beta-transducin repeat containing E3 ubiquitin protein ligase (BTRC), transcript variant X4, mRNA**	Manis pentadactyla	30.2	30.2	93%	606	100.00%	5179
XM_036898777.1	**PREDICTED: Manis pentadactyla beta-transducin repeat containing E3 ubiquitin protein ligase (BTRC), transcript variant X3, mRNA**	Manis pentadactyla	30.2	30.2	93%	606	100.00%	5119
XR_005027378.1	**PREDICTED: Manis pentadactyla beta-transducin repeat containing E3 ubiquitin protein ligase (BTRC), transcript variant X2, misc_RNA**	Manis pentadactyla	30.2	30.2	93%	606	100.00%	4641
XM_036898776.1	**PREDICTED: Manis pentadactyla beta-transducin repeat containing E3 ubiquitin protein ligase (BTRC), transcript variant X1, mRNA**	Manis pentadactyla	30.2	30.2	93%	606	100.00%	5257
…….	…….	…….	…….	…..	……	……	……	…….
CP003192.1	**Legionella pneumophila subsp. pneumophila ATCC 43290, complete genome**	**Legionella pneumophila subsp. pneumophila ATCC 43290**	30.2	30.2	93%	606	100.00%	3359001

.....	.....	.....	.....	.....	.....	.....	.....	.....
XM_036437885.1	**PREDICTED: Pipistrellus kuhlii beta-transducin repeat containing E3 ubiquitin protein ligase (BTRC), transcript variant X13, mRNA**	**Pipistrellus kuhlii**	30.2	30.2	93%	606	100.00%	6318
XM_036437884.1	**PREDICTED: Pipistrellus kuhlii beta-transducin repeat containing E3 ubiquitin protein ligase (BTRC), transcript variant X12, mRNA**	Pipistrellus kuhlii	30.2	30.2	93%	606	100.00%	6330
XR_004987975.1	**PREDICTED: Pipistrellus kuhlii beta-transducin repeat containing E3 ubiquitin protein ligase (BTRC), transcript variant X11, misc_RNA**	Pipistrellus kuhlii	30.2	30.2	93%	606	100.00%	1886
XM_036437883.1	**PREDICTED: Pipistrellus kuhlii beta-transducin repeat containing E3 ubiquitin protein ligase (BTRC), transcript variant X10, mRNA**	Pipistrellus kuhlii	30.2	30.2	93%	606	100.00%	6257
XM_036437882.1	**PREDICTED: Pipistrellus kuhlii beta-transducin repeat containing E3 ubiquitin protein ligase (BTRC), transcript variant X9, mRNA**	Pipistrellus kuhlii	30.2	30.2	93%	606	100.00%	6408
XM_036437881.1	**PREDICTED: Pipistrellus kuhlii beta-transducin repeat containing E3 ubiquitin protein ligase (BTRC), transcript variant X8, mRNA**	Pipistrellus kuhlii	30.2	30.2	93%	606	100.00%	6359
XM_036437879.1	**PREDICTED: Pipistrellus kuhlii beta-transducin repeat containing E3 ubiquitin protein ligase (BTRC), transcript variant X7, mRNA**	Pipistrellus kuhlii	30.2	30.2	93%	606	100.00%	6438
XM_036437878.1	**PREDICTED: Pipistrellus kuhlii beta-transducin repeat containing E3 ubiquitin protein ligase (BTRC), transcript variant X6, mRNA**	Pipistrellus kuhlii	30.2	30.2	93%	606	100.00%	6444
XM_036437877.1	**PREDICTED: Pipistrellus kuhlii beta-transducin repeat containing E3 ubiquitin protein ligase (BTRC), transcript variant X5, mRNA**	Pipistrellus kuhlii	30.2	30.2	93%	606	100.00%	6462
XM_036437876.1	**PREDICTED: Pipistrellus kuhlii beta-transducin repeat containing E3 ubiquitin protein ligase (BTRC), transcript variant X4, mRNA**	Pipistrellus kuhlii	30.2	30.2	93%	606	100.00%	6422
XM_036437875.1	**PREDICTED: Pipistrellus kuhlii beta-transducin repeat containing E3 ubiquitin protein ligase (BTRC), transcript variant X3, mRNA**	Pipistrellus kuhlii	30.2	30.2	93%	606	100.00%	6474
XM_036437874.1	**PREDICTED: Pipistrellus**	Pipistrellus kuhlii	30.2	30.2	93%	606	100.00%	6516

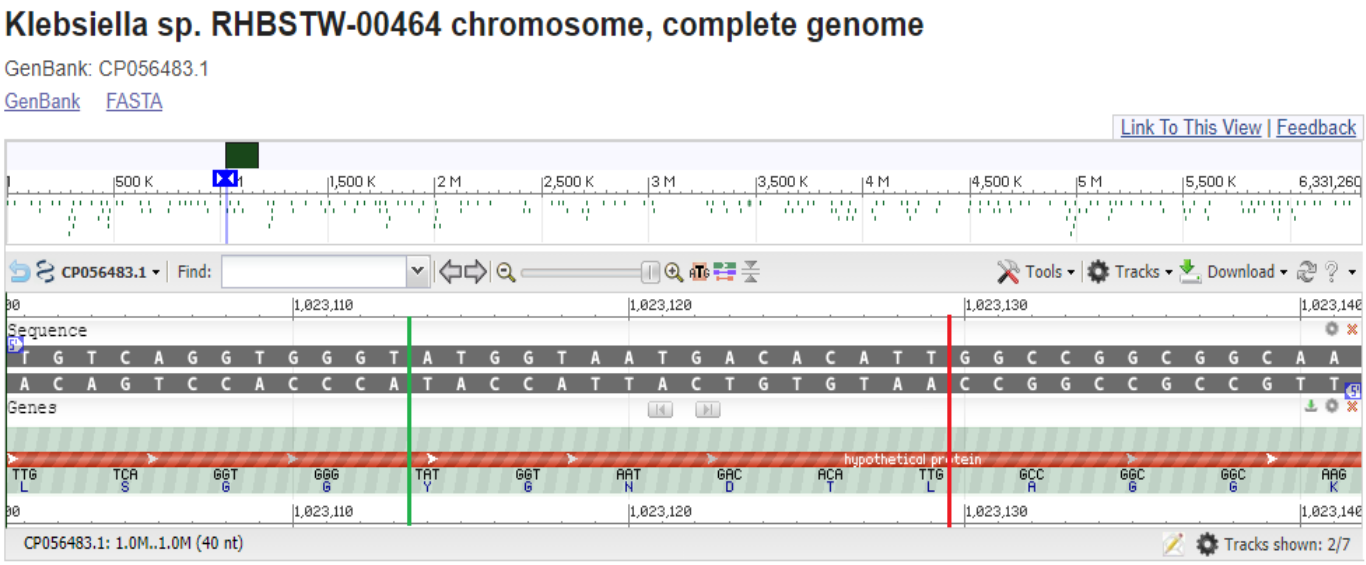
XM_036437873.1	**kuhlii beta-transducin repeat containing E3 ubiquitin protein ligase (BTRC), transcript variant X2, mRNA PREDICTED: Pipistrellus kuhlii beta-transducin repeat containing E3 ubiquitin protein ligase (BTRC), transcript variant X1, mRNA**	Pipistrellus kuhlii	30.2	30.2	93%		606	100.00%	6552
.....	.....	.....	.....	.....	.....	.....		.....	.....
CP061840.1	**Legionella pneumophila subsp. pneumophila strain AW-13-4 chromosome, complete genome**	Legionella pneumophila subsp. pneumophila	30.2	30.2	93%		606	100.00%	3512088
.....	.....	.....	.....	.....	.....	.....		.....	.....
CP059400.1	**Legionella sp. PC1000 chromosome, complete genome**	**Legionella sp. PC1000**	30.2	30.2	93%		606	100.00%	4081644
.....	.....	.....	.....	.....	.....	.....		.....	.....
LR862370.1	**Pipistrellus pipistrellus genome assembly, chromosome: 14**	**Pipistrellus pipistrellus**	30.2	60.5	100%		606	100.00%	51672279
LR862369.1	**Pipistrellus pipistrellus genome assembly, chromosome: 13**	Pipistrellus pipistrellus	30.2	30.2	93%		606	100.00%	53281546
LR862367.1	**Pipistrellus pipistrellus genome assembly, chromosome: 11**	Pipistrellus pipistrellus	30.2	120	93%		606	100.00%	69386144
LR862365.1	**Pipistrellus pipistrellus genome assembly, chromosome: 9**	Pipistrellus pipistrellus	30.2	30.2	93%		606	100.00%	72542072
LR862363.1	**Pipistrellus pipistrellus genome assembly, chromosome: 7**	Pipistrellus pipistrellus	30.2	30.2	93%		606	100.00%	84548104
.....	.....	.....	.....	.....	.....	.....		.....	.....
FR749891.1	**Legionella pneumophila LPS gene cluster, strain ATCC 33215 serogroup 6**	Legionella pneumophila subsp. pneumophila ATCC 33215	30.2	30.2	93%		606	100.00%	28119
FR749890.1	**Legionella pneumophila LPS gene cluster, strain ATCC 43283 serogroup 10**	**Legionella pneumophila subsp. pneumophila ATCC 43283**	30.2	30.2	93%		606	100.00%	30773
FR747827.1	**Legionella pneumophila LPS gene cluster, strain ATCC 43736 serogroup 13**	Legionella pneumophila	30.2	30.2	93%		606	100.00%	28726
FR747826.1	**Legionella pneumophila LPS gene cluster, strain ATCC 43290 serogroup 12**	Legionella pneumophila subsp. pneumophila ATCC 43290	30.2	30.2	93%		606	100.00%	28322
FR733637.1	**Legionella pneumophila LPS gene cluster strain ATCC43073 serogroup 14**	Legionella pneumophila	30.2	30.2	93%		606	100.00%	30771
.....	.....	.....	.....	.....	.....	.....		.....	.....
CP001828.1	**Legionella pneumophila 2300/99 Alcoy, complete genome**	**Legionella pneumophila 2300/99 Alcoy**	30.2	30.2	93%		606	100.00%	3516334
.....	.....	.....	.....	.....	.....	.....		.....	.....
CP000675.2	**Legionella pneumophila str. Corby, complete genome**	**Legionella pneumophila str. Corby**	30.2	30.2	93%		606	100.00%	3576470
.....	.....	.....	.....	.....	.....	.....		.....	.....
CR628336.1	**Legionella pneumophila str. Paris complete genome**	**Legionella pneumophila str. Paris**	30.2	30.2	93%		606	100.00%	3503610
.....	.....	.....	.....	.....	.....	.....		.....	.....

AF131870.1	***Simian immunodeficiency virus complete genome***	**Simian immunodeficiency virus**	30.2	30.2	93%	606	100.00%	10006

Klebsiella sp. RHBSTW-00464 chromosome, complete genome
Sequence ID: CP056483.1 Length: 6331260 Number of Matches: 1
Range 1: 1023114 to 1023129 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
32.2 bits(16)	153	16/16(100%)	0/16(0%)	Plus/Plus

```
CDS: Putative 1   1            G  N  D  T
Query             1        ATGGTAATGACACATT  16
                           ||||||||||||||||
Sbjct             1023114  ATGGTAATGACACATT  1023129
CDS: hypothetica  93       Y  G  N  D  T  L
```

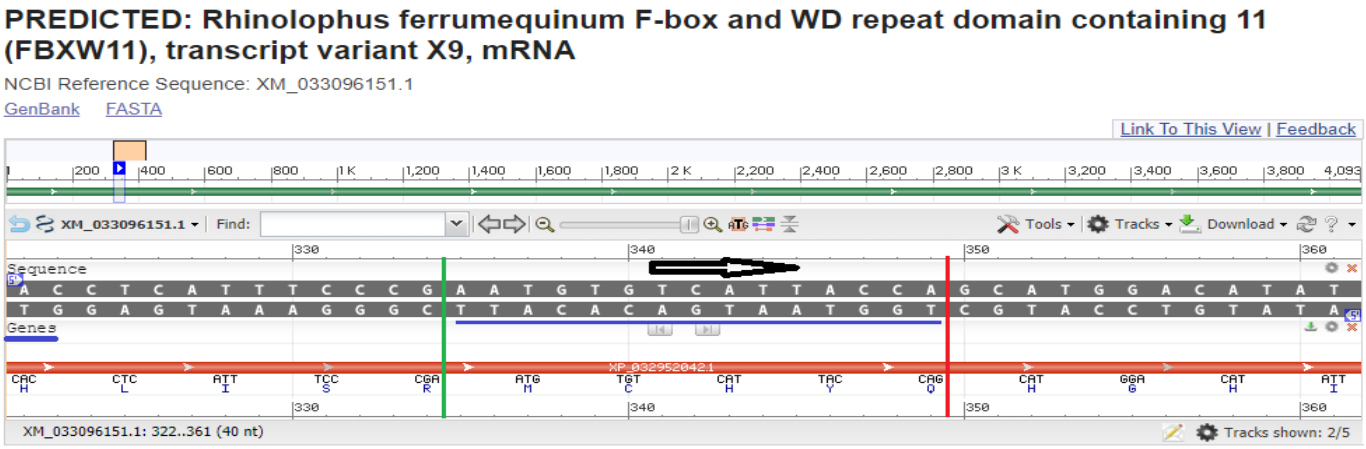


The amino acid sequence Y G N D T L of **Klebsiella sp.** is almost totally similar (differs only for the first amino acid) to the N G N D T L sequence related to **HIV-1 isolate KRC414204 from South Korea envelope glycoprotein (env) gene** and **HIV-1 isolate 17 from Zambia envelope glycoprotein (env) gene** (see pages 194, 195, and 196).

PREDICTED: Rhinolophus ferrumequinum F-box and WD repeat domain containing 11 (FBXW11), transcript variant X9, mRNA
Sequence ID: XM_033096151.1 Length: 4093 Number of Matches: 1
Range 1: 335 to 349 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
30.2 bits(15)	606	15/15(100%)	0/15(0%)	Plus/Minus

```
CDS: Putative 1       4       Y  H  C  M
Query                 2    TGGTAATGACACATT  16
                           ||||||||||||||||
Sbjct                 349  TGGTAATGACACATT  335
CDS:F-box/WD repeat-  69   Q  Y  H  C  M
```

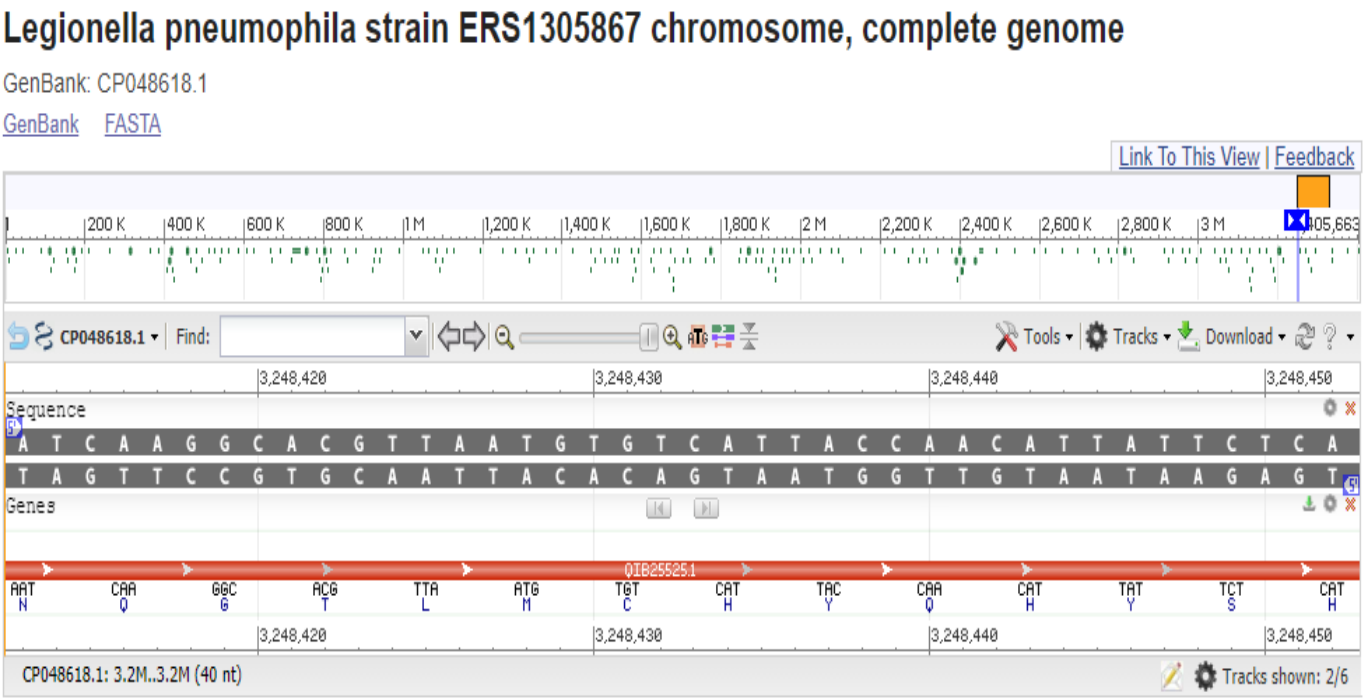


βTrCP2 (beta-transducin repeat containing protein 2; also known as **Fbxw11** or HOS) is a protein that in humans is encoded by the ***FBXW11*** (***F-box and WD repeat domain containing 11***) gene (https://en.wikipedia.org/wiki/FBXW11#cite_note-Koch,_A._et_al.-38). This gene encodes a member of the ***F-box protein*** family which is characterized by an approximately 40 residue structural motif, the **F-box**.
Members of the ***WDR11*** family are involved in a variety of cellular processes, including cell cycle progression, signal transduction, apoptosis, and gene regulation (https://www.genecards.org/cgi-bin/carddisp.pl?gene=WDR11#:~:text=WD%20Repeat%20Domain%2011&text=This%20gene%20encodes%20a%20member,of%20heterotrimeric%20or%20multiprotein%20complexes). The **F-box** proteins constitute one of the four subunits of ***ubiquitin protein ligase complex*** called ***SCFs*** (Skp1-Cul1-F-box protein), which often, but not always, recognize substrates in a *phosphorylation-dependent* manner. ***Human βTrCP*** (referred to both βTrCP1 and βTrCP2) was originally identified as a **cellular ubiquitin ligase** that is bound by the ***HIV-1 Vpu*** *viral protein* [accessory protein that in ***HIV*** is encoded by the Vpu gene. This gene is found exclusively in ***HIV-1*** and some ***HIV-1-related Simian Immunodeficiency Virus*** (SIV)] **to eliminate cellular CD4** (Cluster of differentiation 4, glycoprotein found on the surface of immune cells such as T helper cells, monocytes, macrophages, and dendritic cells) by connecting it to the proteolytic machinery (https://www.sciencedirect.com/science/article/abs/pii/S1286457908001755#:~:text=Vpu%20is%20an%20accessory%20viral,of%20virion%20release%20from%20cells).
βTrCP behaves as an **oncoprotein** in some tissues. Elevated levels of βTrCP expression have been found in **colorectal**, **pancreatic**, **hapatoblastoma**, and **breast cancers** (https://en.wikipedia.org/wiki/FBXW11#cite_note-Koch,_A._et_al.-38).

Legionella pneumophila strain ERS1305867 chromosome, complete genome
Sequence ID: CP048618.1 Length: 3405663 Number of Matches: 1
Range 1: 3248426 to 3248440 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
30.2 bits(15)	606	15/15(100%)	0/15(0%)	Plus/Minus

```
CDS: Putative 1        4          Y  H  C  M
Query                  2        TGGTAATGACACATT  16
                                ||||||||||||||
Sbjct                  3248440  TGGTAATGACACATT  3248426
CDS:SAM-dependent me   261      Q  Y  H  C  M
```



S-adenosyl-L-methionine-dependent methyltransferase (*SAM-dependent methyltransferase*)

This entry represents the ***S-adenosyl-L-methionine-dependent methyltransferase*** (***SAM MTase***) superfamily. *Methyltransferases* transfer a **methyl group from a donor to an acceptor**. *SAM-binding methyltransferases* utilise the *ubiquitous methyl donor SAM* as a cofactor to methylate proteins, small molecules, lipids, and nucleic acids. All *SAM MTases* contain a structurally conserved SAM-binding domain consisting of a central seven-stranded beta-sheet that is flanked by three alpha-helices per side of the sheet (https://www.ebi.ac.uk/interpro/entry/InterPro/IPR029063/).

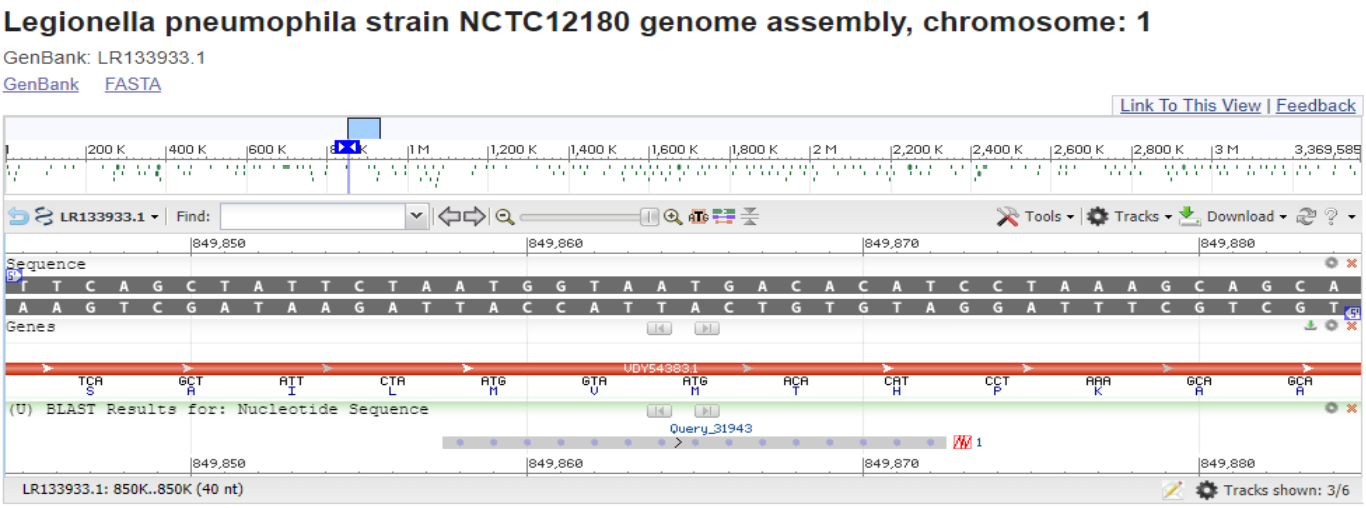
Legionella pneumophila strain NCTC12180 genome assembly, chromosome: 1
Sequence ID: LR133933.1 Length: 3369585 Number of Matches: 1
Range 1: 849858 to 849872 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
30.2 bits(15)	606	15/15(100%)	0/15(0%)	Plus/Plus

```
CDS: Putative 1        1          M  V  M  T  H
Query                  1          ATGGTAATGACACAT  15
                                  |||||||||||||||
Sbjct                  849858     ATGGTAATGACACAT  849872
CDS:putative NAD dep   215        M  V  M  T  H
```

ATTENTION:

It is interesting to observe that the codon ACA, encoding ***Threonine*** (**T**) in the **Legionella pneumophila** sequence, in the ***RNA*** sequence "**GT GAA ATG GTC ATG TGT GGC GG**" [taken from the **RdRp** (*pSARS-F*) **Gene** and object of our study] has (in the same position) *mirror bases* (**TGT**) encoding **Cysteine** (**C**).



NAD-dependent epimerase/dehydratase

This domain is found in **proteins that utilise NAD** as a cofactor and use *nucleotide-sugar substrates* for a variety of chemical reactions. One of the best studied of these *proteins* is ***UDP-galactose 4-epimerase*** which catalyses the *conversion of UDP-galactose* to *UDP-glucose* during *galactose metabolism* (https://www.ebi.ac.uk/interpro/entry/InterPro/IPR001509/).

Legionella pneumophila subsp. pascullei strain NCTC12273 genome assembly, chromosome: 1
Sequence ID: LR134380.1Length: 3458917Number of Matches: 1
Range 1: 864956 to 864970 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
30.2 bits(15)	606	15/15(100%)	0/15(0%)	Plus/Plus

```
CDS: Putative 1       1          M  V  M  T  H
Query                 1          ATGGTAATGACACAT  15
                                 ||||||||||||||
Sbjct                 864956     ATGGTAATGACACAT  864970
CDS:putative NAD dep  215        M  V  M  T  H
```

Legionella pneumophila subsp. pneumophila strain Albuquerque 1 (D-7474) chromosome, complete genome
Sequence ID: CP021286.1Length: 3508676Number of Matches: 1
Range 1: 3341670 to 3341684 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
30.2 bits(15)	606	15/15(100%)	0/15(0%)	Plus/Minus

```
Query  2         TGGTAATGACACATT  16
                 ||||||||||||||
Sbjct  3341684   TGGTAATGACACATT  3341670
```

Legionella pneumophila subsp. pneumophila strain Flint 2 (D-7477) chromosome, complete genome
Sequence ID: CP021281.1 Length: 3588028 Number of Matches: 1
Range 1: 3421015 to 3421029 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
30.2 bits(15)	606	15/15(100%)	0/15(0%)	Plus/Minus

```
Query  2         TGGTAATGACACATT  16
                 ||||||||||||||
Sbjct  3421029   TGGTAATGACACATT  3421015
```

Legionella pneumophila subsp. fraseri strain F-4198 chromosome, complete genome
Sequence ID: CP021279.1 Length: 3461540 Number of Matches: 1
Range 1: 3299576 to 3299590 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
30.2 bits(15)	606	15/15(100%)	0/15(0%)	Plus/Minus

```
Query  2         TGGTAATGACACATT  16
                 ||||||||||||||
Sbjct  3299590   TGGTAATGACACATT  3299576
```

Legionella pneumophila strain D-4040 chromosome, complete genome
Sequence ID: CP021274.1 Length: 3503573 Number of Matches: 1
Range 1: 3340382 to 3340396 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
30.2 bits(15)	606	15/15(100%)	0/15(0%)	Plus/Minus

```
Query  2        TGGTAATGACACATT  16
               ||||||||||||||
Sbjct  3340396  TGGTAATGACACATT  3340382
```

Legionella pneumophila subsp. pneumophila strain Birmingham 1 (D-7470) chromosome, complete genome
Sequence ID: CP021268.1 Length: 3436658 Number of Matches: 1
Range 1: 3291340 to 3291354 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
30.2 bits(15)	606	15/15(100%)	0/15(0%)	Plus/Minus

```
Query  2        TGGTAATGACACATT  16
               ||||||||||||||
Sbjct  3291354  TGGTAATGACACATT  3291340
```

Legionella pneumophila strain NCTC11985 genome assembly, chromosome: 1
Sequence ID: LT906452.1 Length: 3514598 Number of Matches: 1
Range 1: 889867 to 889881 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
30.2 bits(15)	606	15/15(100%)	0/15(0%)	Plus/Plus

```
CDS: Putative 1       1          M  V  M  T  H
Query                 1       ATGGTAATGACACAT  15
                              ||||||||||||||
Sbjct                 889867   ATGGTAATGACACAT  889881
CDS:putative NAD dep  215        M  V  M  T  H
```

Legionella pneumophila subsp. fraseri strain Dallas 1E, complete genome
Sequence ID: CP017458.1 Length: 3500943 Number of Matches: 2
Range 1: 938986 to 939000 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
30.2 bits(15)	606	15/15(100%)	0/15(0%)	Plus/Plus

```
CDS: Putative 1       1          M  V  M  T  H
Query                 1       ATGGTAATGACACAT  15
                              ||||||||||||||
Sbjct                 938986   ATGGTAATGACACAT  939000
CDS:NAD-dependent de  215        M  V  M  T  H
```

Range 2: 3338979 to 3338993 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
30.2 bits(15)	606	15/15(100%)	0/15(0%)	Plus/Minus

```
CDS: Putative 1       4           Y  H  C  M
Query                 2        TGGTAATGACACATT  16
                               ||||||||||||||
Sbjct                 3338993  TGGTAATGACACATT  3338979
CDS:SAM-dependent me  261      Q  Y  H  C  M
```

It is interesting to note how **Legionella pneumophila subsp. fraseri strain** uses the ***same sequence*** to ***encode different amino acids*** that refer respectively to **catalytic activity** (*NAD-dependent dehydratase*) and to a ***methyltransferase*** (*SAM-dependent methyltransferase*).

PREDICTED: Manis javanica beta-transducin repeat containing E3 ubiquitin protein ligase (BTRC), transcript variant X7, mRNA
Sequence ID: XM_017678656.2 Length: 6347 Number of Matches: 1
Range 1: 377 to 391 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
30.2 bits(15)	606	15/15(100%)	0/15(0%)	Plus/Minus

```
CDS: Putative 1       4        Y  H  C  M
Query                 2     TGGTAATGACACATT  16
                            ||||||||||||||
Sbjct                 391   TGGTAATGACACATT  377
CDS:F-box/WD repeat-  99    Q  Y  H  C  M
```

PREDICTED: Manis javanica beta-transducin repeat containing E3 ubiquitin protein ligase (BTRC), transcript variant X2, misc_RNA
Sequence ID: XR_001856019.2 Length: 5913 Number of Matches: 1
Range 1: 563 to 577 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
30.2 bits(15)	606	15/15(100%)	0/15(0%)	Plus/Minus

```
Query  2    TGGTAATGACACATT  16
            ||||||||||||||
Sbjct  577  TGGTAATGACACATT  563
```

PREDICTED: Manis pentadactyla beta-transducin repeat containing E3 ubiquitin protein ligase (BTRC), transcript variant X7, mRNA
Sequence ID: XM_036898782.1 Length: 4933 Number of Matches: 1
Range 1: 386 to 400 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
30.2 bits(15)	606	15/15(100%)	0/15(0%)	Plus/Minus

```
CDS: Putative 1       4       Y   H   C   M
Query                 2    TGGTAATGACACATT  16
                           ||||||||||||||||
Sbjct                 400  TGGTAATGACACATT  386
CDS:F-box/WD repeat-  102  Q   Y   H   C   M
```

PREDICTED: Pipistrellus kuhlii beta-transducin repeat containing E3 ubiquitin protein ligase (BTRC), transcript variant X13, mRNA
Sequence ID: XM_036437885.1 Length: 6318 Number of Matches: 1
Range 1: 387 to 401 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
30.2 bits(15)	606	15/15(100%)	0/15(0%)	Plus/Minus

```
CDS: Putative 1       4       Y   H   C   M
Query                 2    TGGTAATGACACATT  16
                           ||||||||||||||||
Sbjct                 401  TGGTAATGACACATT  387
CDS:F-box/WD repeat-  98   Q   Y   H   C   M
```

Pipistrellus pipistrellus genome assembly, chromosome: 14
Sequence ID: LR862370.1 Length: 51672279 Number of Matches: 2
Range 1: 33468446 to 33468460 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
30.2 bits(15)	606	15/15(100%)	0/15(0%)	Plus/Plus

```
Query  2         TGGTAATGACACATT  16
                 ||||||||||||||||
Sbjct  33468446  TGGTAATGACACATT  33468460
```

Range 2: 50266727 to 50266741 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
30.2 bits(15)	606	15/15(100%)	0/15(0%)	Plus/Minus

```
Query  1         ATGGTAATGACACAT  15
                 ||||||||||||||||
Sbjct  50266741  ATGGTAATGACACAT  50266727
```

Pipistrellus pipistrellus genome assembly, chromosome: 13
Sequence ID: LR862369.1 Length: 53281546 Number of Matches: 1
Range 1: 38306039 to 38306053 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
30.2 bits(15)	606	15/15(100%)	0/15(0%)	Plus/Minus

```
Query  2         TGGTAATGACACATT  16
                 ||||||||||||||
Sbjct  38306053  TGGTAATGACACATT  38306039
```

Pipistrellus pipistrellus genome assembly, chromosome: 11
Sequence ID: LR862367.1 Length: 69386144 Number of Matches: 4
Range 1: 18422230 to 18422244 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
30.2 bits(15)	606	15/15(100%)	0/15(0%)	Plus/Minus

```
Query  1         ATGGTAATGACACAT  15
                 ||||||||||||||
Sbjct  18422244  ATGGTAATGACACAT  18422230
```

Range 2: 62870006 to 62870020 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
30.2 bits(15)	606	15/15(100%)	0/15(0%)	Plus/Plus

```
Query  1         ATGGTAATGACACAT  15
                 ||||||||||||||
Sbjct  62870006  ATGGTAATGACACAT  62870020
```

Range 3: 62909060 to 62909074 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
30.2 bits(15)	606	15/15(100%)	0/15(0%)	Plus/Plus

```
Query  1         ATGGTAATGACACAT  15
                 ||||||||||||||
Sbjct  62909060  ATGGTAATGACACAT  62909074
```

Range 4: 62981149 to 62981163 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
30.2 bits(15)	606	15/15(100%)	0/15(0%)	Plus/Minus

```
Query  1         ATGGTAATGACACAT  15
                 ||||||||||||||
Sbjct  62981163  ATGGTAATGACACAT  62981149
```

Pipistrellus pipistrellus genome assembly, chromosome: 9
Sequence ID: LR862365.1 Length: 72542072 Number of Matches: 1
Range 1: 49224238 to 49224252 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
30.2 bits(15)	606	15/15(100%)	0/15(0%)	Plus/Minus

```
Query  2         TGGTAATGACACATT  16
                 ||||||||||||||
Sbjct  49224252  TGGTAATGACACATT  49224238
```

Pipistrellus pipistrellus genome assembly, chromosome: 7
Sequence ID: LR862363.1 Length: 84548104 Number of Matches: 1
Range 1: 51662018 to 51662032 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
30.2 bits(15)	606	15/15(100%)	0/15(0%)	Plus/Plus

```
Query  1         ATGGTAATGACACAT  15
                 ||||||||||||||
Sbjct  51662018  ATGGTAATGACACAT  51662032
```

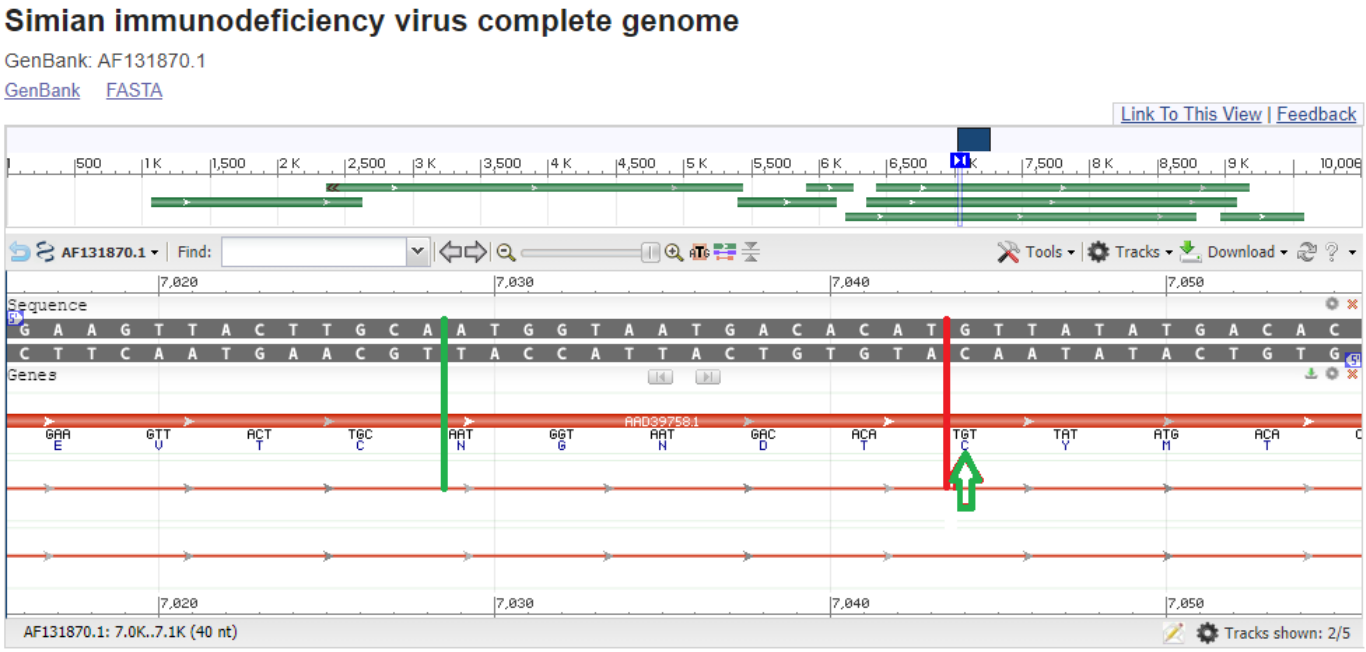
Simian immunodeficiency virus complete genome

Sequence ID: AF131870.1 Length: 10006 Number of Matches: 1
Range 1: 7029 to 7043 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
30.2 bits(15)	606	15/15(100%)	0/15(0%)	Plus/Plus

```
CDS: Putative 1        1          G  N  D  T
Query                  1      ATGGTAATGACACAT  15
                              ||||||||||||||
Sbjct                  7029   ATGGTAATGACACAT  7043
CDS:envelope glycopr   203    N  G  N  D  T
```

The **Simian immunodeficiency virus** N G N D T amino acid sequence is almost totally similar (differs only for the last amino acid, see the green arrow in the graph shown below) to the sequence N G N D T L, sequence related to **HIV-1 isolate KRC414204 from South Korea envelope glycoprotein (env) gene** and **HIV-1 isolate 17 from Zambia envelope glycoprotein (env) gene** (see pages 194, 195, and 196).



Of the **5000** sequences selected by the BLAST search, the sequences in the list just shown are just some of those considered most important and significant for this study.
Before continuing, I would like to remind you that the base sequence "**ATGGTAATGACA CATT**" (on which this last BLAST search was carried out) is obtained from the sequence "**GGCAATGACACAC**" [part of **Sequence n ° 1/3** (**CTCGGGCAATGACACACAAGAA**) identified as a correspondent to the ***HIV-1 Virus*** and referred to the **"misc_feature" region**], to which a letter has been replaced (the "**C**" with the "**T**", so as to obtain an amino acid of **V**aline instead of one of **A**lanine) and to which the final "**C**" has been omitted and then two "**T**" added. As we know, **Sequence n ° 1/3** is one of the *modified sequences* of the **Original Sequence**, the **22-base RNA sequence** "**GTGAAATGGTCATGTGTGGCGG**" taken from the **SARS-CoV-2 Gene RdRp** (see the summary table shown below).

Original Sequence	**G**	**T**	**G**	**A**	**A**	**A**	**T**	**G**	**G**	**T**	**C**	**A**	**T**	**G**	**T**	**G**	**T**	**G**	**G**	**C**	**G**	**G**
Sequence n° 1/3 (86,36%)	**C**	**T**	**C**	**G**	**G**	**G**	**C**	**A**	**A**	**T**	**G**	**A**	**C**	**A**	**C**	**A**	**C**	**A**	**A**	**G**	**A**	**A**
Sequence n° 1/3 (HIV-1)					**G**	**G**	**C**	**A**	**A**	**T**	**G**	**A**	**C**	**A**	**C**	**A**	**C**	**A**	**A**	**G**	**A**	
Sequ. n° 1/3 (HIV-1) - Modified			**A**	**T**	**G**	**G**	**T**	**A**	**A**	**T**	**G**	**A**	**C**	**A**	**C**	**A**	**T**	**T**				

The results of the BLAST research on **Sequence 1/3 (HIV-1) - Modified** (see the table shown on the previous page) are particularly interesting. The **16 bases** considered identified a significant number of "organisms" and, among these, the **two "organisms" object of this research**: **three species** of **bats** [**Rhinolophus** (*ferrumequinum*) and **Pipistrellus** (*pipistrellus* and *kuhlii*)] and **two species** of **pangolins** [**Manis javanica** (*malay pangolin*) and **Manis pentadactyla** (*chinese pangolin*)].
These very important *significant alignments* reveal the presence of significant links between the **sequences** of the **Gene RdRp** of **SARS-CoV-2**, of the **Bats**, of the **Pangolins** and, *unexpectedly and surprisingly*, also and above all of a "**misc_feature**" ***region*** of the ***HIV-1 Virus*** (**region** defined as "**nonfunctional envelope glycoprotein due to mutation**" of the **gene Env**).

Many other *significant alignments* should be reported, but I limit myself to the following ones because the "organisms" to which they refer are implicated in mainly **infectious, feverish** and **pulmonary** symptoms: **Klebsiella sp.**, **Legionella pneumophila** (*subsp. Pascullei, subsp. Pneumophila, subsp . phraseri, subsp. pneumophila ATCC 33215, subsp. pneumophila str. Thunder Bay, subsp. pneumophila ATCC 43290, sp. PC1000, subsp. pneumophila ATCC 43283, str. Corby, str. Paris* and *2300/99 Alcoy*) and **Simian immunodeficiency virus**.

Klebsiella is a ***Gram-negative bacterium*** that can ***cause destructive changes to human and animal lungs if aspirated***, specifically to the alveoli resulting in bloody, brownish or yellow colored jelly like sputum. ***Klebsiella*** illness affects *middle-aged* and *older men* with *debilitating diseases* (https://en.wikipedia.org/wiki/Klebsiella_pneumoniae). This patient population is believed to have ***impaired respiratory host defenses***, including persons with *diabetes, alcoholism, malignancy, liver disease, chronic obstructive pulmonary diseases, glucocorticoid therapy, kidney failure*, and *certain occupational exposures*. Many of these infections are obtained when a person is in the hospital for some other reason. The most common condition caused by *Klebsiella* bacteria outside the hospital is *pneumonia*, typically in the form of bronchopneumonia and also bronchitis. These patients have an increased tendency to develop lung abscess, cavitation, empyema, and pleural adhesions. It has a death rate around 50%, even with antimicrobial therapy. It has a death rate around 50%, even with antimicrobial therapy. Individuals with *Klebsiella* pneumonia tend to cough up a characteristic sputum, as well as having fever, nausea, tachycardia, and vomiting (https://en.wikipedia.org/wiki/Klebsiella_pneumoniae). Treatment for *Klebsiella* pneumonia is by antibiotics such as **aminoglycosides** and **cephalosporins**, the choice depending upon the person's health condition, medical history

and severity of the disease. In addition to *pneumonia*, ***Klebsiella*** can also cause infections in the urinary tract, lower biliary tract, and surgical wound sites. The range of clinical diseases includes pneumonia, thrombophlebitis, urinary tract infection, cholecystitis, diarrhea, upper respiratory tract infection, wound infection, osteomyelitis, meningitis, and bacteremia, and sepsis. Sepsis and septic shock can follow entry of the bacteria into the blood (https://en.wikipedia.org/wiki/Klebsiella_pneumoniae).

Legionella pneumophila is a thin, aerobic, pleomorphic, flagellated, non-spore-forming, **Gram-negative bacterium** of the genus Legionella and is the causative agent of Legionnaires' disease, also known as **legionellosis** (https://en.wikipedia.org/wiki/Legionella_pneumophila). To obtain *amino acids* (the amino acids are the primary carbon and energy source of *L. pneumophila*, that have almost 12 classes of *ABC-transporters*, *amino acid permeases*, and many *proteases*, to exploit it), *L. pneumophila* uses the **AnkB F-Box effector** (https://en.wikipedia.org/wiki/AnkB_F-Box), is a protein excreted by *Legionella pneumophila* which contains a *domain F-box*. F-box proteins are involved in the ***ubiquitination*** of **proteins targeted for degradation by the proteasome**. **AnkB F-box** is a *protein* that assembles host cell *polyubiquitinated proteins* on the cytoplasmatic side of *Legionella Containing Vacuole* (LCV) (https://en.wikipedia.org/wiki/AnkB_F-Box). This effector is required for **intracellular proliferation** within human and amoeba cells and for intrapulmonary Legionella proliferation in mice (https://en.wikipedia.org/wiki/AnkB_F-Box). The presence of the *F-box domain* is an example of **molecular mimicry** (https://en.wikipedia.org/wiki/Molecular_mimicry, topic of crucial importance in the study of ***autoimmunity*** and of which I recommend a specific study also for a greater understanding of *SARS-CoV-2* infection) of *eukaryotic F-box* by which *Legionella* exploits the proteasomal degradation capacity of the host cell for its own sake (https://en.wikipedia.org/wiki/Molecular_mimicry). However, **promotion of proteasomal degradation** for the obtention of amino acids may not be the only *virulence strategy* to obtain carbon and energy sources from the host. **Type II–secreted degradative enzymes** may provide an *additional strategy* to generate carbon and energy sources. **Macrolides** (*azithromycin* or *clarithromycin*) or **fluoroquinolones** (*levofloxacin* or *moxifloxacin*) are the standard treatment for *Legionella* pneumonia in humans, with **levofloxacin** being considered first line with increasing resistance to **azithromycin** (https://en.wikipedia.org/wiki/Legionella_pneumophila).

Simian immunodeficiency virus (**SIV**) is a species of retrovirus (type of *virus* that inserts a copy of its RNA genome into the DNA of a host cell that it invades, thus changing the genome of that cell) that cause persistent infections in at least 45 species of *African non-human primates*. Based on analysis of strains found in four species of monkeys from Bioko Island, which was isolated from the mainland by rising sea levels about 11,000 years ago, it has been concluded that SIV has been present in monkeys and apes for at least 32,000 years, and probably much longer (https://en.wikipedia.org/wiki/Simian_immunodeficiency_virus#:~:text=Simian%20immunodeficiency%20virus%20(SIV)%20is,of%20African%20non%2Dhuman%20primates).

Virus strains from two of these primate species are believed to have *crossed the species barrier into humans*, resulting in **HIV-2** and **HIV-1** respectively, the **two human immunodeficiency viruses**. The most likely route of **transmission of HIV-1 to humans** involves contact with the **blood of chimps** that are often hunted for bushmeat in Africa. Unlike HIV-1 and HIV-2 infections in humans, SIV infections in their *natural African simian non-human hosts* appear in many cases to be *non-pathogenic* due to evolutionary adaptation of the hosts to the virus. However, if the virus infects an *Asian or Indian rhesus macaque*, these non-African simian primates will also develop **simian AIDS** (**SAIDS**), as they, like humans, have not had a prolonged history with the virus. In 2012, researchers reported that initial infection of rhesus monkeys by neutralization-resistant **SIV** strains[28] could be partially prevented through use of an **anti- SIV** $_{\text{SME543}}$ **vaccine obligately including Env protein antigens** (https://en.wikipedia.org/wiki/Simian_immunodeficiency_virus#:~:text=Simian%20immunodeficiency%20virus%20(SIV)%20is,of%20African%20non%2Dhuman%20primates).

It is necessary to point out that the

Sequ. n° 1/3 (HIV-1) - Modified **A T G G T A A T G A C A C A T T**

which has been used here to identify **Simian immunodeficiency virus** (and all the other "organisms" that have already been discussed) is a *minimal modification* of a sequence of a "**misc_feature**" ***region*** defined as "**nonfunctional envelope glycoprotein due to mutation**" of the **Env gene** (a *viral gene* that codes for the protein that forms the **viral envelope**) **of the genome of HIV-1 infected Homo sapiens and not of a monkey genome**. In turn, the sequence of the **Env gene** referred to above was identified starting from a **modified sequence** (following the "*Non-obvious Trend*" n°1) of the ***SARS-CoV-2 RNA***. These relationships will necessarily have to be the subject of future studies and insights.

9.8 FROM THE SEQUENCE OF AMINO ACIDS M V M T H TO THE SEQUENCE OF AMINO ACIDS M V M T H F

This paragraph shows some results of the BLAST search carried out on the sequence the amino acids **M V M T H F**, sequence obtained by adding an amino acid of ***Phenylalanine*** (**F**) to the already known **M V M T H** (sequence which, as seen in the previous paragraphs, derives from the modification of the sequence **M A M T H**). The choice to extend the sequence **M V M T H** with another amino acid is justified by the search for potential amino acids that could bring out *significant relationships* between the sequence **M V M T H** and the **bat**, **pagolino** or **other "organisms"** of interest for the current study.

QUERY: M V M T H F

Search Parameters	
Program	blastp
Word size	2
Expect value	200000
Hitlist size	5000
Gapcosts	9,1
Matrix	PAM30
Filter string	F
Genetic Code	1
Window Size	40
Threshold	11
Composition-based stats	0

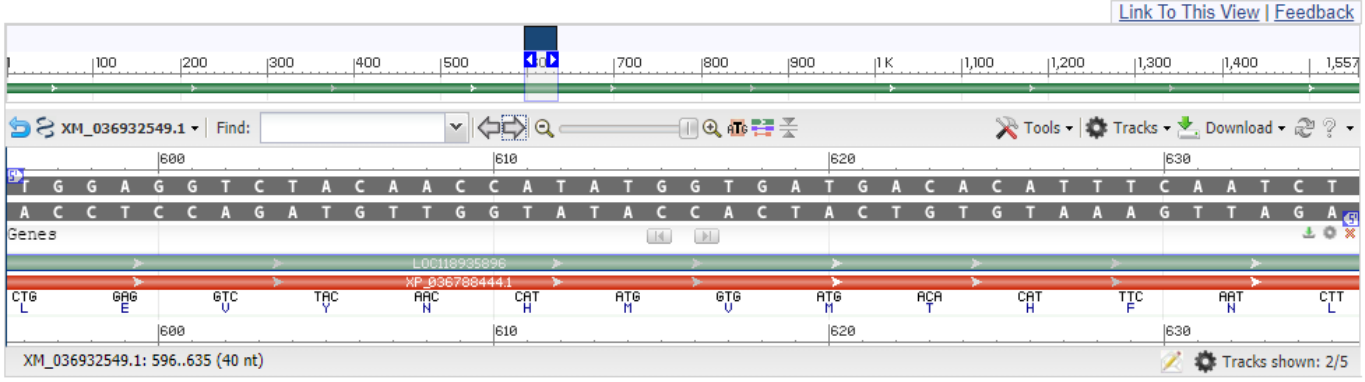
LOW QUALITY PROTEIN: olfactory receptor 6Y1-like [Manis pentadactyla]
Sequence ID: XP_036788444.1 Length: 518 Number of Matches: 1
Range 1: 205 to 210 GenPept Graphics

Score	Expect	Identities	Positives	Gaps
26.1 bits(54)	463	6/6(100%)	6/6(100%)	0/6(0%)

```
Query  1    MVMTHF  6
            MVMTHF
Sbjct  205  MVMTHF  210
```

PREDICTED: Manis pentadactyla olfactory receptor 6Y1-like (LOC118935896), mRNA

NCBI Reference Sequence: XM_036932549.1
GenBank FASTA



Olfactory receptor 6Y1-like

Olfactory receptors (**ORs**) are expressed in the *cell membranes* of *olfactory receptor neurons* and are responsible for the *detection of odorants*. Activated *olfactory receptors* trigger nerve impulses which transmit information about odor to the brain. These receptors are members of the ***class A rhodopsin-like family of G protein-coupled receptors (GPCRs)***. The **olfactory receptors** form a multigene family consisting of around 800 genes in humans and 1400 genes in mice (https://en.wikipedia.org/wiki/Olfactory_receptor). **Olfactory receptor subfamily 6B and related proteins are member of the class A family of seven-transmembrane G protein-coupled receptors (7tmA_OR6B-like)**. This group includes human olfactory receptor 6B, 6A, **6Y**, 6P, and related proteins in other mammals, sauropsids, and amphibians. More than 60% of human ORs are *non-functional pseudogenes* compared to only 20% in mouse. Each OR can recognize structurally similar odorants, and a single odorant can be detected by several ORs. Binding of an odorant to the olfactory receptor induces a *conformational change* that leads to the **activation of the olfactory-specific G protein** (*Golf*). The *G protein* (Golf and/or Gs) in turn stimulates *adenylate cyclase* to make cAMP (https://www.ncbi.nlm.nih.gov/Structure/cdd/cddsrv.cgi?uid=320352). The cAMP opens cyclic nucleotide-gated ion channels, which allow the influx of calcium and sodium ions, resulting in depolarization of the olfactory receptor neuron and triggering an action potential which transmits this information to the brain. A consensus nomenclature system based on evolutionary divergence is used here to classify the olfactory receptor family. The nomenclature begins with the root name OR, followed by an integer representing a family, a letter denoting a subfamily, and an integer representing the individual gene within the subfamily (https://www.ncbi.nlm.nih.gov/Structure/cdd/cddsrv.cgi?uid=320352).

LOW QUALITY PROTEIN: olfactory receptor 6Y1-like [Manis javanica]
Sequence ID: XP_036882530.1 Length: 336 Number of Matches: 1
Range 1: 28 to 33 GenPept Graphics

Score	Expect	Identities	Positives	Gaps
26.1 bits(54)	467	6/6(100%)	6/6(100%)	0/6(0%)

```
Query  1   MVMTHF  6
           MVMTHF
Sbjct  28  MVMTHF  33
```

PREDICTED: Manis javanica olfactory receptor 6Y1-like (LOC118973961), mRNA
gi|1926243006|ref|XM_037026635.1|

olfactory receptor 6Y1-like [Manis pentadactyla]
Sequence ID: XP_036788446.1 Length: 331 Number of Matches: 1
Range 1: 32 to 37 GenPept Graphics

Score	Expect	Identities	Positives	Gaps
26.1 bits(54)	467	6/6(100%)	6/6(100%)	0/6(0%)

```
Query  1   MVMTHF  6
           MVMTHF
Sbjct  32  MVMTHF  37
```

LOW QUALITY PROTEIN: F-BAR domain only protein 2 [Manis javanica]
Sequence ID: XP_036882044.1 Length: 1171 Number of Matches: 1
Range 1: 362 to 367 GenPept Graphics

Score	Expect	Identities	Positives	Gaps
22.7 bits(46)	7551	5/6(83%)	5/6(83%)	0/6(0%)

```
Query  1   MVMTHF  6
           MVM HF
Sbjct  362 MVMAHF  367
```

F-BAR domain

The F-BAR domain is emerging as an important player in ***membrane remodeling*** pathways. F-BAR domain proteins couple membrane remodeling with **actin** dynamics associated with *endocytic pathways* and *filopodium* formation (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2889966/). F-BAR domain containing proteins can be categorized into *five subfamilies* based on their phylogeny which is consistent with the additional protein domains they possess, for example, *RhoGAP domains, Cdc42 binding sites, SH3 domains and tyrosine kinase domains*. *Dynamin1/2*, *N-WASP*, *Huntingtin*, *intersectin* and *Cdc42* are central nodes influencing F-BAR domain protein function (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2889966/).

F-BAR domain only protein 2 isoform X1 [Pipistrellus kuhlii]
NCBI Reference Sequence: XP_036265115.1

Identical Proteins FASTA Graphics

Go to:

```
LOCUS       XP_036265115              6 aa            linear   MAM 01-OCT-2020
DEFINITION  F-BAR domain only protein 2 isoform X1 [Pipistrellus kuhlii].
ACCESSION   XP_036265115 REGION: 1..6
VERSION     XP_036265115.1
DBLINK      BioProject: PRJNA666119
DBSOURCE    REFSEQ: accession XM_036409222.1
KEYWORDS    RefSeq.
```

SOURCE Pipistrellus kuhlii (Kuhl's pipistrelle)

ORGANISM Pipistrellus kuhlii

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Laurasiatheria; Chiroptera; Microchiroptera; Vespertilionidae; Pipistrellus.

COMMENT MODEL REFSEQ: This record is predicted by automated computational analysis. This record is derived from a genomic sequence (NW_023425540.1) annotated using gene prediction method: Gnomon, supported by mRNA evidence.

Also see:

Documentation of NCBI's Annotation Process

##Genome-Annotation-Data-START##

Annotation Provider :: NCBI

Annotation Status :: Full annotation

Annotation Name :: Pipistrellus kuhlii Annotation Release 100

Annotation Version :: 100

Annotation Pipeline :: NCBI eukaryotic genome annotation pipeline

Annotation Software Version :: 8.5

Annotation Method :: Best-placed RefSeq; Gnomon

Features Annotated :: Gene; mRNA; CDS; ncRNA

##Genome-Annotation-Data-END##

COMPLETENESS: full length.

FEATURES Location/Qualifiers

source 1..6

/organism="Pipistrellus kuhlii"

/isolate="mPipKuh1"

/db_xref="taxon:59472"

/chromosome="Unknown"

/sex="male"

/tissue_type="flight muscle, heart, brain, liver, kidney"

/dev stage="adult"

/country="Italy: Bergamo"

/lat lon="45.7430 N 9.5831 E"

/collection date="2017"

/collected_by="Emma Teeling, Andrea Locatelli"

Protein 1..>18

/product="F-BAR domain only protein 2 isoform X1"

/calculated_mol_wt=90865

CDS 1..830

/gene="FCHO2"

/coded by="XM 036409222.1:122..2614"

/db_xref="GeneID:118702814"

```
ORIGIN
        1 mvmahf
//
```

F-BAR domain only protein 2 isoform X1 [Pipistrellus kuhlii]
Sequence ID: XP_036265115.1 Length: 830 Number of Matches: 1
Range 1: 1 to 6 GenPept Graphics

Score	Expect	Identities	Positives	Gaps
22.7 bits(46)	7566	5/6(83%)	5/6(83%)	0/6(0%)

```
Query  1  MVMTHF  6
          MVM HF
Sbjct  1  MVMAHF  6
```

F-BAR domain only protein 2 isoform X2 [Pipistrellus kuhlii]
Sequence ID: XP_036265116.1 Length: 815 Number of Matches: 1
Range 1: 1 to 6 GenPept Graphics

Score	Expect	Identities	Positives	Gaps
22.7 bits(46)	7567	5/6(83%)	5/6(83%)	0/6(0%)

```
Query  1  MVMTHF  6
          MVM HF
Sbjct  1  MVMAHF  6
```

F-BAR domain only protein 2 [Rhinolophus ferrumequinum]
Sequence ID: XP_032966415.1 Length: 811 Number of Matches: 1
Range 1: 1 to 6 GenPept Graphics

Score	Expect	Identities	Positives	Gaps
22.7 bits(46)	7615	5/6(83%)	5/6(83%)	0/6(0%)

```
Query  1  MVMTHF  6
          MVM HF
Sbjct  1  MVMAHF  6
```

F-BAR domain only protein 2 isoform X3 [Pipistrellus kuhlii]
Sequence ID: XP_036265117.1 Length: 810 Number of Matches: 1
Range 1: 1 to 6 GenPept Graphics

Score	Expect	Identities	Positives	Gaps
22.7 bits(46)	7567	5/6(83%)	5/6(83%)	0/6(0%)

```
Query  1  MVMTHF  6
          MVM HF
Sbjct  1  MVMAHF  6
```

F-BAR domain only protein 2 [Manis pentadactyla]
Sequence ID: XP_036741956.1 Length: 806 Number of Matches: 1
Range 1: 1 to 6 GenPept Graphics

Score	Expect	Identities	Positives	Gaps
22.7 bits(46)	7567	5/6(83%)	5/6(83%)	0/6(0%)

```
Query  1  MVMTHF  6
          MVM HF
Sbjct  1  MVMAHF  6
```

F-BAR domain only protein 2 isoform X4 [Pipistrellus kuhlii]
Sequence ID: XP_036265119.1 Length: 648 Number of Matches: 1
Range 1: 1 to 6 GenPept Graphics

Score	Expect	Identities	Positives	Gaps
22.7 bits(46)	7580	5/6(83%)	5/6(83%)	0/6(0%)

```
Query  1  MVMTHF  6
          MVM HF
Sbjct  1  MVMAHF  6
```

FCH and mu domain containing endocytic adaptor 2 [Pipistrellus kuhlii]
Sequence ID: KAF6366193.1 Length: 341 Number of Matches: 1
Range 1: 1 to 6 GenPept Graphics

Score	Expect	Identities	Positives	Gaps
22.7 bits(46)	7639	5/6(83%)	5/6(83%)	0/6(0%)

```
Query  1  MVMTHF  6
          MVM HF
Sbjct  1  MVMAHF  6
```

The results of the BLAST research on the sequence **M V M T H F** showed significant alignments with **pagolins** (*Manis pentadactyla and javanica*) and with **bats** (*Rhinolophus ferrumequinum and Pipistrellus kuhlii*).
This result are so relevant for **two reasons**.

The **first reason** is that the sequence of amino acids **M V M T H F**, as we know, has direct links with the sequence of amino acids **M A M T H** from which it derives [by replacing the amino acid of ***Alanine*** (**A**) with one of ***Valine*** (**V**) and by adding at the end the amino acid ***Phenylalanine*** (**F**)] and with the corresponding base sequence "**ATG GCA ATG ACA CAC**". This base sequence is part of a segment (**GGCAATGACACACAAGA**) of **Sequence n ° 1/3** (**CTCGGGCAATGACACACAAGAA**), that is the sequence identified as corresponding to the ***HIV-1 Virus*** [***region*** "**misc_feature**" **- nonfunctional envelope glycoprotein due to mutation** of the **Env gene**] and obtained from the modification (following the "*Non-Obvious Trend*" n° 1) of the **RNA** sequence **"GTGAA ATG GTC ATG TGT GGC GG"** (corresponding, in one of its segments, to the sequence of amino acids **M V M C G**) taken from the **RdRp Gene** of **SARS- CoV-2**.

The **second reason** why this result are so important is determined by the **type** of some sequences found and reported in this paragraph. In this case I am referring to **olfactory receptor 6Y1-like** [**Manis pentadactyla**] and [**Manis javanica**], a type of *olfactory receptor* also present in *humans*. The clinical reports of many people who have fallen ill with **Covid-19** show a temporary involvement of the loss of smell. Of course, I have no evidence of the possible involvement of this olfactory receptor in the disorder complained by patients suffering from **Covid-19**, but I considered it useful to report this "coincidence" to enable other interested researchers to investigate further.

9.9 SEQUENCE OF AMINO ACIDS "M A M T H K"

This paragraph shows some results of the BLAST search carried out on the sequence of amino acids **M A M T H K**, the sequence already considered in ***paragraph 9.5*** and to which the amino acid **Lysine** (**K**) determined by the next codon, **AAG** ***codon***.

QUERY: M A M T H K (atg gca atg aca cac aag)

Search Parameters	
Program	blastp
Word size	2
Expect value	200000
Hitlist size	**1000**
Gapcosts	9,1
Matrix	PAM30
Filter string	F
Genetic Code	1
Window Size	40
Threshold	11
Composition-based stats	0

1000 sequences selected:

Accession Seq	Description	Comomm Name	Max Score	Total Score	Query C.	E value	Per. Id.	Acc. Len
SBT73617.1	RNA pseudouridylate synthase, putative [Plasmodium ovale]	Plasmodium ovale	24.8	24.8	100%	1309	100.00%	4584
.....	.....	.....	.....	.....	.....	.....	.....	.....
MAJ70411.1	S9 family peptidase [Alteromonadaceae bacterium]	Alteromonadaceae bacterium	24.8	24.8	100%	1318	100.00%	804
.....	.....	.....	.....	.....	.....	.....	.....	.....
EFS5897567.1	L-serine ammonia-lyase [Campylobacter coli]	Campylobacter coli	24.8	24.8	100%	1327	100.00%	454
.....	.....	.....	.....	.....	.....	.....	.....	.....
VXA76768.1	adenosine deaminase [Aeromonas salmonicida]	Aeromonas salmonicida	24.8	24.8	100%	1333	100.00%	346
.....	.....	.....	.....	.....	.....	.....	.....	.....
WP_101048850.1	alpha/beta hydrolase [Macromonas sp. BK-30]	Macromonas sp. BK-30	24.8	24.8	100%	1337	100.00%	291
.....	.....	.....	.....	.....	.....	.....	.....	.....
WP_098544546.1	DNA helicase PcrA [Bacillus cereus]	Bacillus cereus	22.3	22.3	100%	10726	83.33%	743
.....	.....	.....	.....	.....	.....	.....	.....	.....
XP_023340278.1	zinc finger protein 131-like isoform X4 [Eurytemora affinis]	Eurytemora affinis	22.3	22.3	100%	10828	83.33%	322
.....	.....	.....	.....	.....	.....	.....	.....	.....
MAH17558.1	glycosyl transferase family 2 [Euryarchaeota archaeon]	Euryarchaeota archaeon	22.3	22.3	100%	10905	83.33%	225
.....	.....	.....	.....	.....	.....	.....	.....	.....
KAF6326367.1	hypothetical protein mPipKuh1_008369 [Pipistrellus kuhlii]	Pipistrellus kuhlii	21.8	21.8	100%	15143	83.33%	2762
XP_036293424.1	thyroglobulin [Pipistrellus kuhlii]	Pipistrellus kuhlii	21.8	21.8	100%	15143	83.33%	2761
.....	.....	.....	.....	.....	.....	.....	.....	.....

RNA pseudouridylate synthase, putative, partial [Plasmodium ovale]
Sequence ID: SBT73617.1 Length: 4584 Number of Matches: 1
Range 1: 2733 to 2738 GenPept Graphics

Score	Expect	Identities	Positives	Gaps
24.8 bits(51)	1309	6/6(100%)	6/6(100%)	0/6(0%)

```
Query  1     MAMTHK  6
             MAMTHK
Sbjct  2733  MAMTHK  2738
```

Pseudouridylate Synthase

A **Pseudouridylate Synthase** is an enzyme that catalyzes a chemical reaction in which *Uracil* and *D-ribose 5-phosphate* are involved. This enzyme belongs to the family of **lyases** (an enzyme that catalyzes the breaking of ***various chemical bonds by means*** other than *hydrolysis* and *oxidation*), specifically the **hydro-lyases**, which cleave carbon-oxygen bonds. This enzyme participates in *pyrimidine metabolism* (https://en.wikipedia.org/wiki/Pseudouridylate_synthase).

S9 family peptidase [Alteromonadaceae bacterium]
Sequence ID: MAJ70411.1Length: 804Number of Matches: 1

See 1 more title(s) See all Identical Proteins(IPG)

Related Information
Identical Proteins-Identical proteins to MAJ70411.1
Range 1: 133 to 138 GenPept Graphics

Score	Expect	Identities	Positives	Gaps
24.8 bits(51)	1318	6/6(100%)	6/6(100%)	0/6(0%)

```
Query  1     MAMTHK  6
             MAMTHK
Sbjct  133   MAMTHK  138
```

Peptidase and S9 family peptidase

A ***peptidase*** (also called **protease** or **proteinase**) is an **enzyme** that *catalyzes* (increases the rate of) **proteolysis**, the breakdown of proteins into smaller polypeptides or single amino acids. They do this by cleaving the *peptide bonds* within proteins by **hydrolysis**, a reaction where **water** breaks bonds (https://en.wikipedia.org/wiki/Protease). **Proteases** occur in all organisms, from prokaryotes to eukaryotes to viruses. These enzymes are involved in a multitude of physiological reactions from simple digestion of food proteins to highly regulated cascades (e.g., the ***blood-clotting* cascade**, the *complement system*, *apoptosis pathways*, and the *invertebrate prophenoloxidase*-activating cascade). The activity can be a *destructive change* (abolishing a protein's function or digesting it to its principal components), it can be an *activation of a function*, or it can be a *signal in a signalling pathway*. Proteases determine the *lifetime of other proteins* playing important physiological role like *hormones*, *antibodies*, or *other enzymes*. This is **one of the fastest "switching on" and "switching off" regulatory mechanisms in the physiology of an organism**. **By complex cooperative action the proteases may proceed as cascade reactions, which result in rapid and efficient amplification of an organism's response to a physiological signal** (https://en.wikipedia.org/wiki/Protease). **Bacteria** contain proteases responsible for general protein quality control by *degrading unfolded or misfolded proteins*. A *secreted bacterial protease* may also act as an exotoxin, and be an example of a **virulence factor in bacterial pathogenesis**. Bacterial exotoxic proteases destroy extracellular structures. The genomes of some **viruses** encode one massive polyprotein, which needs a protease to cleave this into functional units. These proteases have high specificity and only cleave a very restricted set of substrate sequences (https://en.wikipedia.org/wiki/Protease). They are therefore a common target for **protease inhibitors** ((https://en.wikipedia.org/wiki/Protease_inhibitor_(pharmacology)), a class of *antiviral drugs* that are widely used to treat **HIV/AIDS** and **hepatitis C**. **Protease inhibitors** prevent **viral replication** by selectively *binding to viral proteases* (e.g. *HIV-1 protease*) and *blocking proteolytic cleavage* of protein precursors that are necessary for the production of infectious viral particles ((https://en.wikipedia.org/wiki/Protease_inhibitor_(pharmacology)). Proteases were first grouped into **84 families** according to their evolutionary relationship, and classified under **four catalytic types**: **serine**, **cysteine**, **aspartic**, and **metallo proteases** (https://en.wikipedia.org/wiki/Protease). For the interests covered in this document, the **Serine Protease (Serine** is an amino acid that can be used for the production of metabolic energy) is the most important. **Peptidase family S9** contains a varied set of **serine-dependent peptidases**. Many peptidases in the family are believed to be important for the **degradation of biologically active peptide**. Members of the family are found in *different cellular locations*: *prolyl oligopeptidase in intracellular, DDP-IV and fibroblast activation protein α* are expressed on cell surfaces; and *oligopeptidase B* is secreted to the bacterial periplasm (https://www.ebi.ac.uk/merops/cgi-bin/famsum?family=S9).

L-serine ammonia-lyase [Campylobacter coli]
Sequence ID: EFS5897567.1 Length: 454 Number of Matches: 1
Range 1: 408 to 413 GenPept Graphics

Score	Expect	Identities	Positives	Gaps
24.8 bits(51)	1327	6/6(100%)	6/6(100%)	0/6(0%)

```
Query  1    MAMTHK  6
            MAMTHK
Sbjct  408  MAMTHK  413
```

L-serine ammonia-lyase

L-serine ammonia-lyase is an enzyme that catalyzes a chemical reaction that has one substrate, **L-serine**, and two products, *pyruvate* and NH_3. This enzyme belongs to the family of **lyases**, specifically *ammonia lyases*, which cleave *carbon-nitrogen* bonds. This enzyme participates in *glycine*, *serine* and *threonine metabolism* and *cysteine metabolism*. (https://en.wikipedia.org/wiki/L-serine_ammonia-lyase).

adenosine deaminase [Aeromonas salmonicida]
Sequence ID: VXA76768.1 Length: 346 Number of Matches: 1
Range 1: 117 to 122 GenPept Graphics

Score	Expect	Identities	Positives	Gaps
24.8 bits(51)	1333	6/6(100%)	6/6(100%)	0/6(0%)

```
Query  1    MAMTHK  6
            MAMTHK
Sbjct  117  MAMTHK  122
```

Adenosine deaminase

Adenosine deaminase (also known as **adenosine aminohydrolase**, or **ADA**) is an enzyme involved in *purine metabolism*. It is needed for the breakdown of adenosine from food and for the turnover of nucleic acids in tissues. Its primary function in humans is the *development and maintenance of the immune system*, but the full physiological role of ADA is not yet completely understood. The enzyme has been found in *bacteria, plants, invertebrates, vertebrates, and mammals*, with high conservation of amino acid sequence. **Zinc** is the only cofactor necessary for activity. **Elevated levels of ADA** has been associated with **AIDS**. However, **deficient levels of ADA** have been associated with ***pulmonary inflammation***, *thymic cell death*, and *defective T-cell receptor signaling* (https://en.wikipedia.org/wiki/Adenosine_deaminase).

alpha/beta hydrolase [Macromonas sp. BK-30]
Sequence ID: WP_101048850.1 Length: 291 Number of Matches: 1
Range 1: 246 to 251 GenPept Graphics

Score	Expect	Identities	Positives	Gaps
24.8 bits(51)	1337	6/6(100%)	6/6(100%)	0/6(0%)

```
Query  1    MAMTHK  6
            MAMTHK
Sbjct  246  MAMTHK  251
```

Alpha/beta hydrolase

The ***alpha/beta hydrolase superfamily*** is superfamily of **hydrolytic enzymes** of widely differing phylogenetic origin and catalytic function that shares a *common fold* that includes *proteases, lipases, peroxidases, esterases, epoxide hydrolases and dehalogenases*. The enzymes are believed to have diverged from a common ancestor, retaining *little obvious sequence similarity*, but *preserving the arrangement of the* ***catalytic residues***. All have a *catalytic triad*, the elements of which are borne on loops, which are the best-conserved structural features of the fold (https://en.wikipedia.org/wiki/Alpha/beta_hydrolase_superfamily).

DNA helicase PcrA [Bacillus cereus]
Sequence ID: WP_098544546.1Length: 743Number of Matches: 1

See 2 more title(s) See all Identical Proteins(IPG)

Related Information
Identical Proteins-Identical proteins to WP_098544546.1
Range 1: 5 to 10 GenPept Graphics

Score	Expect	Identities	Positives	Gaps
22.3 bits(45)	10726	5/6(83%)	5/6(83%)	0/6(0%)

```
Query  1   MAMTHK  6
           M MTHK
Sbjct  5   MSMTHK  10
```

DNA helicase PcrA

Plasmid copy reduced (**PcrA**) is a **helicase** [enzymes vital involved in the DNA replication, transcription, translation, recombination, DNA repair, and ribosome biogenesis, all processes involving the separation of nucleic acid (https://en.wikipedia.org/wiki/Helicase)]. Helicase is essential for *plasmid rolling-circle replication* (process of unidirectional nucleic acid replication that can rapidly synthesize multiple copies of circular molecules of DNA or RNA, such as plasmids, the genomes of bacteriophages, and the circular RNA genome of viroids https://en.wikipedia.org/wiki/Rolling_circle_replication) and *repair of DNA damage* caused by exposure to ultraviolet radiation. Helicase catalyzes the unwinding of double-stranded plasmid DNA that has been nicked at the replication origin by the *replication initiation protein*. Genetic and biochemical studies have also shown that the helicase plays an important **role in cell-survival by regulating the levels of RecA-mediated recombination in Gram-positive bacteria** (https://en.wikipedia.org/wiki/PcrA).

zinc finger protein 131-like isoform X4 [Eurytemora affinis]
Sequence ID: XP_023340278.1 Length: 322 Number of Matches: 1
Related Information
Gene-associated gene details
Genome Data Viewer-aligned genomic context
Range 1: 297 to 302 GenPept Graphics

Score	Expect	Identities	Positives	Gaps
22.3 bits(45)	10828	5/6(83%)	5/6(83%)	0/6(0%)

```
Query  1     MAMTHK  6
             M MTHK
Sbjct  297   MSMTHK  302
```

Zinc-finger proteins (ZNFs)

"**Zinc finger protein 131-like isoform X4**" is a *structurally modified **C2H2 Zn finger protein***. The **Zinc finger** are specific protein **regions capable of binding DNA**. The **Zinc finger C2H2** is a classical zinc finger domain. C2H2-type zinc fingers are ubiquitous; more than 1% of all mammalian proteins are predicted to contain at least one zinc finger. They often function as *DNA or protein binding structural motifs*, such as in *eukaryotic transcription factors*, and therefore they play important roles in **cellular processes** such as **development**, **differentiation**, and **oncosuppression**. C2H2 zinc finger proteins contain from 1 to more than 30 zinc finger repeats (https://www.ncbi.nlm.nih.gov/Structure/cdd/cddsrv.cgi?uid=275371). There are **many superfamilies of Znf motifs**, varying in both *sequence* and *structure*. They display considerable versatility in binding modes, even between members of the same class (e.g. some bind DNA, others protein), suggesting that *Znf motifs* are **stable scaffolds that have evolved specialised functions**. For example, *Znf-containing proteins* function in **gene transcription, translation, mRNA trafficking, cytoskeleton organisation, epithelial development, cell adhesion, protein folding, chromatin remodelling and zinc sensing**, to name but a few. **Zinc-binding motifs are stable structures, and they rarely undergo conformational changes upon binding their target** (http://www.ebi.ac.uk/interpro/entry/InterPro/IPR036236/).
Zinc finger protein 131 plays a role during **development** and **organogenesis**, as well as in the *function of the adult central nervous system*. It may be involved in **transcriptional regulation as a repressor of ERalpha signaling** (https://www.ebi.ac.uk/interpro/entry/InterPro/IPR027758/), mechanism of estrogen-induced transcriptional repression by using the **gene Erα** (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2738295/). It has been shown that **ZNFs** could also act as *recruiters of chromatin modifiers*, as *co-factors*, or as *structural proteins involved in cell migration and invasion.* In particular, the role of ZNFs in **cancer development, progression and metastasis** is becoming an interesting research issue (https://www.nature.com/articles/cddiscovery201771). **Zinc finger C2H2** can have a variety of functions such as binding RNA and mediating protein-protein interactions, but is best known for its role in *sequence-specific DNA-binding proteins* such as Zif268 (*Egr1*) (https://en.wikipedia.org/wiki/Zinc_finger). **EGR-1** (Early growth response protein 1) also known as **ZNF268** (*zinc finger protein 268*) or **NGFI-A** (nerve growth factor-induced protein A) is a protein that in humans is encoded by the **EGR1 gene**. **Egr1** is a *nuclear protein and functions as a transcriptional regulator*, especially cerebral. Beside its function in the ***nervous system***, there is significant evidence that EGR-1 along with its paralog EGR-2 is **induced in fibrotic diseases** (including **pulmonary fibrosis**) has key functions in fibrinogenesis and is **necessary for experimentally induced fibrosis in mice** (https://en.wikipedia.org/wiki/EGR1). The modular nature of the zinc finger motif allows for a large number of **combinations of DNA and RNA sequences to be bound with high degree of affinity and specificity**, and is therefore **ideally suited for engineering protein that can be targeted to and bind specific DNA sequences** (https://en.wikipedia.org/wiki/Zinc_finger).

glycosyl transferase family 2 [Euryarchaeota archaeon]
Sequence ID: MAH17558.1 Length: 225 Number of Matches: 1

See 1 more title(s) See all Identical Proteins(IPG)

Related Information
Identical Proteins-Identical proteins to MAH17558.1

Range 1: 123 to 128 GenPept Graphics

Score	Expect	Identities	Positives	Gaps
22.3 bits(45)	10905	5/6(83%)	5/6(83%)	0/6(0%)

```
Query  1    MAMTHK  6
            M MTHK
Sbjct  123  MSMTHK  128
```

Glycosyltransferases (**GTFs**, **Gtfs**)

Glycosyltransferases (**GTFs**, **Gtfs**) are enzymes that establish **natural glycosidic linkages**. They catalyze the transfer from an *glycosyl donor* to a *glycosyl acceptor*. **GTFs** have been widely used in the both targeted synthesis of *specific glycoconjugates* as well as the *synthesis of differentially glycosylated libraries of drugs*, biological probes or natural products in the context of drug discovery and drug development. Suitable enzymes can be isolated from natural sources or produced recombinantly (https://en.wikipedia.org/wiki/Glycosyltransferase).

hypothetical protein mPipKuh1_008369 [Pipistrellus kuhlii]

GenBank: KAF6326367.1

Identical Proteins FASTA Graphics

```
LOCUS       KAF6326367                6 aa            linear   MAM 07-AUG-2020
DEFINITION  hypothetical protein mPipKuh1 008369 [Pipistrellus kuhlii].
ACCESSION   KAF6326367 REGION: 1976..1981
VERSION     KAF6326367.1
DBLINK      BioProject: PRJNA628567
            BioSample: SAMN14734388
DBSOURCE    accession JACAGB010000014.1
KEYWORDS    .
SOURCE      Pipistrellus kuhlii (Kuhl's pipistrelle)
  ORGANISM  Pipistrellus kuhlii
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Laurasiatheria; Chiroptera; Microchiroptera;
            Vespertilionidae; Pipistrellus.
REFERENCE   1  (residues )
  AUTHORS   Jebb,D., Huang,Z., Pippel,M., Hughes,G.M., Lavrichenko,K.,
            Devanna,P., Winkler,S., Jermiin,L.S., Skirmuntt,E.C.,
            Katzourakis,A., Burkitt-Gray,L., Ray,D.A., Sullivan,K.A.M.,
            Roscito,J.G., Kirilenko,B.M., Davalos,L.M., Corthals,A.P.,
            Power,M.L., Jones,G., Ransome,R.D., Dechmann,D.K.N.,
            Locatelli,A.G., Puechmaille,S.J., Fedrigo,O., Jarvis,E.D.,
            Hiller,M., Vernes,S.C., Myers,E.W. and Teeling,E.C.
```

```
  TITLE     Six reference-quality genomes reveal evolution of bat adaptations
  JOURNAL   Nature 583 (7817), 578-584 (2020)
   PUBMED   32699395
REFERENCE   2  (residues )
  AUTHORS   Pippel,M., Jebb,D., Huang,Z., Huges,G., Lavrichenko,K., Devanna,P.,
            Winkler,S., Ray,D., Davalos,L., Hiller,M., Vernes,S., Myers,E. and
            Teeling,E.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-JUN-2020) Center for Systems Biology, Bat1K,
            Pfotenhauerstrasse 108, Dresden 01307, Germany
COMMENT     ##Genome-Assembly-Data-START##
            Assembly Date          :: 16-APR-2019
            Assembly Method        :: DAmar v. Apr 19; Bionano Solve DLS v.
                                      3.3; Salsa2 HiC v. git commit: e3ae7d8;
                                      GenomicConsensus v. git commit: 038de5c;
                                      longranger align v. 2.2.0; Freebayes v.
                                      1.2.0; HiGlass manual curation v. 1.5
            Assembly Name          :: mPipKuh1.p
            Genome Representation  :: Full
            Expected Final Version :: No
            Genome Coverage        :: 80.8x
            Sequencing Technology  :: PacBio Sequel CLR; 10X Genomics chromium
                                      linked reads; Bionano Genomics; Phase
                                      Genomics HiC; PacBio Sequel IsoSeq
            ##Genome-Assembly-Data-END##
            Method: conceptual translation.
FEATURES             Location/Qualifiers
     source          1..6
                     /organism="Pipistrellus kuhlii"
                     /isolate="mPipKuh1"
                     /db xref="taxon:59472"
                     /sex="male"
                     /tissue_type="flight muscle, heart, brain, liver, kidney"
                     /dev stage="adult"
                     /country="Italy: Bergamo"
                     /lat lon="45.7430 N 9.5831 E"
                     /collection date="2017"
                     /collected_by="Emma Teeling, Andrea Locatelli"
     Protein         <1..>18
                     /product="hypothetical protein"
     CDS             1..2762
                     /locus tag="mPipKuh1 008369"
                     /coded_by="complement(join(JACAGB010000014.1:2240776..2240
                     894,JACAGB010000014.1:2241844..2242034,
                     JACAGB010000014.1:2242480..2242614,
```

JACAGB010000014.1:2246900..2247007,
JACAGB010000014.1:2247844..2248025,
JACAGB010000014.1:2257005..2257172,
JACAGB010000014.1:2257998..2258162,
JACAGB010000014.1:2300308..2300513,
JACAGB010000014.1:2300944..2301100,
JACAGB010000014.1:2303310..2303403,
JACAGB010000014.1:2303757..2303976,
JACAGB010000014.1:2308051..2308215,
JACAGB010000014.1:2309052..2309186,
JACAGB010000014.1:2317381..2317443,
JACAGB010000014.1:2320248..2320391,
JACAGB010000014.1:2321291..2321370,
JACAGB010000014.1:2322200..2322311,
JACAGB010000014.1:2323980..2324156,
JACAGB010000014.1:2325150..2325287,
JACAGB010000014.1:2326454..2326534,
JACAGB010000014.1:2327257..2327322,
JACAGB010000014.1:2331050..2331214,
JACAGB010000014.1:2333332..2333523,
JACAGB010000014.1:2336493..2336601,
JACAGB010000014.1:2337878..2337993,
JACAGB010000014.1:2338855..2338971,
JACAGB010000014.1:2340643..2340813,
JACAGB010000014.1:2343501..2343650,
JACAGB010000014.1:2346138..2346359,
JACAGB010000014.1:2346755..2346911,
JACAGB010000014.1:2348000..2348154,
JACAGB010000014.1:2348817..2349029,
JACAGB010000014.1:2349606..2349806,
JACAGB010000014.1:2350841..2350943,
JACAGB010000014.1:2352126..2352238,
JACAGB010000014.1:2352686..2352763,
JACAGB010000014.1:2353074..2353205,
JACAGB010000014.1:2356543..2356782,
JACAGB010000014.1:2358175..2358765,
JACAGB010000014.1:2359776..2360858,
JACAGB010000014.1:2362365..2362550,
JACAGB010000014.1:2363176..2363319,
JACAGB010000014.1:2363841..2363947,
JACAGB010000014.1:2369030..2369189,
JACAGB010000014.1:2370085..2370288,
JACAGB010000014.1:2370675..2370772,
JACAGB010000014.1:2371050..2371158,
JACAGB010000014.1:2372530..2372596))"

/note="gene symbol: LOC00913"

ORIGIN

1 **mamshk**

//

hypothetical protein mPipKuh1_008369 [Pipistrellus kuhlii]
Sequence ID: KAF6326367.1 Length: 2762 Number of Matches: 1
Range 1: 1976 to 1981 GenPept Graphics

Score	Expect	Identities	Positives	Gaps
21.8 bits(44)	15143	5/6(83%)	5/6(83%)	0/6(0%)

```
Query  1     MAMTHK  6
             MAM HK
Sbjct  1976  MAMSHK  1981
```

thyroglobulin [Pipistrellus kuhlii]

NCBI Reference Sequence: XP_036293424.1

Identical Proteins FASTA Graphics

Go to:

LOCUS XP_036293424 6 aa linear MAM **01-OCT-2020**
DEFINITION thyroglobulin [Pipistrellus kuhlii].
ACCESSION XP_036293424 REGION: 1976..1981
VERSION XP_036293424.1
DBLINK BioProject: PRJNA666119
DBSOURCE REFSEQ: accession XM_036437531.1
KEYWORDS RefSeq.
SOURCE Pipistrellus kuhlii (Kuhl's pipistrelle)
ORGANISM Pipistrellus kuhlii
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Laurasiatheria; Chiroptera; Microchiroptera;
Vespertilionidae; Pipistrellus.
COMMENT MODEL REFSEQ: This record is predicted by automated computational
analysis. This record is derived from a genomic sequence
(NW_023425449.1) annotated using gene prediction method: Gnomon.
Also see:
Documentation of NCBI's Annotation Process

##Genome-Annotation-Data-START##
Annotation Provider :: NCBI

```
            Annotation Status           :: Full annotation
            Annotation Name             :: Pipistrellus kuhlii Annotation
                                           Release 100
            Annotation Version          :: 100
            Annotation Pipeline         :: NCBI eukaryotic genome annotation
                                           pipeline
            Annotation Software Version :: 8.5
            Annotation Method           :: Best-placed RefSeq; Gnomon
            Features Annotated          :: Gene; mRNA; CDS; ncRNA
            ##Genome-Annotation-Data-END##
            COMPLETENESS: full length.
FEATURES             Location/Qualifiers
     source          1..6
                     /organism="Pipistrellus kuhlii"
                     /isolate="mPipKuh1"
                     /db_xref="taxon:59472"
                     /chromosome="Unknown"
                     /sex="male"
                     /tissue_type="flight muscle, heart, brain, liver, kidney"
                     /dev_stage="adult"
                     /country="Italy: Bergamo"
                     /lat lon="45.7430 N 9.5831 E"
                     /collection_date="2017"
                     /collected_by="Emma Teeling, Andrea Locatelli"
     Protein         <1..>18
                     /product="thyroglobulin"
                     /calculated mol wt=294162
     CDS             1..2761
                     /gene="TG"
                     /coded by="XM 036437531.1:1..8286"
                     /db_xref="GeneID:118718652"
ORIGIN
        1 mamshk
//
```

thyroglobulin [Pipistrellus kuhlii]
Sequence ID: XP_036293424.1 Length: 2761 Number of Matches: 1
Range 1: 1976 to 1981 GenPept Graphics

Score	Expect	Identities	Positives	Gaps
21.8 bits(44)	15143	5/6(83%)	5/6(83%)	0/6(0%)

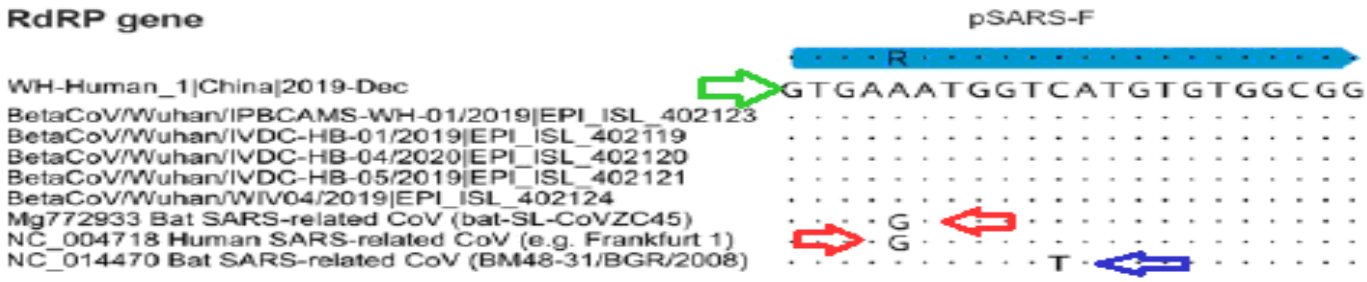
```
Query  1     MAMTHK  6
             MAM HK
Sbjct  1976  MAMSHK  1981
```

Thyroglobulin

Thyroglobulin (**Tg**) is a **dimeric glycoprotein** produced by the follicular cells of the thyroid and used entirely within the ***thyroid gland***. Thyroglobulin is in all vertebrates the main precursor to ***thyroid hormones***, which are produced when ***thyroglobulin's tyrosine residues*** are combined with *iodine* and the protein is subsequently cleaved (https://en.wikipedia.org/wiki/Thyroglobulin). It has been shown that ***thyroglobulin*** interacts with ***Binding immunoglobulin protein (BiP)***. ***BiP*** located in the lumen of the ***endoplasmic reticulum*** (**ER**), is an essential component of the ***translocation machinery***, as well as playing a role in *retrograde transport across the ER membrane of aberrant proteins destined for degradation by the proteasome*. ***BiP*** is an abundant protein under all growth conditions, but its synthesis is markedly induced under conditions that lead to the accumulation of unfolded *polypeptides in the ER* (https://pubmed.ncbi.nlm.nih.gov/10597629/). BiP has potent *immunological activity*, modulates some *cardiovascular* and *neurodegenerative diseases*, improves *metabolism*, is involved in ***bacterial DNA replication*** and therefore can be useful for the development of antibiotics. In contrast, **viruses rely on host BiP to successfully replicate, largely by infecting cells through cell-surface BiP, stimulating BiP expression to chaperone viral proteins, and suppressing the ER stress death response** (https://en.wikipedia.org/wiki/Binding_immunoglobulin_protein).

From the information reported in this paragraph, an important aspect determined by the addition of the *amino acid* **Lysine** (**K**) to the **M A M T H** sequence seems to concern the involvement of **enzymes** that perform ***catalytic functions***, in particular the ***lyase*** family (e.g., ***L-Serine***), ***adenosine deaminase***, ***peptidases*** (or ***proteases***), ***hydrolases***, ***glycosyltransferases***, i.e. many enzymes of ***the alpha/beta hydrolase superfamily***. The importance of some **proteins** involved in the processes of *DNA replication*, in *transcription*, *recombination* and *DNA repair* and in the *biogenesis of ribosomes*, all processes that involve the action of ***helicases*** and involve the separation of nucleic acid strands, is also evident. as well as in the ***mechanisms of translocation***.

Noteworthy are the **zing finger motifs** among the *significant alignments* repeatedly found in this and also in other subsequent chapters. These are proteins that, due to their modular nature, allow for binding a large number of **combinations of DNA and RNA sequences with a high degree of affinity and specificity, therefore ideal for engineering proteins that can be targeted and bind DNA sequences specifications** (https://en.wikipedia.org/wiki/Zinc_finger).
Still with reference to the **zinc finger motifs**, it is very important to underline that they have also been identified among the *significant alignments* [*Egretta garzetta zinc finger protein 639 (ZNF639), transcript variant X1, X2 and X3, mRNA*] of the RNA sequence "**GTGAAATGGTTATGTGTGGCGG**" of the *RdRp Gene of "Bat coronavirus BM48-31/BGR/2008, complete genome"*. This sequence is important because it is almost totally identical to the "**GTGAAATGGTCATGTGTGGCGG**" sequence, the RNA sequence studied in this document. Furthermore, as already seen in Chapter III in Fig. 3.1 (of which I report the relevant segment below), the sequence "**GTGAAATGGTTATGTGTGGCGG**" (in the segment, sequence indicated by the blue arrow) it is one of the sequences of the **RdRp gene** (*RNA-dependent RNA polymerase*, the motor of the **transcription** process of *SARS-CoV-2*) whose significant alignments of BLAST researches identify the ***Bat*** and the ***Pangolin*** (***Malay*** and ***Chinese***).



In addition to these "organisms", the other "organism" identified is precisely the ***Little Egretta***, with its significant alignments *Egretta garzetta zinc finger protein 639 (ZNF639), transcript variant X1, X2 and X3, mRNA*. It should now appear evident that the frequent presence of ***zinc finger protein*** among the significant alignments of this chapter (and also of the others) certainly cannot be considered random, but determined by the action of the *"Non-Obvious Trend"* of **Sequence n° 1/3** (and of other **modified sequences**).

9.10 SEQUENCE OF AMINO ACIDS "M A M T H K S Q"

This paragraph shows some results of the BLAST research carried out on the sequence of amino acids **M A M T H K S Q**, the sequence already considered in the ***previous paragraph*** and to which the amino acids **Serine** (**S**) and **Glutamine** (**Q**) have been added, respectively from the **AGT** and **CAA** ***codons***. The bases of these last two codons do not belong to **Sequence n° 1/3** (the sequence obtained from the modification of the RNA sequence studied in this document), but to the **HIV-1** sequence identified by **Sequence n° 1/3** and object of detailed study in these last paragraphs. For consistency, it was decided to continue studying subsequent codons until meeting the STOP codon.

QUERY: M A M T H K S Q (atg gca atg aca cac aag agt caa)

Search Parameters	
Program	blastp
Word size	2
Expect value	200000
Hitlist size	**10**
Gapcosts	9,1
Matrix	PAM30
Filter string	F
Genetic Code	1
Window Size	40
Threshold	11
Composition-based stats	0

10 sequences selected:

Accession Seq	Description	Comomm Name	Max Score	Total Score	Query C.	E value	Per. Id.	Acc. Len
NVK54495.1	**S9 family peptidase [Alteromonadaceae bacterium]**	Alteromonadaceae bacterium	28.2	28.2	100 %	130	87.50%	804
KRX08649.1	**Immunoglobulin E-set [Pseudocohnilembus persalinus]**	Pseudocohnilembus persalinus	27.8	27.8	100 %	185	87.50%	384
MBC8272682.1	**flagellar basal body rod protein FlgB [Gammaproteobacteria bacterium]**	Gammaproteobacteria bacterium	27.8	27.8	100 %	191	87.50%	132
EFS5897567.1	***L-serine ammonia-lyase [Campylobacter coli]***	Campylobacter coli	27.4	27.4	87%	262	100.00 %	454
WP_071431232.1	**phage late control D family protein [Tatumella sp. TA1]**	Tatumella sp. TA1	27.4	27.4	87%	263	100.00 %	354
WP_013071004.1	**TM1802 family CRISPR-associated protein [Zunongwangia profunda]**	Zunongwangia profunda	26.9	26.9	100 %	370	87.50%	669
AOS87527.1	**neuraminidase [Influenza A virus (A/mallard/Alberta/177/2004(H7N9))]**	**Influenza A virus (A/mallard/Alberta/177/2004(H7N9))**	26.5	26.5	100 %	526	87.50%	470
.....	.....	.....	...	...	...	...	.....	...

S9 family peptidase [Alteromoaceae bacterium]
Sequence ID: NVK54495.1 Length: 804 Number of Matches: 1
Range 1: 133 to 140 GenPept Graphics

Score	Expect	Identities	Positives	Gaps
28.2 bits(59)	130	7/8(88%)	7/8(87%)	0/8(0%)

```
Query  1    MAMTHKSQ  8
            MAMTHK Q
Sbjct  133  MAMTHKAQ  140
```

Peptidase and S9 family peptidase

See the discussion on page 229.

Immunoglobulin E-set [Pseudocohnilembus persalinus]
Sequence ID: KRX08649.1 Length: 384 Number of Matches: 1
Range 1: 269 to 276 GenPept Graphics

Score	Expect	Identities	Positives	Gaps
27.8 bits(58)	185	7/8(88%)	7/8(87%)	0/8(0%)

```
Query  1    MAMTHKSQ  8
            M MTHKSQ
Sbjct  269  MTMTHKSQ  276
```

Immunoglobulin E-set

Immunoglobulin E-set are **immunoglobulin-like proteins (Ig)** very common in animals. The "early" set **(E-set)** domains are possibly related to the **immunoglobulin** (IPR007110) and/or **fibronectin type III** (IPR003961) ***Ig-like protein superfamilies***. *Ig-like E-set domains* include also part of the superfamily of *enzymes* called **O-Glycosyl hydrolases** (3.2.1.) [(https://www.ebi.ac.uk/interpro/entry/InterPro/IPR014756/); (https://www.ebi.ac.uk/interpro/entry/InterPro/IPR004185/); see also what is written on page 230 on the ***alpha/beta hydrolase superfamily*** and on page 233 on ***glycosyltransferases***] and ***Arrestin family*** of proteins. These latter proteins are important for ***regulating signal transduction at G protein-coupled receptors***. The **Sequence ID: KRX08649.1** *"Immunoglobulin E-set"* identified above refers to a part of a sequence of "***Arrestin (or S-antigen), N-terminal domain***". **Arrestins** were first discovered as a part of a conserved two-step mechanism for *regulating the activity of G protein-coupled receptors (GPCRs) in the visual rhodopsin* (http://pfam.xfam.org/family/Arrestin_N).

flagellar basal body rod protein FlgB [Gammaproteobacteria bacterium]
Sequence ID: MBC8272682.1 Length: 132 Number of Matches: 1
Range 1: 62 to 69 GenPept Graphics

Score	Expect	Identities	Positives	Gaps
27.8 bits(58)	191	7/8(88%)	8/8(100%)	0/8(0%)

```
Query  1   MAMTHKSQ  8
           MAMTHKS+
Sbjct  62  MAMTHKSH  69
```

Flagellar basal body rod protein FlgB, N-terminal

Many ***bacterial species*** swim actively by means of ***flagella***. A ***bacterial-type flagellum*** is a motor complex composed of an extracellular helical protein filament coupled to a rotary motor embedded in the cell envelope (https://www.ebi.ac.uk/QuickGO/GTerm?id=GO:0071973). One of the parts of the ***flagellum organelle*** is the *basal body*. In *Salmonella typhimurium* and related organisms the *rod* has been shown to consist of **four different**, **yet evolutionary related proteins**: in the *distal portion* of the rod there are about 26 subunits of **protein flgG** and in the *proximal portion* there are about six subunits each of **proteins flgB, flgC, and flgF**. These **four proteins** contain a *highly conserved asparagine-rich domain at their N terminus* (https://www.ebi.ac.uk/interpro/entry/InterPro/IPR001444/).

L-serine ammonia-lyase [Campylobacter coli]
Sequence ID: EFS5897567.1 Length: 454 Number of Matches: 1
Range 1: 408 to 414 GenPept Graphics

Score	Expect	Identities	Positives	Gaps
27.4 bits(57)	262	7/7(100%)	7/7(100%)	0/7(0%)

```
Query  1    MAMTHKS  7
            MAMTHKS
Sbjct  408  MAMTHKS  414
```

L-serine ammonia-lyase

See the discussion on page 230.

phage late control D family protein [Tatumella sp. TA1]
Sequence ID: WP_071431232.1 Length: 354 Number of Matches: 1

See 1 more title(s) See all Identical Proteins(IPG)

Related Information
Identical Proteins-Identical proteins to WP_071431232.1
Range 1: 1 to 7 GenPept Graphics

Score	Expect	Identities	Positives	Gaps
27.4 bits(57)	263	7/7(100%)	7/7(100%)	0/7(0%)

```
Query  1  MAMTHKS  7
          MAMTHKS
Sbjct  1  MAMTHKS  7
```

Phage late control gene D protein (GPD)

This family includes a ***number of phage*** (or ***Bacteriophages***) ***late control gene D proteins and related bacterial sequences***. This family also includes ***Bacteriophage Mu P proteins*** and related sequences (https://www.ncbi.nlm.nih.gov/Structure/cdd/PF05954). ***Bacteriophage*** is a virus that infects and replicates within *bacteria* and *archaea* (https://en.wikipedia.org/wiki/Bacteriophage). **Phages** have been used as an alternative to antibiotics and they are seen as a possible therapy against multi-drug-resistant strains of many bacteria and **useful for counteracting bioweapons and toxins,** such as **anthrax and botulism**. **Bacteriophage researchers have been developing engineered viruses to overcome antibiotic resistance, and engineering the phage genes responsible for coding enzymes that degrade the biofilm matrix, phage structural proteins**, **and the enzymes responsible for lysis of the bacterial cell wall**. On the other hand, some specie of phages have been shown to complicate **biofilms** involved in ***pneumonia*** and ***cystic fibrosis and to shelter the bacteria from drugs meant to eradicate disease, thus promoting persistent infection*** (https://en.wikipedia.org/wiki/Bacteriophage).

TM1802 family CRISPR-associated protein [Zunongwangia profunda]
Sequence ID: WP_013071004.1 Length: 669 Number of Matches: 1

See 1 more title(s) See all Identical Proteins(IPG)

Related Information
Identical Proteins-Identical proteins to WP_013071004.1
Range 1: 548 to 555 GenPept Graphics

Score	Expect	Identities	Positives	Gaps
26.9 bits(56)	370	7/8(88%)	7/8(87%)	0/8(0%)

```
Query  1    MAMTHKSQ  8
            M MTHKSQ
Sbjct  548  MDMTHKSQ  555
```

CRISPR/Cas system-associated protein Cas8a1

CRISPR (*Clustered Regularly Interspaced Short Palindromic Repeats*) and ***associated Cas proteins*** comprise a ***system for heritable host defense*** by prokaryotic cells against ***phage*** and other ***foreign DNA***; Large proteins, some contain **Zn-finger** domain; signature gene for I-A subtype; also known as TM1802 family (https://www.ncbi.nlm.nih.gov/Structure/cdd/cddsrv.cgi?uid=304414).

neuraminidase [Influenza A virus (A/mallard/Alberta/177/2004(H7N9))]

Sequence ID: AOS87527.1 Length: 470 Number of Matches: 1

Range 1: 310 to 317 GenPept Graphics

Score	Expect	Identities	Positives	Gaps
26.5 bits(55)	526	7/8(88%)	7/8(87%)	0/8(0%)

```
Query  1    MAMTHKSQ  8
            MAMTH SQ
Sbjct  310  MAMTHTSQ  317
```

Neuraminidase (Sialidase)

Neuraminidase are **glycoside hydrolase** enzymes that cleave (cut) the glycosidic linkages of *neuraminic acids*. *Neuraminidase enzymes* are a *large family*, found in a range of organisms. The best-known neuraminidase is the **viral neuraminidase**, a drug target for the prevention of the spread of **influenza infection**. The *viral neuraminidases* are frequently used as **antigenic** determinants found on the **surface of the influenza virus**. **Some variants of the influenza neuraminidase confer more virulence to the virus than others. Sialidases may act as pathogenic factors in microbial infections** (https://en.wikipedia.org/wiki/Neuraminidase). **Viral neuraminidase** is a type of *neuraminidase* found on the ***surface of influenza viruses*** that enables the virus to be released from the host cell. ***Neuraminidases*** are enzymes that ***cleave sialic acid*** (also called *neuraminic acid*) groups from **glycoproteins**. ***Neuraminidase inhibitors*** are *antiviral agents that inhibit influenza viral neuraminidase activity and are of major importance in the control of influenza* (https://en.wikipedia.org/wiki/Viral_neuraminidase). ***Viral neuraminidases*** are the members of the ***Glycoside hydrolase family* 34** (CAZY GH_34) which comprises enzymes with only one known activity; **sialidase** or **neuraminidase** (EC 3.2.1.18). *Neuraminidases* cleave the terminal sialic acid residues from carbohydrate chains in *glycoproteins*. **Sialic acid** is a negatively charged sugar associated with the protein and lipid portions of lipoproteinsm (https://en.wikipedia.org/wiki/Viral_neuraminidase). **To infect a host cell, *influenza virus* attaches to the exterior cell surface using hemagglutinin, a molecule found on the surface of the virus that binds to sialic acid groups**. **Sialic acids** are found on various *glycoproteins* at the host cell surface. The *virus* then moves from *sialic acid group* to *sialic acid group* until it finds the **proper cell surface receptor** (whose identity remains unknown). *Neuraminadase* enables this movement by *cleaving sialic acid groups that hemagglutinin was attached to*. **After the virus has entered the cell and has replicated, new viral particles bud from the host cell membrane** (https://en.wikipedia.org/wiki/Viral_neuraminidase). The **hemagglutinin** on new viral particles remains attached to sialic acid groups of glycoproteins on the external cell surface and on the surface of other viral

particles; neuraminadase cleaves these groups and thereby allows the release of viral particles and prevents self-aggregation (https://en.wikipedia.org/wiki/Viral_neuraminidase). ***Neuraminidase inhibitors*** (**NAIs**) are a class of drugs which block the *neuraminidase enzyme*. They are commonly used as **antiviral drugs** because they *block the function of viral neuraminidases of the influenza virus, by preventing its reproduction by budding from the host cell* (https://en.wikipedia.org/wiki/Neuraminidase_inhibitor). *Oseltamivir (Tamiflu) un profarmaco, Zanamivir (Relenza), Laninamivir (Inavir)* and Peramivir belong to this class of substances (https://en.wikipedia.org/wiki/Viral_neuraminidas). After the ***H1N1*** *pandemic* in 2009 (**Influenza A virus subtype H1N1** is a subtype of ***Influenza A virus***. Well known outbreaks of ***H1N1*** strains in humans include the 2009 *swine flu pandemic*, the 1977 *Russian flu pandemic* as well as the 1918 *flu pandemic*), the effectiveness of **early treatment with neuraminidase inhibitors** in reducing serious cases and deaths was reported in various countries (https://en.wikipedia.org/wiki/Influenza_A_virus_subtype_H1N1; https://en.wikipedia.org/wiki/Neuraminidase_inhibitor).
Influenza A virus causes influenza in *birds* and some *mammals*, and is the only species of the genus *Alphainfluenzavirus* of the virus family *Orthomyxoviridae*. Strains of all subtypes of *influenza A virus* have been isolated from ***wild birds***, although disease is uncommon. Some isolates of *influenza A virus* cause severe disease both in ***domestic poultry*** and, rarely, in ***humans***. Occasionally, viruses are transmitted from ***wild aquatic birds*** to ***domestic poultry***, and this may cause an outbreak or give rise to **human influenza pandemics** (https://en.wikipedia.org/wiki/Influenza_A_virus). ***Influenza A viruses*** are *negative-sense*, single-stranded, segmented RNA viruses. The **several subtypes** are labeled according to an **H number** (for the type of hemagglutinin) and an **N number** (for the type of neuraminidase). There are 18 different known H antigens (H1 to H18) and 11 different known N antigens (N1 to N11). **H17N10** was isolated from ***fruit bats*** in 2012. ***H18N11*** was discovered in a ***Peruvian bat*** in 2013 (https://en.wikipedia.org/wiki/Influenza_A_virus). Each virus subtype has mutated into a variety of strains with differing pathogenic profiles; some are pathogenic to one species but not others, some are pathogenic to multiple species. A filtered and purified ***influenza A vaccine*** for humans has been developed and many countries have stockpiled it to allow a quick administration to the population in the event of an ***avian influenza pandemic***. In 2011, researchers reported the discovery of **an antibody effective against all types of the influenza A virus**. It is **FI6**, an **antibody** that targets a protein found on the surface of all influenza A viruses called *hemagglutinin*. **FI6** is the only known antibody found to bind all 16 subtypes of the influenza A virus hemagglutinin and is hoped to be useful for a **universal influenza virus therapy** (https://en.wikipedia.org/wiki/Influenza_A_virus; https://en.wikipedia.org/wiki/FI6_(antibody)).

The addition of the *amino acids* **Serine** (**S**) and **Glutamine** (**Q**) to the **M A M T H** sequence confirms the importance, already mentioned in the previous paragraph, of the **enzymes** that perform ***catalytic functions***, in particular the ***peptidases*** (or ***proteases***), ***L-serine ammonia-lyase*** and ***O-Glycosyl hydrolase***. Also important is the presence of *significant alignments* concerning the ***proteins of the gene D of late phage control*** (or ***bacteriophages***), the latter organisms used **in the development of engineered viruses responsible for encoding the enzymes that degrade the biofilm matrix of bacteria**. However, in some cases, the opposite has also been shown, namely that they complicate the biofilms involved in ***pneumonia*** and ***cystic fibrosis*** and ***protect the bacteria from drugs designed to eradicate the disease, thus promoting persistent infection***. Finally, very important for the purposes of our study, the significant alignment with **Neuraminidase (*Sialidase*)**, an enzyme of the **glycoside hydrolase** family involved in **Influenzavirus A**. **Some variants of the influenza neuraminidase confer more virulence to the virus than others. Sialidases may act as pathogenic factors in microbial infections**.

9.11 SEQUENCE OF AMINO ACIDS "M A M T H K S Q M K"

This paragraph shows some results of the BLAST research carried out on the sequence of amino acids **M A M T H K S Q M K**, the sequence already considered in the ***previous paragraph*** and to which the amino acids **Methionine** (**M**) and **Lysine** (**K**) have been added, determined respectively by the **ATG** and **AAA *codons***.

QUERY: M A M T H K S Q M K (**atg gca atg aca cac aag agt caa atg aaa**)

Search Parameters	
Program	blastp
Word size	2
Expect value	200000
Hitlist size	1000
Gapcosts	9,1
Matrix	PAM30
Filter string	F
Genetic Code	1
Window Size	40
Threshold	11
Composition-based stats	0

1000 sequences selected:

Accession Seq	Description	Comomm Name	Max Score	Total Score	Query C.	E value	Per. Id.	Acc. Len
.....	.....	.....	.....	.....	.....	....	.....	.....
EQB47646.1	GlcNAc-PI de-N-acetylase [Colletotrichum gloeosporioides Cg-14]	NA	29.1	29.1	90%	100	77.78%	698
.....	.....	.....	.....	.....	.....	....	.....	.....
KAF8849934.1	LmbE-like protein [Acephala macrosclerotiorum]	NA	29.1	29.1	90%	101	81.82%	285
.....	.....	.....	.....	.....	.....	....	.....	.....
GAF02461.1	MlotiK1 channel [Saccharicrinis fermentans DSM 9555 = JCM 21142]	NA	28.6	28.6	80%	143	87.50%	223
.....	.....	.....	.....	.....	.....	....	.....	.....
RKP30872.1	DNA/RNA polymerase [Metschnikowia bicuspidata]	NA	28.2	28.2	100%	200	72.73%	1007
MBF0136189.1	bacteriohemerythrin [Magnetococcales bacterium]	NA	28.2	28.2	70%	200	100.00%	957
.....	.....	.....	.....	.....	.....	....	.....	.....
WP_013071004.1	TM1802 family CRISPR-associated protein [Zunongwangia profunda]	NA	28.2	28.2	100%	200	80.00%	669
WP_008041925.1	hybrid sensor histidine kinase/response regulator [Reinekea blandensis]	NA	28.2	28.2	70%	201	100.00%	579
.....	.....	.....	.....	.....	.....	....	.....	.....
WP_057784003.1	glycoside hydrolase family 65 protein [Pediococcus parvulus]	NA	27.8	27.8	100%	284	70.00%	753

WP_169115272.1	2-methylaconitate cis-trans isomerase PrpF [Duganella sp. AF9R3]	NA	27.8	27.8	80%	286	87.50%	393
WP_107184309.1	pirin family protein [Photobacterium leiognathi]	NA	27.8	27.8	80%	287	87.50%	276
PES82498.1	glycosyl transferase family 2 [Bacillus anthracis]	NA	27.4	27.4	100%	402	70.00%	1115
WP_098905904.1	glycosyltransferase [Bacillus cereus]	NA	27.4	27.4	100%	402	70.00%	1115
TVY91158.1	Nuclear GTPase SLIP-GC [Lachnellula willkommii]	NA	26.5	26.5	70%	811	100.00%	1050
WP_132469210.1	serine hydrolase [Novosphingobium sp. ST904]	NA	26.5	26.5	90%	814	77.78%	473

GlcNAc-PI de-N-acetylase [Colletotrichum gloeosporioides Cg-14]
Sequence ID: EQB47646.1 Length: 698 Number of Matches: 1
Range 1: 282 to 290 GenPept Graphics

Score	Expect	Identities	Positives	Gaps
29.1 bits(61)	100	7/9(78%)	7/9(77%)	0/9(0%)

```
Query  1    MAMTHKSQM  9
           M M HKSQM
Sbjct  282  MTMAHKSQM  290
```

GlcNAc-PI de-N-acetylase

Members of this family are related to ***PIG-L an N-acetylglucosaminylphosphatidylinositol de-N-acetylase*** that catalyses the second step in GPI biosynthesis (https://pfam.xfam.org/family/PF02585). Deepak K. Sharm e Michael A. J. Ferguson explain that "***Glycosyl-phosphatidylinositol*** (**GPI**) **membrane anchors** are present in organisms at most stages of eukaryotic evolution, including protozoa, yeast, slime moulds, invertebrates and vertebrates, and are found on a diverse range of *proteins*. They are primarily responsible for the **anchoring of cell-surface proteins in the plasma membrane** (or in some cases to the topologically equivalent lumenal surface of secretory vesicles) and may be considered as an alternative to the hydrophobic transmembrane polypeptide anchor of type-1 integral membrane proteins. Many other functions have been proposed (though some remain controversial) for **GPI anchors**, including roles in **intracellular sorting**, **transmembrane signalling**, and **novel endocytic processes**. Most of these proposed functions are dependent on the hypothesis that a **GPI anchor** might allow proteins to associate in **specialised membrane microdomains** (https://link.springer.com/chapter/10.1007/978-3-642-61126-1_28)".

LmbE-like protein [Acephala macrosclerotiorum]
Sequence ID: KAF8849934.1 Length: 285 Number of Matches: 1
Range 1: 249 to 259 GenPept Graphics

Score	Expect	Identities	Positives	Gaps
29.1 bits(61)	101	9/11(82%)	9/11(81%)	2/11(18%)

```
Query  2    AMT--HKSQMK  10
            AMT  HKSQMK
Sbjct  249  AMTTAHKSQMK  259
```

LmbE-like protein

S. Viars, J. Valentine, and M. Hernick explain that "The **LmbE-like superfamily** is comprised of a **series of enzymes** that use a *single catalytic metal* ion to catalyze the **hydrolysis** of various *substrates*. These *substrates* are often key ***metabolites*** for eukaryotes and prokaryotes, which makes the **LmbE-like enzymes** important targets for drug development….this is the newest **superfamily of metallohydrolases**…..The majority of **enzymes** identified to date are *metal-dependent deacetylases* that catalyze the **hydrolysis** of a *N-acetylglucosamine moiety* on substrate using a combination of amino acid side chains and a single bound metal ion, predominantly **zinc**. The **catalytic zinc** is coordinated to proteins via **His2-Asp-solvent binding site**" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4101496/). Members of this family include *N-acetylglucosaminyl-phosphatidylinositol de-N-acetylase* (3.5.1.89) that catalyses the second step in ***glycosylphosphatidylinositol (GPI) biosynthesis*** and *1D-myo-inositol 2-acetamido-2-deoxy-alpha-D-glucopyranoside deacetylase* (3.5.1.103), involved in the **biosynthesis of mycothiol** (an unusual *thiol compound* found in **Actinobacteria**) (https://www.ebi.ac.uk/interpro/entry/InterPro/IPR003737/).

MlotiK1 channel [Saccharicrinis fermentans DSM 9555 = JCM 21142]
Sequence ID: GAF02461.1 Length: 223 Number of Matches: 1
Range 1: 1 to 8 GenPept Graphics

Score	Expect	Identities	Positives	Gaps
28.6 bits(60)	143	7/8(88%)	7/8(87%)	0/8(0%)

```
Query  3  MTHKSQMK  10
          MTHK QMK
Sbjct  1  MTHKNQMK  8
```

MlotiK1 channel

Cyclic nucleotide-regulated ion channels (*nucleotide* del *Mesorhizobium* loti , MlotiK1) are present in bacteria, plants, vertebrates, and humans. In higher organisms, they are closely involved in **signaling networks of vision** and **olfaction**. Binding of cAMP or cGMP favors the activation of these ion channels (https://www.pnas.org/content/108/51/20802). To deepen the study of the **role played by these ion channels in visual and olfactory functions**, consult this interesting article: https://journals.physiology.org/doi/pdf/10.1152/physrev.00008.2002.

DNA/RNA polymerase [Metschnikowia bicuspidata]
Sequence ID: RKP30872.1 Length: 1007 Number of Matches: 1
Range 1: 366 to 376 GenPept Graphics

Score	Expect	Identities	Positives	Gaps
28.2 bits(59)	200	8/11(73%)	8/11(72%)	1/11(9%)

```
Query  1    MAMTHKS-QMK  10
            MAM HK  QMK
Sbjct  366  MAMAHKTGQMK  376
```

DNA-directed RNA polymerase N-terminal (pfam14700)

This is the ***N-terminal domain of DNA-directed RNA polymerase***. This domain has a role in interaction with ***regions of upstream promoter DNA and the nascent RNA chain***, leading to the **processivity of the enzyme**. In order to make **mRNA transcripts** the **RNA polymerase** undergoes a *transition* from the *initiation phase* (which only makes short fragments of RNA) to an *elongation phase*. This domain undergoes a **structural change** in the transition from initiation to elongation phase. The structural change results in **abolition of the promoter binding site**, creation of a **channel accommodating the heteroduplex in the active site and formation of an exit tunnel which the RNA transcript passes through after peeling off the heteroduplex** (https://www.ncbi.nlm.nih.gov/Structure/cdd/cddsrv.cgi?uid=317144; http://www.ebi.ac.uk/interpro/entry/InterPro/IPR002092/).

bacteriohemerythrin [Magnetococcales bacterium]
Sequence ID: MBF0136189.1 Length: 957 Number of Matches: 1
Range 1: 58 to 64 GenPept Graphics

Score	Expect	Identities	Positives	Gaps
28.2 bits(59)	200	7/7(100%)	7/7(100%)	0/7(0%)

```
Query  3   MTHKSQM  9
           MTHKSQM
Sbjct  58  MTHKSQM  64
```

bacteriohemerythrin

Hemerythrin is an **oligomeric protein** responsible for **oxygen (O_2) transport** in the marine invertebrate phyla of sipunculids, priapulids, brachiopods, and in a single annelid worm genus, *Magelona*. Recent evidence has revealed hemerythrin to be a **multi-functional protein** and contributing to **innate immunity** and anterior **tissue regeneration** in worms. **Hemerythrin/HHE** (H-HxxxE-HxxxH-HxxxxD) **proteins** found in **bacteria** are implicated in **signal transduction** and **chemotaxis**. More distantly related ones include H-HxxxE-H-HxxxE proteins (including the **E3 ligase**) and animal **F-box proteins** (H-HExxE-H-HxxxE) (https://en.wikipedia.org/wiki/Hemerythrin).

TM1802 family CRISPR-associated protein [Zunongwangia profunda]
Sequence ID: WP_013071004.1 Length: 669 Number of Matches: 1

See 1 more title(s) See all Identical Proteins(IPG)

Related Information
Identical Proteins-Identical proteins to WP_013071004.1
Range 1: 548 to 557 GenPept Graphics

Score	Expect	Identities	Positives	Gaps
28.2 bits(59)	200	8/10(80%)	8/10(80%)	0/10(0%)

```
Query  1    MAMTHKSQMK  10
            M MTHKSQ K
Sbjct  548  MDMTHKSQQK  557
```

See the discussion on page 244.

hybrid sensor histidine kinase/response regulator [Reinekea blandensis]
Sequence ID: WP_008041925.1 Length: 579 Number of Matches: 1

See 1 more title(s) See all Identical Proteins(IPG)

Related Information
Identical Proteins-Identical proteins to WP_008041925.1
Range 1: 191 to 197 GenPept Graphics

Score	Expect	Identities	Positives	Gaps
28.2 bits(59)	201	7/7(100%)	7/7(100%)	0/7(0%)

```
Query  3    MTHKSQM  9
            MTHKSQM
Sbjct  191  MTHKSQM  197
```

Histidine kinase (HK)

Histidine kinases are proteins of the **transferase class of enzymes** that play a role in **signal transduction** across the **cellular membrane**. HKs can act as **cellular receptors** *for signaling molecules* in a way analogous to **tyrosine kinase receptors** (RTK). **Multifunctional receptor molecules such as HKs and RTKs** typically have portions on the *outside of the cell* that bind to *hormone- or growth factor-like molecules*, portions that span the *cell membrane*, and portions *within the cell* that contain the enzymatic activity. In **addition to kinase activity**, the intracellular domains typically have regions that bind to a secondary effector molecule or complex of molecules that *further propagate signal transduction within the cell*. **This type of enzyme is involved in signal transduction pathways upstream of many cellular processes including various metabolic, virulence, and homeostatic pathways** (https://en.wikipedia.org/wiki/Histidine_kinase).

family 65 glycosyl hydrolase [Pediococcus sp.]
Sequence ID: HBO47983.1 Length: 753 Number of Matches: 1
Range 1: 219 to 228 GenPept Graphics

Score	Expect	Identities	Positives	Gaps
27.8 bits(58)	284	7/10(70%)	9/10(90%)	0/10(0%)

```
Query  1    MAMTHKSQMK  10
            M MTHKS++K
Sbjct  219  MEMTHKSELK  228
```

family 65 glycosyl hydrolase

Glycoside hydrolase family 65 is a family of **glycoside hydrolases**. Glycoside hydrolases are a widespread group of enzymes that hydrolyse the glycosidic bond between two or more carbohydrates, or between a carbohydrate and a non-carbohydrate moiety (https://en.wikipedia.org/wiki/Glycoside_hydrolase_family_65). **Glycoside hydrolases** (also called **glycosidases** or **glycosyl hydrolases**) catalyze the hydrolysis of glycosidic bonds in complex sugars. They are extremely common enzymes with roles in nature including degradation of biomass such as cellulose, hemicellulose, and starch, in anti-bacterial defense strategies, in pathogenesis mechanisms (e.g., **viral neuraminidases**) and in normal cellular function. **Together with glycosyltransferases, glycosidases form the major catalytic machinery for the synthesis and breakage of glycosidic bonds** (https://en.wikipedia.org/wiki/Glycoside_hydrolase).

2-methylaconitate cis-trans isomerase PrpF [Duganella sp. AF9R3]
Sequence ID: WP_169115272.1Length: 393Number of Matches: 1

See 1 more title(s) See all Identical Proteins(IPG)
Related Information
Identical Proteins-Identical proteins to WP_169115272.1
Range 1: 1 to 8 GenPept Graphics

Score	Expect	Identities	Positives	Gaps
27.8 bits(58)	286	7/8(88%)	7/8(87%)	0/8(0%)

```
Query  3  MTHKSQMK  10
          MTHK QMK
Sbjct  1  MTHKPQMK  8
```

2-methylaconitate cis-trans isomerase PrpF

The **2-methylcitrate cycle** is one of at least **five degradation pathways for propionate** via *propionyl-CoA*. Degradation of propionate toward pyruvate consumes oxaloacetate and releases succinate. This family consists of **PrpF**, an incompletely characterized protein that appears to be an essential accessory *protein for the Fe/S-dependent 2-methylisocitrate dehydratase AcnD*. This protein is related to but distinct from **FldA** (part of Pfam family PF04303), a putative fluorene degradation protein of Sphingomonas sp. LB126 (https://www.ncbi.nlm.nih.gov/genome/annotation_prok/evidence/TIGR02334/).

pirin family protein [Photobacterium leiognathi]
Sequence ID: WP_107184309.1Length: 276Number of Matches: 1

See 1 more title(s) See all Identical Proteins(IPG)

Related Information
Identical Proteins-Identical proteins to WP_107184309.1
Range 1: 93 to 100 GenPept Graphics

Score	Expect	Identities	Positives	Gaps
27.8 bits(58)	287	7/8(88%)	7/8(87%)	0/8(0%)

```
Query  3   MTHKSQMK  10
           M HKSQMK
Sbjct  93  MAHKSQMK  100
```

Pirin family protein (Pirin_C)

Pirin C-terminal cupin domain is found the **C-terminal** half of the **Pirin** protein (https://www.ncbi.nlm.nih.gov/Structure/cdd/cddsrv.cgi?uid=331824; https://pfam.xfam.org/family/PF05726). *Eukaryotic pirins* are highly conserved nuclear proteins that may **function as transcriptional regulators with a role in apoptosis**. *Prokaryotic homologue*s have also been identified. Both *bacterial* and *human* **pirins** have been shown to possess **quercetinase activity**, although this is not universally true for all family members - YhaK (SWISSPROT), for example, displays no such enzymatic activity. Pirin is composed of *two structurally similar domains* arranged face to face (http://pfam.xfam.org/family/PF05726). Although the two domains are similar, the *C-terminal domain* of pirin differs from the *N-terminal domain* as it does not contain a *metal binding site* and its sequence does not contain the conserved *metal-coordinating residue*. **Pirin** is considered a member of the "**cupin" superfamily** on the basis of primary sequence and structural similarity. The **presence of a metal binding site** in the **N-terminal beta-barrel of pirin**, may be significant in its interaction with **Bcl-3 and nuclear factor I (NFI) and role in regulating NF-kappaB transcription factor activity** (http://pfam.xfam.org/family/PF05726).

glycosyl transferase family 2 [Bacillus anthracis]
Sequence ID: PES82498.1 Length: 1115 Number of Matches: 1
Range 1: 337 to 346GenPeptGraphics

Score	Expect	Identities	Positives	Gaps
27.4 bits(57)	402	7/10(70%)	7/10(70%)	0/10(0%)

```
Query  1    MAMTHKSQMK  10
            MAM H S MK
Sbjct  337  MAMAHNSDMK  346
```

Glycosyl transferase family 2

Glycoside hydrolase family 2 is a family of **glycoside hydrolases**. **Glycoside hydrolases** are a widespread group of enzymes that hydrolyse the glycosidic bond between two or more carbohydrates, or between a carbohydrate and a non-carbohydrate moiety. **Glycoside hydrolase family 2** comprises enzymes with several known activities: **beta-galactosidase**; **beta-mannosidase**; **beta-glucuronidase** (https://en.wikipedia.org/wiki/Glycoside_hydrolase_family_2). These enzymes contain a conserved glutamic acid residue which has been shown to be the general acid/base catalyst in the active site of the enzyme. The catalytic domain of Beta-galactosidases have a **TIM barrel core** surrounded several other largely beta domains. The sugar binding domain of these proteins has a jelly-roll fold. These enzymes also include an **immunoglobulin-like beta-sandwich domain** (https://en.wikipedia.org/wiki/Glycoside_hydrolase_family_2).
The **Region 108..408** of the **Sequence ID: PES82498.1** (the sequence identified above) correspondes with the **GH18 (glycosyl hydrolase, family 18) type II**. *These chitinases* hydrolyze **chitin**, an abundant *polymer of beta-1,4-linked N-acetylglucosamine (GlcNAc)* which is a major component of the cell wall of fungi and the exoskeleton of arthropods. Chitinases have been identified in *viruses, bacteria, fungi, protozoan parasites, insects, and plants*. The structure of the *GH18 domain* is an *eight-stranded* ***beta/alpha barrel*** *with a pronounced active-site cleft at the C-terminal end of the beta-barrel*. The *GH18 family* includes *chitotriosidase, chitobiase, hevamine, zymocin-alpha, narbonin, SI-CLP* (stabilin-1 interacting chitinase-like protein), *IDGF* (imaginal disc growth factor), *CFLE* (cortical fragment-lytic enzyme) *spore hydrolase, the type III and type V plant chitinases*, the *endo-beta-N-acetylglucosaminidases*, and the *chitolectins*. The **GH85** (glycosyl hydrolase, family 85) **ENGases** (endo-beta-N-acetylglucosaminidases) are closely related to the **GH18 chitinases** and are included in this alignment model (https://www.ncbi.nlm.nih.gov/Structure/cdd/cddsrv.cgi?uid=324582).

glycosyltransferase [Bacillus cereus]
Sequence ID: WP_098905904.1 Length: 1115 Number of Matches: 1

See 1 more title(s) See all Identical Proteins(IPG)

Related Information
Identical Proteins-Identical proteins to WP_098905904.1
Range 1: 337 to 346 GenPept Graphics

Score	Expect	Identities	Positives	Gaps
27.4 bits(57)	402	7/10(70%)	7/10(70%)	0/10(0%)

```
Query  1    MAMTHKSQMK  10
            MAM H S MK
Sbjct  337  MAMAHNSDMK  346
```

Glycosyltransferase

See the discussion above and on pages 233 and 251.

Nuclear GTPase SLIP-GC, partial [Lachnellula willkommii]
Sequence ID: TVY91158.1 Length: 1050 Number of Matches: 1
Range 1: 909 to 915 GenPept Graphics

Score	Expect	Identities	Positives	Gaps
26.5 bits(55)	811	7/7(100%)	7/7(100%)	0/7(0%)

```
Query  4    THKSQMK  10
            THKSQMK
Sbjct  909  THKSQMK  915
```

GTPase Nucleare SLIP-GC (Speckled-like Pattern in the Germinal Center)

A group of researchers (K. Richter, S. Brar, M. Ray, P. Pisitkun, S. Bolland, L.Verkoczy, and M. Diaz; DOI: 10.1074 / jbc.M109.014506) have "….identified a **novel GTPase**, **SLIP-GC**, with expression limited to a **few tissues**, in particular **germinal center B cells**. It lacks homology to any known proteins, indicating that it may belong to a novel family of GTPases. **SLIP-GC** is expressed in **germinal center B cells** and in **lymphomas derived from germinal center B cells such as large diffuse B cell lymphomas**. In cell lines, *SLIP-GC* is expressed in *lymphomas* that express **activation-induced deaminase** (**AID**) and that likely undergo ***somatic hypermutation***. SLIP-GC is a **nuclear protein**, and it localizes to **replication factories**. ***Reduction of SLIP-GC levels*** in the Burkitt *lymphoma cell line Raji* and in *non-Hodgkin lymphoma* cell lines resulted in an **increase in DNA breaks** and **apoptosis that was AID-dependent**, as simultaneous reduction of AID abrogated the deleterious effects of ***SLIP-GC reduction***. These results strongly suggest that SLIP-GC is a replication-related protein in germinal center B cells whose reduction is toxic to cells through an AID-dependent mechanism".

serine hydrolase [Novosphingobium sp. ST904]
Sequence ID: WP_132469210.1 Length: 473 Number of Matches: 1

See 1 more title(s) See all Identical Proteins(IPG)

Related Information
Identical Proteins-Identical proteins to WP_132469210.1
Range 1: 419 to 427 GenPept Graphics

Score	Expect	Identities	Positives	Gaps
26.5 bits(55)	814	7/9(78%)	7/9(77%)	0/9(0%)

```
Query  1    MAMTHKSQM  9
            MAM H SQM
Sbjct  419  MAMGHNSQM  427
```

Serine hydrolase

Serine hydrolases are one of the largest known enzyme classes comprising approximately ~200 enzymes or 1% of the genes in the human proteome. A defining characteristic of these enzymes is the presence of a **nucleophilic serine** in their *active site*, which is used for the **hydrolysis of substrates**. *Catalysis* proceeds by the formation of an acyl-enzyme intermediate through this **serine**, followed by water/hydroxide-induced saponification of the intermediate and regeneration of the enzyme. Unlike other *non-catalytic serines*, the **nucleophilic serine** of *these hydrolases* is typically activated by a **proton relay** involving a **catalytic triad** consisting of the **serine**, an acidic residue (e.g. *aspartate* or *glutamate*) and a **basic residue** (usually *histidine*), although variations on this mechanism exist. Superfamilies of **serine hydrolases** includes: *Serine proteases, Extracellular lipases, Intracellular lipases, Cholinesterases, Small molecule thioesterases, Some phospholipases, Protein and glycan hydrolases, Some amidases, Some peptidases* (https://en.wikipedia.org/wiki/Serine_hydrolase).

Among the sequences identified through BLAST research and reported in this paragraph, I recall ***Glycosyl-phosphatidylinositol*** (**GPI**) ***membrane anchors***, which are primarily responsible for the anchoring of cell-surface proteins in the plasma membrane and perform functions in *intracellular sorting*, *transmembrane signalling*, and *novel endocytic processes*.

A role in *signal transduction* (and *chemotaxis*) is also played by the ***bacteriohemerythrin*** sequence.

The same can be said for **Histidine kinase**, an enzyme involved in the *signal transduction pathways* upstream of many cellular processes including various metabolic, virulence, and homeostatic pathways.

Significant alignments with the **family 65 glycosyl hydrolase, Glycosyl transferase family 2** [more precisely, *GH18 (glycosyl hydrolase, family 18) type II*], **serine hydrolase** (very important hydrolase and which will be considered in the following chapters) and with the ***LmbE-like protein*** (**metallohydrolase** involving **zinc**) propose the importance of the **hydrolase** families (already discussed in the previous paragraphs).

The significant alignment with a ***LmbE-like protein*** re-proposes the importance of **hydrolase** (already discussed in the previous paragraphs), more precisely of **metallohydrolase** and the role of **zinc**.

The **Zn-finger** domain is also present in the **TM1802** family sequence (*CRISPR* and *associated Cas proteins*), family involved in the host defense against *phages* or other *foreign DNA*.

A *metal binding site* is also present in the *N-terminal beta barrel of pyrin* [found in significant alignment with the sequence of a ***protein of the pyrine family (Pirin_C)***], the presence of which may be significant in interaction with *Bcl-3* and *Nuclear Factor I* (NFI) and role in regulating the activity of the *transcription factor NF-kappaB* (*eukaryotic pyrins* are highly conserved nuclear proteins that can function as transcriptional regulators with a role in **apoptosis**).

Apoptosis that is **AID-dependent** (*Activation-Induced Deaminase*) and **DNA breaks** are involved in the ***Nuclear GTPase SLIP-GC*** (the latter protein related to *replication in the B cells of the germinal center*).

The product of another sequence is instead "***2-methylaconitate cis-trans isomerase PrpF***". In this case, it is the role of ***dehydratase*** that gains importance, as *PrpF is an essential accessory protein for the 2-methylisocitrate* ***dehydratase*** *AcnD Fe / S-dependent*.

The alignment with a sequence of ***MlotiK1 channel***, involved in **signaling networks of vision** and **olfaction**, instead, it emphasizes the importance of the **olfactory system**.

Finally, very important is the identification of a sequence that refers to the ***N-terminal domain of DNA-directed RNA polymerase*** (domain that undergoes a structural change that favors the passage of the *RNA transcript*) which is particularly interesting because the sequence of amino acids **M A M T H K S Q M K** is correlated, through a segment of the **Sequence n° 1/3**, to the **original sequence**, the known sequence (**GTGAAATGGTCATGTGTGGCGG**) of the **RdRp gene** [or ***RNA-dependent RNA polymerase***] of **SARS-CoV-2** (the one being studied in this document).

Before concluding, I would like to draw attention again to the significant alignment of the sequences (***which are very numerous***, although I have reported only one as an example for all) whose product is the ***Glycosyl transferase family 2***, because it refers to **Bacillus anthracis**. This **bacterium**, which will also be identified from other sequences obtained from the modifications of the **original sequence**, due to its **lethality** and its massive involvement in **lung injury**, will be the subject of extensive discussion in the following chapters.

9.12 SEQUENCE OF AMINO ACIDS "MAMTHKSQMKPRQMAP"

This paragraph shows some results of the BLAST research carried out on the sequence of amino acids **MAMTHKSQMKPRQMAP**, the sequence already considered in the ***previous paragraph*** and to which the amino acids **Proline** (**P**), **Arginine** (**R**), **Glutamine** (**Q**), **Methionine** (**M**), **Alanine** (**A**), and **Proline** (**P**), determined by the ***codons*** **CCT**, **CGT**, **CAA**, **ATG**, **GCA**, and **CCA** respectively.

QUERY: MAMTHKSQMKPRQMAP

(atg gca atg aca cac aag agt caa atg aaa cct cgt caa atg gca cca taa)

Search Parameters	
Program	Blastp
Word size	2
Expect value	200000
Hitlist size	1000
Gapcosts	9,1
Matrix	PAM30
Filter string	F
Genetic Code	1
Window Size	40
Threshold	11
Composition-based stats	0

1000 sequences selected:

Accession Seq	Description	Comomm Name	Max Score	Total Score	Query C.	E value	Per. Id.	Acc. Len
.....	.....	.....	.....	.....	.....	....	.....	.....
CAD7463592.1	unnamed protein product [Timema tahoe]	Timema tahoe	32.9	65.8	68%	14	81.82%	385
.....	.....	.....	.....	.....	.....	....	.....	.....
WP_011927984.1	FAD binding domain-containing protein [Arthrobacter sp. Rue61a]	Arthrobacter sp. Rue61a	29.9	29.9	56%	155	88.89%	290
.....	.....	.....	.....	.....	.....	....	.....	.....
WP_038204185.1	coniferyl aldehyde dehydrogenase [Vibrio tubiashii]	Vibrio tubiashii	29.9	29.9	56%	155	88.89%	478
.....	.....	.....	.....	.....	.....	....	.....	.....
XP_032397351.1	desmoplakin-like [Etheostoma spectabile]	Etheostoma spectabile	29.5	29.5	56%	220	88.89%	1675
.....	.....	.....	.....	.....	.....	....	.....	.....
KAB5551371.1	helix-loop-helix DNA-binding domain-containing protein [Coniochaeta sp. 2T2.1]	Coniochaeta sp. 2T2.1	28.6	28.6	56%	440	88.89%	395
.....	.....	.....	.....	.....	.....	....	.....	.....
WP_105377509.1	glycoside hydrolase family 3 C-terminal domain-containing protein [Massilia phosphatilytica]	Massilia phosphatilytica	28.6	28.6	50%	440	87.50%	462
.....	.....	.....	.....	.....	.....	....	.....	.....
NXP71248.1	MRPP3 ribonuclease [Ramphastos sulfuratus]	Ramphastos sulfuratus	28.6	28.6	50%	440	87.50%	468

unnamed protein product, partial [Timema tahoe]

GenBank: CAD7463592.1

Identical Proteins FASTA Graphics

Go to:

LOCUS CAD7463592 385 aa linear INV **11-DEC-2020**

DEFINITION unnamed protein product, partial [Timema tahoe].

ACCESSION CAD7463592

VERSION CAD7463592.1

DBLINK BioProject: PRJEB31411

DBSOURCE embl accession OE008534.1

KEYWORDS .

SOURCE Timema tahoe

ORGANISM Timema tahoe

Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Polyneoptera; Phasmatodea; Timematodea; Timematoidea; Timematidae; Timema.

REFERENCE 1

AUTHORS Tran Van,P.

TITLE Direct Submission

JOURNAL **Submitted (22-NOV-2020) UNIL, DEE, 1015 Lausanne, Switzerland**

FEATURES Location/Qualifiers

source 1..385
/organism="**Timema tahoe**"
/db_xref="taxon:61484"

Protein 1..>385
/name="**unnamed protein product**"

CDS 1..385
/locus tag="TTEB3V08 LOCUS11474"
/coded_by="complement(join(OE008534.1:<3806..4288,
OE008534.1:4845..5327,OE008534.1:6878..7006,
OE008534.1:7853..7912))"
/note="source:maker;
ID:TTE 10322-RA:cds"

ORIGIN

```
        1 mldqdgmfea lalrrkkstv lesqpmdiec vsllhdihss nisnynyrgy tilgqndkyr
       61 eikmyhalgp krslallmwh altgcdptls frgrskkmaf nawkknpdsv teailqimek
      121 pfmeidmgsy elkhldsymi ilysatvgdk inglrrkify hkyqnpeimp ptqdcdlqhc
      181 kravfqasiw amshkskmkp rnpskygwnv kenilfelsy kleemyhalg pkrslallmw
      241 haltgcdptl sfrgrskkma fnawkknpds vteailqime kpfmeidmgs yelkhldsym
      301 iilysatvgd kinglrrkif yhkyqnpeim pptqdcdlqh ckravfqasi wamshkskmk
      361 prnpskygwn vkenilfels yklee
```

//

unnamed protein product, partial [Timema tahoe]
Sequence ID: CAD7463592.1 Length: 385 Number of Matches: 2
Range 1: 191 to 201 GenPept Graphics

Score	Expect	Identities	Positives	Gaps
32.9 bits(70)	13	9/11(82%)	9/11(81%)	0/11(0%)

```
Query  2    AMTHKSQMKPR  12
            AM HKS MKPR
Sbjct  191  AMSHKSKMKPR  201
```

Range 2: 352 to 362 GenPept Graphics

Score	Expect	Identities	Positives	Gaps
32.9 bits(70)	13	9/11(82%)	9/11(81%)	0/11(0%)

```
Query  2    AMTHKSQMKPR  12
            AM HKS MKPR
Sbjct  352  AMSHKSKMKPR  362
```

FAD binding domain-containing protein [Arthrobacter sp. Rue61a]
Sequence ID: WP_011927984.1 Length: 290 Number of Matches: 1

See 2 more title(s) See all Identical Proteins(IPG)

Related Information
Identical Proteins-Identical proteins to WP_011927984.1
Range 1: 38 to 46 GenPep tGraphics

Score	Expect	Identities	Positives	Gaps
29.9 bits(63)	153	8/9(89%)	8/9(88%)	0/9(0%)

```
Query  8   QMKPRQMAP  16
           QMKPRQ AP
Sbjct  38  QMKPRQVAP  46
```

FAD binding domain-containing protein (CO dehydrogenase flavoprotein C-terminal domain)

Proteins containing this domain form structural complexes with other *known families*, such as ***Aldehyde oxidase/xanthine dehydrogenase*** (IPR008274), and ***2Fe-2S ferredoxin-type iron-sulfur binding domain*** (IPR001041). The *carbon monoxide* (CO) *dehydrogenase* of Oligotropha carboxidovorans is a heterotrimeric complex composed of a **apoflavoprotein**, a *molybdoprotein*, and an **iron-sulphur protein**. It can be dissociated with *sodium dodecylsulphate*. CO dehydrogenase catalyzes the oxidation of CO according to the following equation: CO + H2O = CO2 + 2e + 2H+. **Subunit S** represents the *iron-sulphur protein of CO dehydrogenase* and is clearly divided into a **C-** and an **N-terminal domain**, each binding a [2Fe-2S] cluster. The structure of the **C-terminal domain** has a **beta-alpha** fold arranged in an **alpha+beta sandwich** (http://www.ebi.ac.uk/interpro/entry/InterPro/IPR036683/).

coniferyl aldehyde dehydrogenase [Vibrio tubiashii]
Sequence ID: WP_038204185.1 Length: 478 Number of Matches: 1

See 1 more title(s) See all Identical Proteins(IPG)

Related Information
Identical Proteins-Identical proteins to WP_038204185.1
Range 1: 330 to 338 GenPept Graphics

Score	Expect	Identities	Positives	Gaps
29.9 bits(63)	153	8/9(89%)	8/9(88%)	0/9(0%)

```
Query  8    QMKPRQMAP  16
            Q KPRQMAP
Sbjct  330  QVKPRQMAP  338
```

Coniferyl aldehyde dehydrogenase

A **Coniferyl-aldehyde dehydrogenase** is an **enzyme** that catalyzes a chemical reaction. The **4 substrates** of this enzyme are **coniferyl aldehyde**, **H2O**, **NAD+**, and **NADP+**, whereas its **4 products** are **ferulate**, **NADH**, **NADPH**, and **H+**. This enzyme belongs to the family of *oxidoreductases*, specifically those acting on the *aldehyde* or *oxo* group of donor with *NAD+ or NADP+ as acceptor*. The systematic name of this enzyme class is **coniferyl aldehyde:NAD(P)+oxidoreductase** (https://en.wikipedia.org/wiki/Coniferyl-aldehyde_dehydrogenase).

desmoplakin-like [Etheostoma spectabile]
Sequence ID: XP_032397351.1 Length: 1675 Number of Matches: 1
Related Information
Gene-associated gene details
Genome Data Viewer-aligned genomic context
Range 1: 1041 to 1049 GenPept Graphics

Score	Expect	Identities	Positives	Gaps
29.5 bits(62)	217	8/9(89%)	8/9(88%)	0/9(0%)

```
Query  3     MTHKSQMKP  11
             MTHKSQ KP
Sbjct  1041  MTHKSQIKP  1049
```

Desmoplakin

Desmoplakin is a **protein** in humans that is encoded by the *DSP* gene. Desmoplakin is a critical component of *desmosome structures* in *cardiac muscle* and *epidermal cells*, which ***function to maintain the structural integrity at adjacent cell contacts***. In **cardiac muscle**, desmoplakin is localized to *intercalated discs* which mechanically couple cardiac cells to function in a coordinated syncytial structure. **Mutations in desmoplakin** have been shown to play a role in *dilated cardiomyopathy*, *arrhythmogenic right ventricular cardiomyopathy*, *striate palmoplantar keratoderma*, *Carvajal syndrome* and *paraneoplastic pemphigus* (https://en.wikipedia.org/wiki/Desmoplakin).

helix-loop-helix DNA-binding domain-containing protein [Coniochaeta sp. 2T2.1]
Sequence ID: KAB5551371.1Length: 395Number of Matches: 1
Range 1: 104 to 112 GenPept Graphics

Score	Expect	Identities	Positives	Gaps
28.6 bits(60)	435	8/9(89%)	8/9(88%)	0/9(0%)

```
Query  8    QMKPRQMAP  16
            QMKPRQ AP
Sbjct  104  QMKPRQQAP  112
```

Myc-type, basic helix-loop-helix (bHLH) domain

A number of **eukaryotic proteins**, which probably are **sequence specific DNA-binding proteins that act as transcription factors**, share a conserved domain of 40 to 50 amino acid residues. It has been proposed that this domain is formed of two amphipathic helices joined by a variable length linker region that could form a loop. This "**helix-loop-helix**" (**HLH**) domain **mediates protein dimerization**. Most of these proteins (see this websit: https://www.ebi.ac.uk/interpro/entry/InterPro/IPR011598/) have an *extra basic* region of about 15 amino acid residues that is *adjacent to the HLH domain and specifically binds to DNA. They are referred as basic helix-loop-helix proteins (bHLH), and are classified in two groups: class A (ubiquitous) and class B (tissue-specific).* Members of the bHLH family bind variations on the core sequence "CANNTG", also referred to as the **E-box motif** (https://www.ebi.ac.uk/interpro/entry/InterPro/IPR011598/).
. The homo- or heterodimerization mediated by the HLH domain is independent of, but necessary for DNA binding, as two basic regions are required for DNA binding activity. The HLH proteins lacking the basic domain (Emc, Id) function as **negative regulators**, since they form heterodimers, ***but fail to bind DNA***. The hairy-related proteins (hairy, E(spl), deadpan) also repress transcription although they can bind DNA. The proteins of this subfamily act together with *co-repressor proteins*, like *groucho*, through their **-terminal motif WRPW** (https://www.ebi.ac.uk/interpro/entry/InterPro/IPR011598/).

glycoside hydrolase family 3 C-terminal domain-containing protein [Massilia phosphatilytica]
Sequence ID: WP_105377509.1 Length: 462 Number of Matches: 1

See 1 more title(s) See all Identical Proteins(IPG)

Related Information
Identical Proteins-Identical proteins to WP_105377509.1
Range 1: 455 to 462 GenPept Graphics

Score	Expect	Identities	Positives	Gaps
28.6 bits(60)	435	7/8(88%)	8/8(100%)	0/8(0%)

```
Query  9    MKPRQMAP  16
            MKPR+MAP
Sbjct  455  MKPREMAP  462
```

Glycoside hydrolase family 3 C-terminal domain-containing protein

Glycoside hydrolase family 3 is a family of **glycoside hydrolases** (See what has been discussed about hydrolases in the previous paragraphs and link http://pfam.xfam.org/family/PF01915.22). Glycoside hydrolase family 3 (CAZY GH 3) comprises enzymes with a number of known activities; *beta-glucosidase; beta-xylosidase; N-acetyl beta-glucosaminidase; glucan beta-1,3-glucosidase; cellodextrinase; exo-1,3-1,4-glucanase*. These enzymes are *two-domain globular proteins* that are *N-glycosylated at three sites* (http://pfam.xfam.org/family/PF00933; for important information, please also consult the link https://www.ncbi.nlm.nih.gov/genome/annotation_prok/evidence/NF014026/).

MRPP3 ribonuclease [Ramphastos sulfuratus]
Sequence ID: NXP71248.1 Length: 468 Number of Matches: 1
Range 1: 299 to 306 GenPept Graphics

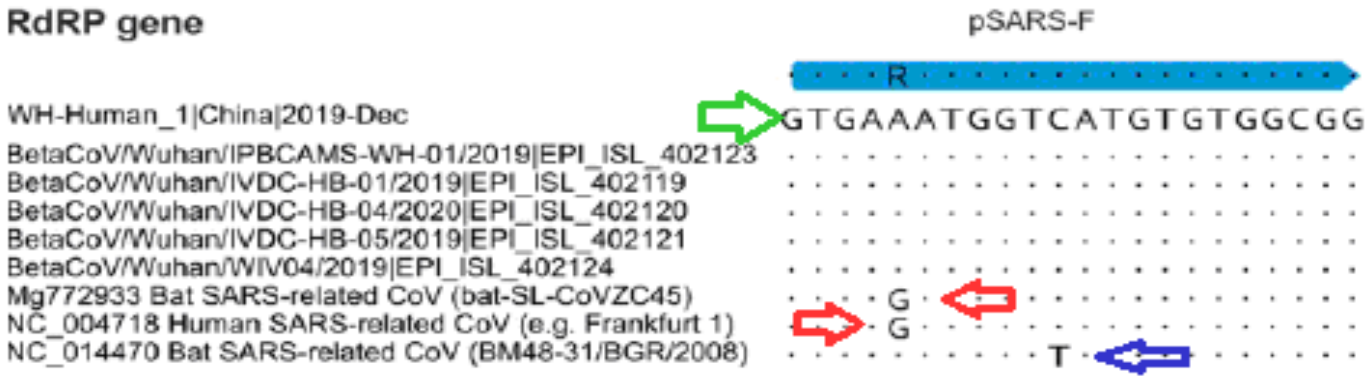
Score	Expect	Identities	Positives	Gaps
28.6 bits(60)	435	7/8(88%)	8/8(100%)	0/8(0%)

```
Query  7    SQMKPRQM  14
            S+MKPRQM
Sbjct  299  SHMKPRQM  306
```

MRPP3 ribonuclease

PRORPs (protein-only RNase P) are a class of *RNA processing enzymes* that catalyze maturation of the 5' end of precursor tRNAs in Eukaryotes. Arabidopsis thaliana contains PRORP enzymes (PRORP1, PRORP2 and PRORP3) where PRORP1 localizes to mitochondria as well as chloroplasts, while PRORP2 and PRORP3 are found in the nucleus. In humans and most other metazoans, mt-RNase P is composed of **three protein subunits** (*mitochondrial RNase P proteins 1-3;* **MRPP1-3**), homologs to the Arabidopsis thaliana PRORP1-3. This domain corresponds to the *metallonuclease* domain of PRORPs. PRORP1 has 22% sequence identity to the human homologue **MRPP3**. PRORP1 crystal structure shows a V-shaped tripartite structure with a C-terminal metallonuclease domain of the NYN (N4BL1, YacP-like nuclease) family, with a typical and functional two-metal-ion catalytic site that has conserved aspartate residues (https://pfam.xfam.org/family/PRORP#tabview=tab0). *L. Reinhard, S. Sridhara* and *B. M. Hällberg* explain us that "**Mitochondrial RNA polymerase** produces long polycistronic precursors that contain the *mRNAs, rRNAs and tRNAs needed for mitochondrial translation*. Mitochondrial RNase P (mt-RNase P) initiates the maturation of the precursors by cleaving at the 5' ends of the tRNAs. Human mt-RNase P is only active as a tripartite complex (mitochondrial RNase P proteins 1-3; MRPP1-3), whereas plant and trypanosomal RNase Ps (PRORPs)-albeit homologous to MRPP3-are active as single proteins. The reason for this discrepancy has so far remained obscure…**the crystal structure of human MRPP3, which features a remarkably distorted and hence non-productive….(switch)….to a fully productive state only upon association with MRPP1, MRPP2 and pre-tRNA substrate. We suggest a mechanism in which MRPP1 and MRPP2 both deliver the pre-tRNA substrate and activate MRPP3 through an induced-fit process**" (https://pubmed.ncbi.nlm.nih.gov/25953853/).

The **first two significant alignment** identified by the BLAST research on the sequence "**MAMTHKSQMKPRQMAP**" (sequence object of further study in this paragraph) is an "***unnamed protein product, partial***" of the **Timema tahoe** insect. **Timema** is a genus of relatively short-bodied, stout stick insects native to the far *western United States*, and the sole extant member of the *family Timematidae* (https://en.wikipedia.org/wiki/Timema). According to Tanja Schwander, "*Timema* are indeed the *oldest insects* for which there is good evidence that they have been asexual for long periods of time. Tanja Schwander heads a team of researchers who found that five *Timema* species (*T. douglasi*, *T. monikense*, *T. shepardi*, ***T. tahoe*** and *T. genevievae*) have used only asexual reproduction for more than 500,000 years, with ***T. tahoe*** and *T. genevievae reproducing asexually for over one million years* (https://en.wikipedia.org/wiki/Timema). For the purposes of our research, the importance of this alignment is justified by the fact that (as announced in Chapter II and then shown in Chapter V) ***Timema tahoe*** is among the very few "organisms" identified by the BLAST research on the sequence on the "**GTGAGATGGTCATGTGTGGCGG**". This sequence is important because it is almost totally identical to the "**GTGAAATGGTCATGTGTGGCGG**" sequence, the RNA sequence studied in this document. Furthermore, as already seen in Chapter III in Fig. 3.1 (of which I report the relevant segment below), the sequence "**GTGAGATGGTCATGTGTGGCGG**" (in the segment, the sequence indicated by the red arrow) it is one of the sequences of the **RdRp gene** (*RNA-dependent RNA polymerase*, the motor of the **transcription** process of *SARS-CoV-2*) whose significant alignments of BLAST researches have identified variants of the **Bat** species [***Rhinolophus Sinicus*** (called *Horseshoe*), ***Rhinolophus Pusillus*** e ***Rhinolophus Monoceros***].



The presence of **Timema tahoe** insect among the significant alignments (shown in this and subsequent chapters) certainly cannot be considered random, but determined by the action of the "*Non-Obvious Trend*" of **Sequence n° 1/3** and other **modified sequences** (which will be presented in the following chapters). Furthermore, it is important to underline that the **prediction of the constitutive codons of the segment** of **Sequence n° 1/3** (and the next segment of nitrogenous bases identified – see page 257) corresponds to the **hypothetical** amino acids sequence "**MAMTHKSQMKPRQMAP**" determined by the significant alignments of a sequence ***HIV-1***, the sequence that is object of study in this chapter.

The significant alignment with a sequence of the "***AD binding domain-containing protein (CO dehydrogenase flavoprotein C-terminal domain)***" insistently re-proposes the role of ***dehydrogenases*** and ***metal bonds*** (in this case, **iron-sulfur bonds**).

The role of ***dehydrogenases*** is reaffirmed by the significant alignment with a sequence whose *product* is the "**Coniferyl aldehyde dehydrogenase**".

The product "***Glycoside hydrolase family 3 C-terminal domain-containing protein***" of the significant alignment of another sequence, on the other hand, emphasizes the role of **hydrolase**.

Less clear is the alignment with a sequence whose *product* is a ***Desmoplakin-like protein*** (a critical component of *desmosome structures* in *cardiac muscle* and *epidermal cells*), a protein that works ***to maintain the structural integrity at adjacent cell contacts***. **Mutations in desmoplakin** have been shown to play a role in *paraneoplastic pemphigus* (https://en.wikipedia.org/wiki/Desmoplakin). ***Paraneoplastic pemphigus*** (**PNP**) is an *autoimmune disorder* stemming from an *underlying tumor*. It is hypothesized that *antigens associated with the tumor* trigger an *immune response* resulting in *blistering of the skin* and *mucous membranes*. Roughly 1/3 of the **deaths from PNP** stem from **pulmonary insufficiency which is brought about by the action of PNP on the respiratory mucosa**. It manifests as *dyspnea* and progresses to *bronchiolitis obliterans* (non-reversible obstructive lung disease) via an unknown mechanism (https://en.wikipedia.org/wiki/Paraneoplastic_pemphigus).

The significant alignment relating to the "**Myc-type, basic helix-loop-helix (bHLH) domain**" concerns *DNA-binding proteins* (proteins already discussed in the previous paragraphs) *specific to the sequence that act as* **transcription factors**.

I conclude this paragraph and the entire chapter, underlining the identification of a sequence that refers to the "**MRPP3 ribonuclease**" (of the class of *mitochondrial RNase P proteins 1-3;* **MRPP1-3**), which takes into consideration the importance of **mitochondrial RNA polymerase processing enzymes**.

CHAPTER X

BLAST RESEARCH AND DISCUSSION OF THE RESULTS CARRIED OUT ON ALIGNMENTS OF SEQUENCE N ° 3/1

GOALS OF CHAPTER X

This chapter contains the results of the *significant alignments* of BLAST research carried out on **Sequence n° 3/1**. This sequence is one of many generated by the **Original Sequence** [the sequence "**GTGAAATGGTCATGTGTGGCGG**" of the RdRp Gene (*RNA-dependent RNA polymerase, RdRp_SARSr-F*)] and from which it differs by **81,82%** of the bases.
The results of the significant alignments of the BLAST research (**Hitlist size 100**) covered only one **common "organisms"** (**Chanos chanos**) among the **"organisms"** identified by BLAST research carried out on sequences to which **a base of the original sequence has been replaced** (see Chapters V and VI).
Extending the BLAST research (**Hitlist size 5000**), many **common "organisms"** have been identified (distinguishable in different colours in the following pages) and, in particular, **Manis pentadactyla** (*Chinese Pangolin*).
The same BLAST research (**Hitlist size 5000**) identified six sequences of **Pipistrellus pipistrellus**, four sequences of the **Influenza A virus** and, above all, a considerable number of **Paranthracis and Anthracis Bacilli sequences**.
The latter two "**organisms**" were certainly not expected among expected results, even less than the **HIV-1 Virus** identified through the BLAST research carried out on **Sequence n° 1/3** (see what discussed in the chapter previous one). However, as shown in the previous chapter, the BLAST research carried out on the sequence of amino acids **M A M T H K S Q M K** [determined starting from **Sequence n ° 1/3** (for further information see the previous chapter)] has strangely identified numerous sequences of **Bacillus Anthracis** [whose "products" correspond to *GH18 (glycosyl hydrolase, family 18) type II*].

Bacillus Anthracis, which will also be identified from other sequences obtained from modifications of the **original sequence**, due to its **lethality** and its massive involvement in **lung damage**, will be the subject of extensive discussion in the following chapters.
Paranthracis and Anthracis Bacilli are part of the **Bacillus Cereus group** (http://lifemap-ncbi.univ-lyon1.fr/?tid=2026186), that is a group that also includes foodborne pathogen ***B. Cereus***, and biopesticide ***B. Thuringiensisn*** (*https://en.wikipedia.org/wiki/Bacillus_anthracis*). The BLAST research carried out on **Sequence n ° 3/1** has identified numerous sequences of all these **Bacilli** (https://mbio.asm.org/content/11/1/e00034-20). "The ***Bacillus Cereus group*** comprises 21 closely related species of environmentally **ubiquitous spore-forming** *Gram-positive bacteria*. Some of them are **opportunistic pathogens**. Due to their diverse lifestyles and differences in gene content, some strains of the ***B. Cereus*** group are considered to be **probiotics**" (https://mra.asm.org/content/8/44/e00662-19).

As already discussed in the **Preamble** of this document (and to which I refer for other important information), ***Bacillus Anthracis*** is an infectious gram-positive bacterium, the causative agent of *anthrax in animals* and *anthrax in humans* (https://en.wikipedia.org/wiki/Bacillus_anthracis).

A*nthrax in animals* is a disease of animals, especially herbivores, caused by the ingestion or inhalation of live or spore-bearing bacilli. Affected animals have a high fever and quickly go to die, presenting a blackish spleen, fragile and very rich in bacilli, and dark-looking blood at the post-mortem examination (https://en.wikipedia.org/wiki/Bacillus_anthracis).
A*nthrax* is the human equivalent of **A*nthrax in animals***. Human-animal infection is difficult, and only patients who are in prolonged contact with the host animals of the bacterium or with their products are exposed to it, so it can be considered a typical zoonosis.
Inter-human contagion is impossible (https://it.wikipedia.org/wiki/Bacillus_anthracis).

"**Anthrax** infection occurs in *three forms*:
- the ***cutaneous form***, by far the most common, with a mortality of 20% in untreated cases. It manifests itself with a typical skin lesion. The disease becomes fatal when the bacilli break through the skin barrier and spread into the blood, causing severe septicemiala;
- the ***gastrointestinal form***, rare, contracted through the ingestion of undercooked meat from infected animals. It manifests itself with general gastrointestinal symptoms (nausea, loss of appetite, bloody diarrhea), followed by high fever and intense abdominal pain;
- the ***pulmonary form*** (very rare and present only in risk categories), contracted by inhalation, is **considerably more serious** and has been associated with the widespread fear of a biological warfare (https://it.wikipedia.org/wiki/Bacillus_anthracis)".

The *pulmonary form* of **anthrax infection begins with the typical symptoms of a flu (sore throat, fever, myalgia); followed by cough, chest pain and difficulty in breathing. This form is, from an anatomical-pathological point of view, hemorrhagic lymphadenitis-**mediastinitis**, rather than actual pneumonia, and does not involve the presence of bacilli in the sputum** (https://db0nus869y26v.cloudfront.net/it/Bacillus_anthracis).
Symptoms in all three forms can occur **within 7 days of contact with the bacterium,** with the **exception of the pulmonary form which can present them after 7-42 days**.
Anthrax infection is not easily identifiable, because "the symptoms of the disease (except for the more specific cutaneous form) hardly allow a sure diagnosis. A more precise method is by culturing the bacterium in the blood (https://it.wikipedia.org/wiki/Bacillus_anthracis).
The **anthrax vaccine in animals** created by *Pasteur* has greatly reduced the danger of infections. The **human anthrax** problem is more complex. The existing **vaccine** against the disease is not available to the public, due to the **presence of side effects**; instead, it is intended to be used for those who can potentially come into contact with the bacterium, including **some US armed forces** and the **staff of laboratories where the bacilli are present**.
Anthrax is considered to be in category A (i.e. of highest priority) of infectious agents associated with a potential risk of use as a **biological weapon** (https://it.wikipedia.org/wiki/Bacillus_anthracis).

SEQUENCE N° 3/1

(T=2; C=0; G=3; A=1)

10.1 SEQUENCE N° 3/1 (T=2; C=0; G=3; A=1) – (Hitlist size: 100)

Original Sequence	**G**	**T**	**G**	**A**	**A**	**A**	**T**	**G**	**G**	**T**	**C**	**A**	**T**	**G**	**T**	**G**	**T**	**G**	**G**	**C**	**G**	**G**
Trend n° 3 (T=2; C=0; G=3; A=1)	**3**	**2**	**3**	**1**	**1**	**1**	**2**	**3**	**3**	**2**	**0**	**1**	**2**	**3**	**2**	**3**	**2**	**3**	**3**	**0**	**3**	**3**
Trend n° 3 (I i D d)	**-**	**D**	**I**	**D**	**d**	**d**	**I**	**I**	**i**	**D**	**D**	**I**	**I**	**I**	**D**	**I**	**D**	**I**	**i**	**D**	**I**	**i**
Trend n° 3/1	**2**	**1**	**2**	**0**	**0**	**0**	**1**	**2**	**2**	**1**	**0**	**1**	**2**	**3**	**1**	**2**	**1**	**2**	**2**	**1**	**2**	**2**
Sequence n° 3/1 (81,82%)	**T**	**A**	**T**	**C**	**C**	**C**	**A**	**T**	**T**	**A**	**C**	**A**	**T**	**G**	**A**	**T**	**A**	**T**	**T**	**A**	**T**	**T**

Query: **TATCCCATTACATGATATTATT**

Search Parameters	
Program	Blastn
Word size	7
Expect value	1000
Hitlist size	100
Match/Mismatch scores	1,-3
Gapcosts	5,2
Filter string	F
Genetic Code	1

100 sequences selected: 100 sequences selected:

Accession Seq	Description	Comomm Name	Max Score	Total Score	Query C.	E value	Per. Id.	Acc. Len
XM_033872625.1	PREDICTED: Pecten maximus protein Wnt-10a-like (LOC117317723), transcript variant X4, mRNA	Pecten maximus	40.1	40.1	90%	0.56	100.00%	5199
XM_033872624.1	PREDICTED: Pecten maximus protein Wnt-10a-like (LOC117317723), transcript variant X3, mRNA	Pecten maximus	40.1	40.1	90%	0.56	100.00%	4618
XM_033872623.1	PREDICTED: Pecten maximus protein Wnt-10a-like (LOC117317723), transcript variant X2, mRNA	Pecten maximus	40.1	40.1	90%	0.56	100.00%	5064
XM_033872622.1	PREDICTED: Pecten maximus protein Wnt-10a-like (LOC117317723), transcript variant X1, mRNA	Pecten maximus	40.1	40.1	90%	0.56	100.00%	5205
LR736856.1	Pecten maximus genome assembly, chromosome: 19	Pecten maximus	40.1	195	90%	0.56	100.00%	32483354
CP054800.1	**Bacillus anthracis strain FDAARGOS_702 chromosome**	Bacillus anthracis	38.2	38.2	86%	2.2	100.00%	5272559
CP054797.1	**Bacillus anthracis strain FDAARGOS_703 chromosome**	Bacillus anthracis	38.2	38.2	86%	2.2	100.00%	5261520
CP054568.1	Bacillus thuringiensis strain FDAARGOS_791 chromosome, complete genome	Bacillus thuringiensis	38.2	38.2	86%	2.2	100.00%	5281841
CP050183.1	Bacillus thuringiensis strain HER1410 chromosome, complete genome	Bacillus thuringiensis	38.2	72.4	100%	2.2	100.00%	5585577
CP053938.1	Bacillus thuringiensis strain FDAARGOS_792 chromosome, complete genome	Bacillus thuringiensis	38.2	70.4	86%	2.2	100.00%	5251676
CP053934.1	Bacillus thuringiensis strain FDAARGOS_794 chromosome, complete genome	Bacillus thuringiensis	38.2	70.4	86%	2.2	100.00%	5214223
CP053931.1	Bacillus cereus strain FDAARGOS_797 chromosome, complete genome	Bacillus cereus	38.2	70.4	86%	2.2	100.00%	5413450
CP053997.1	Bacillus cereus strain FDAARGOS_780 chromosome, complete genome	Bacillus cereus	38.2	70.4	86%	2.2	100.00%	5271040
CP053991.1	Bacillus cereus strain FDAARGOS_781 chromosome, complete genome	Bacillus cereus	38.2	70.4	86%	2.2	100.00%	5271029
CP053980.1	Bacillus thuringiensis strain FDAARGOS_795 chromosome, complete genome	Bacillus thuringiensis	38.2	70.4	86%	2.2	100.00%	5228070
CP053954.1	Bacillus cereus strain	Bacillus cereus	38.2	38.2	86%	2.2	100.00%	5381410

CP053965.1	FDAARGOS_798 chromosome, complete genome	Bacillus cereus	38.2	38.2	86%	2.2	100.00%	5342923
CP053955.1	Bacillus tropicus strain FDAARGOS_802 chromosome, complete genome	Bacillus tropicus	38.2	38.2	86%	2.2	100.00%	5263142
CP053951.1	Bacillus cereus strain FDAARGOS_782 chromosome, complete genome	Bacillus cereus	38.2	38.2	86%	2.2	100.00%	5381556
CP053981.1	Bacillus thuringiensis strain FDAARGOS_799 chromosome, complete genome	Bacillus thuringiensis	38.2	38.2	86%	2.2	100.00%	5256259
CP053289.1	Bacillus cereus strain FDAARGOS_793 chromosome, complete genome	Bacillus cereus	38.2	72.4	100%	2.2	100.00%	5328644
CP049978.1 1/2 8/1	Bacillus cereus strain WPySW2 chromosome, complete genome	Bacillus sp. RZ2MS9	38.2	38.2	86%	2.2	100.00%	5357194
CP050973.1	**Bacillus anthracis strain FDAARGOS_699 chromosome**	Bacillus anthracis	38.2	38.2	86%	2.2	100.00%	5227345
CP050972.1	**Bacillus anthracis strain FDAARGOS_697 chromosome**	Bacillus anthracis	38.2	38.2	86%	2.2	100.00%	5240765
CP050971.1	**Bacillus anthracis strain FDAARGOS_696 chromosome**	Bacillus anthracis	38.2	38.2	86%	2.2	100.00%	5255974
CP050970.1	**Bacillus anthracis strain FDAARGOS_694 chromosome**	Bacillus anthracis	38.2	38.2	86%	2.2	100.00%	5251490
CP049019.1	Bacillus tropicus strain AOA-CPS1 chromosome	Bacillus tropicus	38.2	70.4	86%	2.2	100.00%	5246960
CP048687.1	**Bacillus paranthracis strain MN1F chromosome, complete genome**	Bacillus paranthracis	38.2	70.4	86%	2.2	100.00%	5687871
AP022643.1	Bacillus wiedmannii PL1 DNA, complete genome	Bacillus wiedmannii	38.2	38.2	86%	2.2	100.00%	5309441
CP028009.1	Bacillus cereus strain SGAir0260 chromosome	Bacillus cereus	38.2	38.2	86%	2.2	100.00%	5945656
CP027920.1	Bacillus cereus strain SGAir0263 chromosome, complete genome	Bacillus cereus	38.2	38.2	86%	2.2	100.00%	5947158
CP041981.1	**Bacillus paranthracis strain NCCP 15910 chromosome, complete genome**	Bacillus paranthracis	38.2	70.4	86%	2.2	100.00%	5208566
CP041750.1	**Bacillus paranthracis strain NCCP 14796 chromosome, complete genome**	Bacillus paranthracis	38.2	70.4	86%	2.2	100.00%	5017580
CP047131.1	**Bacillus anthracis str. BF1 chromosome, complete genome**	Bacillus anthracis str. BF1	38.2	38.2	86%	2.2	100.00%	5213532
CP047111.1	**Bacillus anthracis strain FDAARGOS_704 chromosome, complete genome**	Bacillus anthracis	38.2	38.2	86%	2.2	100.00%	5227450
CP047107.1	**Bacillus anthracis strain FDAARGOS_705 chromosome, complete genome**	Bacillus anthracis	38.2	38.2	86%	2.2	100.00%	5227452
CP047104.1	**Bacillus anthracis strain FDAARGOS_706 chromosome, complete genome**	Bacillus anthracis	38.2	38.2	86%	2.2	100.00%	5227454
CP047099.1	**Bacillus anthracis strain FDAARGOS_700 chromosome, complete**	Bacillus anthracis	38.2	38.2	86%	2.2	100.00%	5227419

	genome							
CP047098.1	**Bacillus anthracis strain FDAARGOS_701 chromosome, complete genome**	Bacillus anthracis	38.2	38.2	86%	2.2	100.00%	5227419
CP047097.1	**Bacillus anthracis strain FDAARGOS_698 chromosome, complete genome**	Bacillus anthracis	38.2	38.2	86%	2.2	100.00%	5225746
CP047085.1	**Bacillus paranthracis strain BC307 chromosome, complete genome**	Bacillus paranthracis	38.2	70.4	86%	2.2	100.00%	5248370
CP046398.1 **1/2 8/1**	Bacillus sp. A260 chromosome, complete genome	Bacillus sp. A260	38.2	72.4	100%	2.2	100.00%	5286434
CP040344.1	Bacillus albus strain DLOU-Yingkou chromosome, complete genome	Bacillus albus	38.2	38.2	86%	2.2	100.00%	5298275
CP040342.1	Bacillus cereus strain DLOU-Weihai chromosome, complete genome	Bacillus cereus	38.2	72.4	100%	2.2	100.00%	5291741
CP040340.1	Bacillus cereus strain DLOU-Tangshan chromosome, complete genome	Bacillus cereus	38.2	72.4	100%	2.2	100.00%	5542157
CP040334.1	Bacillus cereus strain DLOU-Changhai chromosome, complete genome	Bacillus cereus	38.2	72.4	100%	2.2	100.00%	5389366
CP045777.1	**Bacillus paranthracis strain CFSAN068816 chromosome, complete genome**	Bacillus paranthracis	38.2	70.4	86%	2.2	100.00%	5351395
CP045606.2	Bacillus cereus strain SB1 chromosome, complete genome	Bacillus cereus	38.2	72.4	100%	2.2	100.00%	5461537
CP045537.1	Bacillaceae bacterium C05 chromosome, complete genome	Bacillaceae bacterium C05	38.2	70.4	86%	2.2	100.00%	5258784
CP045533.1	Bacillaceae bacterium C02 chromosome, complete genome	Bacillaceae bacterium C02	38.2	70.4	86%	2.2	100.00%	5188692
CP042270.1	Bacillus thuringiensis strain GA-A07 chromosome, complete genome	Bacillus thuringiensis	38.2	72.4	100%	2.2	100.00%	5272357
CP044978.1	Bacillus thuringiensis strain BT62 chromosome, complete genome	Bacillus thuringiensis	38.2	72.4	100%	2.2	100.00%	5294964
CP032365.1	Bacillus wiedmannii strain SR52 chromosome, complete genome	Bacillus wiedmannii	38.2	38.2	86%	2.2	100.00%	5448361
XM_030889606.1	PREDICTED: Sitophilus oryzae uncharacterized LOC115874445 (LOC115874445), mRNA	rice weevil	38.2	38.2	86%	2.2	100.00%	1156
XM_030911861.1	PREDICTED: Sitophilus oryzae uncharacterized LOC115891418 (LOC115891418), mRNA	rice weevil	38.2	38.2	86%	2.2	100.00%	1572
CP031068.1 **1/2 8/1**	Bacillus sp. JAS24-2 chromosome, complete genome	Bacillus sp. JAS24-2	38.2	38.2	86%	2.2	100.00%	5225480
CP031065.1 **1/2 8/1**	Bacillus sp. SH8-8 chromosome, complete genome	Bacillus sp. SH8-8	38.2	72.4	100%	2.2	100.00%	5241886
CP031062.1 **1/2 8/1**	Bacillus sp. AR4-2 chromosome, complete genome	Bacillus sp. AR4-2	38.2	72.4	100%	2.2	100.00%	5241257
CP042929.1	Bacillus cereus strain G1-1 chromosome, complete genome	Bacillus cereus	38.2	72.4	100%	2.2	100.00%	5544716
CP042874.1	Bacillus cereus strain 09 chromosome, complete genome	Bacillus cereus	38.2	72.4	100%	2.2	100.00%	5250357
LR697109.1 **1/2 2/2 3/3 3/4 8/1**	**Chanos chanos genome assembly, chromosome: 4**	milkfish	38.2	70.4	95%	2.2	100.00%	53909382
CP031778.1	Bacillus cereus strain Co1-1 chromosome	Bacillus cereus	38.2	70.4	100%	2.2	100.00%	6375991
LR597470.1 **1/2 2/2 3/5 3/6**	Sphaeramia orbicularis genome assembly, chromosome: 13	orbiculate cardinalfish	38.2	104	90%	2.2	100.00%	53310899
CP041071.1	Bacillus tropicus strain LM1212-W3 chromosome, complete genome	Bacillus tropicus	38.2	38.2	86%	2.2	100.00%	5631167
CP029454.1	Bacillus cereus strain FORC087 chromosome, complete genome	Bacillus cereus	38.2	38.2	86%	2.2	100.00%	5271204
AP019731.1	**Bacillus anthracis PCr DNA, complete genome**	Bacillus anthracis	38.2	38.2	86%	2.2	100.00%	5209717
CP034551.1	Bacillus cereus ATCC 14579 chromosome, complete genome	Bacillus cereus ATCC 14579	38.2	72.4	100%	2.2	100.00%	5416249

CP029323.1	**Bacillus anthracis strain 17OD930 chromosome, complete genome**	Bacillus anthracis	38.2	38.2	86%	2.2	100.00%	5213297
CP029468.1	Bacillus cereus strain FORC_086 chromosome, complete genome	Bacillus cereus	38.2	72.4	100%	2.2	100.00%	5340498
CP039269.1	Bacillus cereus strain MH19 chromosome, complete genome	Bacillus cereus	38.2	70.4	86%	2.2	100.00%	5247580
CP026607.1 **1/2**	Bacillus sp. DU-106 chromosome, complete genome	Bacillus sp. DU-106	38.2	72.4	100%	2.2	100.00%	5415320
CP028516.1	Bacillus cereus strain 1000305 chromosome, complete genome	Bacillus cereus	38.2	72.4	100%	2.2	100.00%	5330209
LR535833.1	Mastacembelus armatus genome assembly, chromosome: 1	zig-zag eel	38.2	38.2	86%	2.2	100.00%	26083871
CP026608.1	**Bacillus anthracis strain HDZK-BYSB7 chromosome, complete genome**	Bacillus anthracis	38.2	70.4	86%	2.2	100.00%	5294779
CP024684.1	Bacillus wiedmannii bv. thuringiensis strain FCC41 chromosome, complete genome	Bacillus wiedmannii bv. thuringiensis	38.2	38.2	86%	2.2	100.00%	5513449
CP033795.1 **1/2 8/1**	Bacillus sp. FDAARGOS_527 chromosome, complete genome	Bacillus sp. FDAARGOS_527	38.2	70.4	86%	2.2	100.00%	5277296
CP031443.1	Bacillus mobilis strain ML-A2C4 chromosome, complete genome	Bacillus mobilis	38.2	38.2	86%	2.2	100.00%	5468285
CP024771.1	Bacillus thuringiensis LM1212 chromosome, complete genome	Bacillus thuringiensis LM1212	38.2	38.2	86%	2.2	100.00%	5705934
AP018443.1	**Bacillus anthracis CZC5 DNA, complete genome**	Bacillus anthracis	38.2	38.2	86%	2.2	100.00%	5219278
CP031642.1	**Bacillus anthracis strain MCCC 1A02161 chromosome, complete genome**	Bacillus anthracis	38.2	70.4	86%	2.2	100.00%	5231857
CP031643.1	**Bacillus anthracis strain MCCC 1A01412 chromosome, complete genome**	Bacillus anthracis	38.2	70.4	86%	2.2	100.00%	5291783
CP034516.1	Eukaryotic synthetic construct chromosome 13	eukaryotic synthetic construct	38.2	193	86%	2.2	100.00%	96089878
CP034491.1	Eukaryotic synthetic construct chromosome 13	eukaryotic synthetic construct	38.2	193	86%	2.2	100.00%	96089878
CP021436.1	Bacillus thuringiensis strain C15 chromosome, complete genome	Bacillus thuringiensis	38.2	72.4	100%	2.2	100.00%	5637049
CP034548.1	Bacillus albus strain PFYN01 chromosome, complete genome	Bacillus albus	38.2	38.2	86%	2.2	100.00%	4939577
CP029805.1	**Bacillus anthracis strain London_499 chromosome, complete genome**	Bacillus anthracis	38.2	38.2	86%	2.2	100.00%	5228599
CP020383.1	Bacillus cereus strain FORC60 chromosome, complete genome	Bacillus cereus	38.2	72.4	100%	2.2	100.00%	5361178
CP026678.1	Bacillus cereus strain TG1-6 chromosome, complete genome	Bacillus cereus	38.2	72.4	100%	2.2	100.00%	5417196
CP022044.2	**Bacillus anthracis strain FDAARGOS_341 chromosome, complete genome**	Bacillus anthracis	38.2	38.2	86%	2.2	100.00%	5228806
CP026375.1	Bacillus cereus ATCC 10987 chromosome	Bacillus cereus ATCC 10987	38.2	104	100%	2.2	100.00%	4871269
CP026376.1	Bacillus cereus strain G9241 chromosome	Bacillus cereus	38.2	104	100%	2.2	100.00%	4880708
CP025122.1 **1/2 8/1**	Bacillus sp. HBCD-sjtu chromosome, complete genome	Bacillus sp. HBCD-sjtu	38.2	70.4	86%	2.2	100.00%	5230501
CP024655.1	Bacillus cereus strain MLY1 chromosome MLY1.0, complete sequence	Bacillus cereus	38.2	72.4	100%	2.2	100.00%	5561815
CP023727.1	Bacillus cereus strain BHU1 chromosome	Bacillus cereus	38.2	72.4	100%	2.2	100.00%	5204644
CP023726.1	Bacillus cereus strain BHU2 chromosome	Bacillus cereus	38.2	68.4	86%	2.6	100.00%	5186653
CP023179.1	Bacillus cereus strain CC-1 chromosome, complete genome	Bacillus cereus	38.2	70.4	86%	2.6	100.00%	5279524

CP023245.1	Bacillus cereus strain HBL-AI chromosome, complete genome	Bacillus cereus	38.2	72.4	100%	2.6	100.00%	5482892
CP023001.1	**Bacillus anthracis strain 14RA5914, complete genome**	Bacillus anthracis	38.2	38.2	86%	2.6	100.00%	5245242
CP020754.1	Bacillus thuringiensis strain ATCC 10792, complete genome	Bacillus thuringiensis	38.2	72.4	100%	2.6	100.00%	5528419
CP022445.1	Bacillus cereus C1L chromosome, complete genome	Bacillus cereus C1L	38.2	72.4	100%	2.6	100.00%	5312355

ALIGNMENT VIEW FLAT QUERY-ANCHORED WITH LETTERS FOR IDENTITIES

Query range 1: 1 to 22

Query	1	TATCCCATTACATGATATTATT	22
XM_033872625.1	2805	TCCCATTACATGATATTATT	2786
XM_033872624.1	2224	TCCCATTACATGATATTATT	2205
XM_033872623.1	2670	TCCCATTACATGATATTATT	2651
XM_033872622.1	2811	TCCCATTACATGATATTATT	2792
LR736856.1	25169672	TCCCATTACATGATATTATT	25169653
LR736856.1	953986	ATTACATGATATTATT	953971
LR736856.1	20757468	ATTACATGATATTATT	20757453
LR736856.1	2076285	ATTACATGATATTAT	2076271
LR736856.1	15300707	ATTACATGATATTAT	15300693
LR736856.1	24907107	ATTACATGATATTAT	24907121
CP054800.1	**208258**	**TATCCCATTACATGATATT**	**208276**
CP054797.1	**65999**	**TATCCCATTACATGATATT**	**65981**
CP054568.1	4982888	TATCCCATTACATGATATT	4982870
CP050183.1	2588912	TATCCCATTACATGATATT	2588894
CP050183.1	4984958	CATTACATGATATTATT	4984942
CP053938.1	1731233	TATCCCATTACATGATATT	1731215
CP053938.1	587161	CCCATTACATGATATT	587176
CP053934.1	2468315	TATCCCATTACATGATATT	2468333
CP053934.1	3602104	CCCATTACATGATATT	3602089
CP053931.1	4336123	TATCCCATTACATGATATT	4336141
CP053931.1	216870	CCCATTACATGATATT	216855
CP053997.1	615231	TATCCCATTACATGATATT	615249
CP053997.1	1757114	CCCATTACATGATATT	1757099
CP053991.1	2062698	TATCCCATTACATGATATT	2062680
CP053991.1	920826	CCCATTACATGATATT	920841
CP053980.1	4574420	TATCCCATTACATGATATT	4574402
CP053980.1	3420689	CCCATTACATGATATT	3420704
CP053954.1	5056629	TATCCCATTACATGATATT	5056611
CP053965.1	3056974	TATCCCATTACATGATATT	3056992
CP053955.1	3671059	TATCCCATTACATGATATT	3671077
CP053951.1	1001141	TATCCCATTACATGATATT	1001159
CP053981.1	3437422	TATCCCATTACATGATATT	3437404
CP053289.1	2826804	TATCCCATTACATGATATT	2826822
CP053289.1	540310	CATTACATGATATTATT	540326
CP049978.1	2581345	TATCCCATTACATGATATT	2581327
CP050973.1	**1083080**	**TATCCCATTACATGATATT**	**1083098**
CP050972.1	**3047578**	**TATCCCATTACATGATATT**	**3047596**
CP050971.1	**61005**	**TATCCCATTACATGATATT**	**60987**
CP050970.1	**4915760**	**TATCCCATTACATGATATT**	**4915742**
CP049019.1	1935112	TATCCCATTACATGATATT	1935094
CP049019.1	791722	CCCATTACATGATATT	791737
CP048687.1	**5037812**	**TATCCCATTACATGATATT**	**5037830**
CP048687.1	**430393**	**CCCATTACATGATATT**	**430378**
AP022643.1	68533	TATCCCATTACATGATATT	68515
CP028009.1	5140082	TATCCCATTACATGATATT	5140100
CP027920.1	2880954	TATCCCATTACATGATATT	2880936
CP041981.1	**2708055**	**TATCCCATTACATGATATT**	**2708037**
CP041981.1	**1645460**	**CCCATTACATGATATT**	**1645475**
CP041750.1	**2498278**	**TATCCCATTACATGATATT**	**2498260**
CP041750.1	**1427940**	**CCCATTACATGATATT**	**1427955**
CP047131.1	**2458699**	**TATCCCATTACATGATATT**	**2458681**
CP047111.1	**400614**	**TATCCCATTACATGATATT**	**400632**
CP047107.1	**4514812**	**TATCCCATTACATGATATT**	**4514830**
CP047104.1	**231066**	**TATCCCATTACATGATATT**	**231048**

CP047099.1	**2849926**	**TATCCCATTACATGATATT**	**2849908**
CP047098.1	**5146998**	**TATCCCATTACATGATATT**	**5147016**
CP047097.1	**1622506**	**TATCCCATTACATGATATT**	**1622488**
CP047085.1	**1792017**	**TATCCCATTACATGATATT**	**1791999**
CP047085.1	**717567**	**CCCATTACATGATATT**	**717582**
CP046398.1	2475271	TATCCCATTACATGATATT	2475253
CP046398.1	4811602	CATTACATGATATTATT	4811586
CP040344.1	4155946	TATCCCATTACATGATATT	4155928
CP040342.1	3933233	TATCCCATTACATGATATT	3933251
CP040342.1	1632121	CATTACATGATATTATT	1632137
CP040340.1	3334375	TATCCCATTACATGATATT	3334357
CP040340.1	122423	CATTACATGATATTATT	122407
CP040334.1	3724392	TATCCCATTACATGATATT	3724410
CP040334.1	1469776	CATTACATGATATTATT	1469792
CP045777.1	**3106170**	**TATCCCATTACATGATATT**	**3106188**
CP045777.1	**4177604**	**CCCATTACATGATATT**	**4177589**
CP045606.2	2670105	TATCCCATTACATGATATT	2670087
CP045606.2	4981301	CATTACATGATATTATT	4981285
CP045537.1	2495274	TATCCCATTACATGATATT	2495256
CP045537.1	1459815	CCCATTACATGATATT	1459830
CP045533.1	2467973	TATCCCATTACATGATATT	2467955
CP045533.1	1432466	CCCATTACATGATATT	1432481
CP042270.1	3697495	TATCCCATTACATGATATT	3697477
CP042270.1	726299	CATTACATGATATTATT	726283
CP044978.1	3359847	TATCCCATTACATGATATT	3359829
CP044978.1	313889	CATTACATGATATTATT	313873
CP032365.1	2623129	TATCCCATTACATGATATT	2623111
XM_030889606.1	1055	CCCATTACATGATATTATT	1037
XM_030911861.1	1422	CCCATTACATGATATTATT	1404
CP031068.1	2497032	TATCCCATTACATGATATT	2497014
CP031065.1	2530420	TATCCCATTACATGATATT	2530402
CP031065.1	4747176	CATTACATGATATTATT	4747160
CP031062.1	2529186	TATCCCATTACATGATATT	2529168
CP031062.1	4746547	CATTACATGATATTATT	4746531
CP042929.1	2716926	TATCCCATTACATGATATT	2716908
CP042929.1	5050716	CATTACATGATATTATT	5050700
CP042874.1	2565787	TATCCCATTACATGATATT	2565769
CP042874.1	4779721	CATTACATGATATTATT	4779705
LR697109.1	15346309	ATCCCATTACATGATATTA	15346327
LR697109.1	10387640	ATTACATGATATTATT	10387625
CP031778.1	2451680	TATCCCATTACATGATATT	2451662
CP031778.1	231370	ATTACATGATATTATT	231355
LR597470.1	39095160	TCCCATTACATGATATTAT	39095142
LR597470.1	39192313	TCCCATTACATGATATT	39192297
LR597470.1	32074587	ATTACATGATATTATT	32074572
CP041071.1	1106485	TATCCCATTACATGATATT	1106503
CP029454.1	2546892	TATCCCATTACATGATATT	2546874
AP019731.1	**1368346**	**TATCCCATTACATGATATT**	**1368328**
CP034551.1	**2649612**	**TATCCCATTACATGATATT**	**2649594**
CP034551.1	4905565	CATTACATGATATTATT	4905549
CP029323.1	**2458998**	**TATCCCATTACATGATATT**	**2458980**
CP029468.1	2587934	TATCCCATTACATGATATT	2587916
CP029468.1	4844440	CATTACATGATATTATT	4844424
CP039269.1	2521062	TATCCCATTACATGATATT	2521044
CP039269.1	1401018	CCCATTACATGATATT	1401033
CP026607.1	2524526	TATCCCATTACATGATATT	2524508
CP026607.1	4939190	CATTACATGATATTATT	4939174
CP028516.1	2600026	TATCCCATTACATGATATT	2600008
CP028516.1	4838096	CATTACATGATATTATT	4838080
LR535833.1	16826115	CCCATTACATGATATTATT	16826133

CP026608.1	**3594133**	**TATCCCATTACATGATATT**	**3594115**
CP026608.1	**2431380**	**CCCATTACATGATATT**	**2431395**
CP024684.1	2707607	TATCCCATTACATGATATT	2707589
CP033795.1	1774691	TATCCCATTACATGATATT	1774673
CP033795.1	693315	CCCATTACATGATATT	693330
CP031443.1	2734004	TATCCCATTACATGATATT	2733986
CP024771.1	4372606	TATCCCATTACATGATATT	4372624
AP018443.1	**2463024**	**TATCCCATTACATGATATT**	**2463006**
CP031642.1	**2575655**	**TATCCCATTACATGATATT**	**2575637**
CP031642.1	**1399303**	**CCCATTACATGATATT**	**1399318**
CP031643.1	**2546571**	**TATCCCATTACATGATATT**	**2546553**
CP031643.1	**1369983**	**CCCATTACATGATATT**	**1369998**
CP034516.1	10452935	CCCATTACATGATATTATT	10452917
CP034516.1	51256638	ATTACATGATATTATT	51256653
CP034516.1	87641583	ATTACATGATATTATT	87641598
CP034516.1	43658963	TTACATGATATTATT	43658949
CP034516.1	57611062	ATTACATGATATTAT	57611076
CP034516.1	81750531	ATTACATGATATTAT	81750517
CP034491.1	10452935	CCCATTACATGATATTATT	10452917
CP034491.1	51256638	ATTACATGATATTATT	51256653
CP034491.1	87641583	ATTACATGATATTATT	87641598
CP034491.1	43658963	TTACATGATATTATT	43658949
CP034491.1	57611062	ATTACATGATATTAT	57611076
CP034491.1	81750531	ATTACATGATATTAT	81750517
CP021436.1	2717244	TATCCCATTACATGATATT	2717226
CP021436.1	5122524	CATTACATGATATTATT	5122508
CP034548.1	2730671	TATCCCATTACATGATATT	2730689
CP029805.1	**2472333**	**TATCCCATTACATGATATT**	**2472315**
CP020383.1	2518436	TATCCCATTACATGATATT	2518418
CP020383.1	4881901	CATTACATGATATTATT	4881885
CP026678.1	4967235	TATCCCATTACATGATATT	4967253
CP026678.1	2702343	CATTACATGATATTATT	2702359
CP022044.2	**43833**	**TATCCCATTACATGATATT**	**43815**
CP026375.1	2285751	TATCCCATTACATGATATT	2285733
CP026375.1	4404267	CATTACATGATATTATT	4404251
CP026375.1	1252099	CCCATTACATGATATT	1252114
CP026376.1	4850681	TATCCCATTACATGATATT	4850699
CP026376.1	2724168	CATTACATGATATTATT	2724184
CP026376.1	999576	CCCATTACATGATATT	999561
CP025122.1	5166699	TATCCCATTACATGATATT	5166681
CP025122.1	4062674	CCCATTACATGATATT	4062689
CP024655.1	2770639	TATCCCATTACATGATATT	2770657
CP024655.1	4740463	CATTACATGATATTATT	4740447
CP023727.1	2405971	TATCCCATTACATGATATT	2405953
CP023727.1	4705517	CATTACATGATATTATT	4705501
CP064072.1	2498158	TATCCCATTACATGATATT	2498140
CP064072.1	1426675	CCCATTACATGATATT	1426690
CP023726.1	4159070	TATCCCATTACATGATATT	4159052
CP023726.1	3458987	CCATTACATGATATT	3458973
CP023179.1	2700310	TATCCCATTACATGATATT	2700328
CP023179.1	1469206	CCCATTACATGATATT	1469221
CP023245.1	2709891	TATCCCATTACATGATATT	2709873
CP023245.1	4999622	CATTACATGATATTATT	4999606
CP023001.1	**237616**	**TATCCCATTACATGATATT**	**237598**
CP020754.1	990178	TATCCCATTACATGATATT	990196
CP020754.1	4170375	CATTACATGATATTATT	4170391
CP022445.1	2600932	TATCCCATTACATGATATT	2600914
CP022445.1	4846937	CATTACATGATATTATT	4846921
CP022345.1	2273085	TATCCCATTACATGATATT	2273103
CP022345.1	4598157	ATTACATGATATTATT	4598172

ALIGNMENT VIEW FLAT QUERY-ANCHORED WITH LETTERS FOR IDENTITIES

Bacillus anthracis strain FDAARGOS_702 chromosome

GenBank: CP054800.1

FASTA Graphics

Go to:

LOCUS CP054800 19 bp DNA linear BCT **17-JUN-2020**
DEFINITION Bacillus anthracis strain FDAARGOS_702 chromosome.
ACCESSION CP054800 REGION: 208258..208276
VERSION CP054800.1
DBLINK BioProject: PRJNA231221
BioSample: SAMN11056417
KEYWORDS .
SOURCE Bacillus anthracis
ORGANISM Bacillus anthracis
Bacteria; Firmicutes; Bacilli; Bacillales; Bacillaceae; Bacillus;
Bacillus cereus group.
REFERENCE 1 (bases 1 to 19)
AUTHORS Sozhamannan,S., Rosenzweig,N., Tallon,L., Sadzewicz,L.,
Vavikolanu,K., Mehta,A., Aluvathingal,J., Nadendla,S., Yan,Y. and
Sichtig,H.
TITLE FDA dAtabase for Regulatory Grade micrObial Sequences (FDA-ARGOS):
Supporting development and validation of Infectious Disease Dx
tests
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 19)
AUTHORS Sozhamannan,S., Rosenzweig,N., Tallon,L., Sadzewicz,L.,
Vavikolanu,K., Mehta,A., Aluvathingal,J., Nadendla,S., Yan,Y. and
Sichtig,H.
TITLE Direct Submission
JOURNAL **Submitted (25-NOV-2019) Center for Devices and Radiological Health, US Food and Drug Administration, 10903 New Hampshire Avenue, Silver Spring, MD 20993-0002, USA**
COMMENT Assembly not circularized.
The annotation was added by the NCBI Prokaryotic Genome Annotation
Pipeline (PGAP). Information about PGAP can be found here:
https://www.ncbi.nlm.nih.gov/genome/annotation_prok/

##Genome-Assembly-Data-START##
Assembly Method :: SMRT v. 2.3.0, HGAP v. 3, SPAdes v. 3.6.0
Genome Coverage :: 923.93x

```
            Sequencing Technology :: Pacbio; Illumina
            ##Genome-Assembly-Data-END##


            ##Genome-Annotation-Data-START##
            Annotation Provider               :: NCBI
            Annotation Date                   :: 06/10/2020 00:01:37
            Annotation Pipeline               :: NCBI Prokaryotic Genome
                                                 Annotation Pipeline (PGAP)
            Annotation Method                 :: Best-placed reference protein
                                                 set; GeneMarkS-2+
            Annotation Software revision      :: 4.11
            Features Annotated                :: Gene; CDS; rRNA; tRNA; ncRNA;
                                                 repeat_region
            Genes (total)                     :: 5,533
            CDSs (total)                      :: 5,400
            Genes (coding)                    :: 5,101
            CDSs (with protein)               :: 5,101
            Genes (RNA)                       :: 133
            rRNAs                             :: 11, 11, 11 (5S, 16S, 23S)
            complete rRNAs                    :: 11, 11, 11 (5S, 16S, 23S)
            tRNAs                             :: 95
            ncRNAs                            :: 5
            Pseudo Genes (total)              :: 299
            CDSs (without protein)            :: 299
            Pseudo Genes (ambiguous residues) :: 0 of 299
            Pseudo Genes (frameshifted)       :: 215 of 299
            Pseudo Genes (incomplete)         :: 38 of 299
            Pseudo Genes (internal stop)      :: 67 of 299
            Pseudo Genes (multiple problems)  :: 18 of 299
            ##Genome-Annotation-Data-END##
FEATURES             Location/Qualifiers
     source          1..19
                     /organism="Bacillus anthracis"
                     /mol_type="genomic DNA"
                     /strain="FDAARGOS_702"
                     /culture_collection="FDA:FDAARGOS_702"
                     /db_xref="taxon:1392"
                     /country="USA: MD"
                     /collected_by="U.S. Army Edgewood Chemical Biological
                     Center"
                     /note="genetically modified strain"
     gene            complement(<1..>19)
                     /locus_tag="FOC10_01165"
     CDS             complement(<1..>19)
```

/locus_tag="FOC10_01165"
/inference="COORDINATES: similar to AA
sequence:RefSeq:WP_000849936.1"
/note="Derived by automated computational analysis using
gene prediction method: Protein Homology."
/codon_start=3
/transl_table=11
/product="**serine hydrolase**"
/protein_id="**QKU12340.1**"
/translation="MKTKDQIDNIVKETYRHIDFSGVILVKEKKGIVYEEVFGYANRN
ECINNTLQTRFGIASGCKIFTAIGICQLVENGVITFQTKLKECLSVNFPNFNEDITIH
HLLTHSSGIPDYFDESIMDNFEDLWKQTPMYLLKSLKDFLPLFQNRDMMYAPGSKFHY
NNAGFIILGLIIEEQTGLTFTEYIEKNIFNPIGMNHSGYFSLDRLPRQTALGYIKDEI
SQTWRTNAYSIPIKGGSDGGAFITASDMLKFWEALFNYEIISQEYTKILLTPHIQVNN
NQSYGYGIWIETRENKIF**KYHVMGY**DPGVSFRSAVYPDLGITLVIPSNKGAGAEKLMI
EIEGSF"

ORIGIN

1 **tatcccatta catgatatt**

//

serine hydrolase [Bacillus anthracis]

GenBank: QKU12340.1

Identical Proteins FASTA Graphics

LOCUS QKU12340 340 aa linear BCT **17-JUN-2020**
DEFINITION serine hydrolase [Bacillus anthracis].
ACCESSION QKU12340
VERSION QKU12340.1
DBLINK BioProject: PRJNA231221
BioSample: SAMN11056417
DBSOURCE accession CP054800.1
KEYWORDS .
SOURCE Bacillus anthracis
ORGANISM Bacillus anthracis
Bacteria; Firmicutes; Bacilli; Bacillales; Bacillaceae; Bacillus;
Bacillus cereus group.
REFERENCE 1 (residues 1 to 340)
AUTHORS Sozhamannan,S., Rosenzweig,N., Tallon,L., Sadzewicz,L.,
Vavikolanu,K., Mehta,A., Aluvathingal,J., Nadendla,S., Yan,Y. and
Sichtig,H.
TITLE FDA dAtabase for Regulatory Grade micrObial Sequences (FDA-ARGOS):
Supporting development and validation of Infectious Disease Dx
tests
JOURNAL Unpublished

```
REFERENCE   2  (residues 1 to 340)
  AUTHORS   Sozhamannan,S., Rosenzweig,N., Tallon,L., Sadzewicz,L.,
            Vavikolanu,K., Mehta,A., Aluvathingal,J., Nadendla,S., Yan,Y. and
            Sichtig,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-NOV-2019) Center for Devices and Radiological
            Health,US Food and Drug Administration, 10903 New Hampshire
            BbAvenue, Silver Spring, MD 20993-0002, USA
COMMENT     The annotation was added by the NCBI Prokaryotic Genome Annotation
            Pipeline (PGAP). Information about PGAP can be found here:
            https://www.ncbi.nlm.nih.gov/genome/annotation prok/


            ##Genome-Annotation-Data-START##
            Annotation Provider               :: NCBI
            Annotation Date                   :: 06/10/2020 00:01:37
            Annotation Pipeline               :: NCBI Prokaryotic Genome
                                                 Annotation Pipeline (PGAP)
            Annotation Method                 :: Best-placed reference protein
                                                 set; GeneMarkS-2+
            Annotation Software revision      :: 4.11
            Features Annotated                :: Gene; CDS; rRNA; tRNA; ncRNA;
                                                 repeat_region
            Genes (total)                     :: 5,533
            CDSs (total)                      :: 5,400
            Genes (coding)                    :: 5,101
            CDSs (with protein)               :: 5,101
            Genes (RNA)                       :: 133
            rRNAs                             :: 11, 11, 11 (5S, 16S, 23S)
            complete rRNAs                    :: 11, 11, 11 (5S, 16S, 23S)
            tRNAs                             :: 95
            ncRNAs                            :: 5
            Pseudo Genes (total)              :: 299
            CDSs (without protein)            :: 299
            Pseudo Genes (ambiguous residues) :: 0 of 299
            Pseudo Genes (frameshifted)       :: 215 of 299
            Pseudo Genes (incomplete)         :: 38 of 299
            Pseudo Genes (internal stop)      :: 67 of 299
            Pseudo Genes (multiple problems)  :: 18 of 299
            ##Genome-Annotation-Data-END##
FEATURES             Location/Qualifiers
     source          1..340
                     /organism="Bacillus anthracis"
                     /strain="FDAARGOS_702"
                     /culture_collection="FDA:FDAARGOS_702"
```

/db_xref="taxon:1392"
/country="USA: MD"
/collected_by="U.S. Army Edgewood Chemical Biological Center"
/note="genetically modified strain"

Protein 1..340
/product="serine hydrolase"

Region 5..325
/region_name="AmpC"
/note="CubicO group peptidase, beta-lactamase class C family [Defense mechanisms]; COG1680"
/db_xref="CDD:224594"

CDS 1..340
/locus_tag="FOC10_01165"
/coded_by="complement(CP054800.1:208137..209159)"
/inference="COORDINATES: similar to AA sequence:RefSeq:WP_000849936.1"
/note="Derived by automated computational analysis using gene prediction method: Protein Homology."
/transl_table=11

ORIGIN
```
        1 mktkdqidni vketyrhidf sgvilvkekk givyeevfgy anrnecinnt lqtrfgiasg
       61 ckiftaigic qlvengvitf qtklkeclsv nfpnfnedit ihhllthssg ipdyfdesim
      121 dnfedlwkqt pmyllkslkd flplfqnrdm myapgskfhy nnagfiilgl iieeqtgltf
      181 teyieknifn pigmnhsgyf sldrlprqta lgyikdeisq twrtnaysip ikggsdggaf
      241 itasdmlkfw ealfnyeiis qeytkilltp hiqvnnnqsy gygiwietre nkifkyhvmg
      301 ydpgvsfrsa vypdlgitlv ipsnkgagae klmieiegsf
//
```

Pathogen: environmental/food/other sample from Bacillus anthracis

Identifiers

BioSample: SAMN11056417; Sample name: **FDAARGOS_702**; SRA: SRS4855012

Organism

Bacillus anthracis

cellular organisms; Bacteria; Terrabacteria group; Firmicutes; Bacilli; Bacillales; Bacillaceae; Bacillus; Bacillus cereus group

Package

Pathogen: environmental/food/other; version 1.0

Attributes

geographic location	USA:MD
collected by	**U.S. Army Edgewood Chemical Biological Center**
isolation source	not applicable
subsource note	**genetically modified strain**
collection date	Unknown
strain	**FDAARGOS_702**
latitude and longitude	Not applicable
host sex	not applicable
host disease	Not applicable
host age	Not applicable
culture collection	FDA:FDAARGOS_702

host disease outcome	Not applicable
host health state	Not applicable
host description	Not applicable
host disease stage	Not applicable
identification method	Phenotypic and Molecular Methods

BioProject

PRJNA231221

Retrieve all samples from this project

Submission

FDA, Heike Sichtig; **2019-03-05**

Accession:

SAMN11056417

ID:

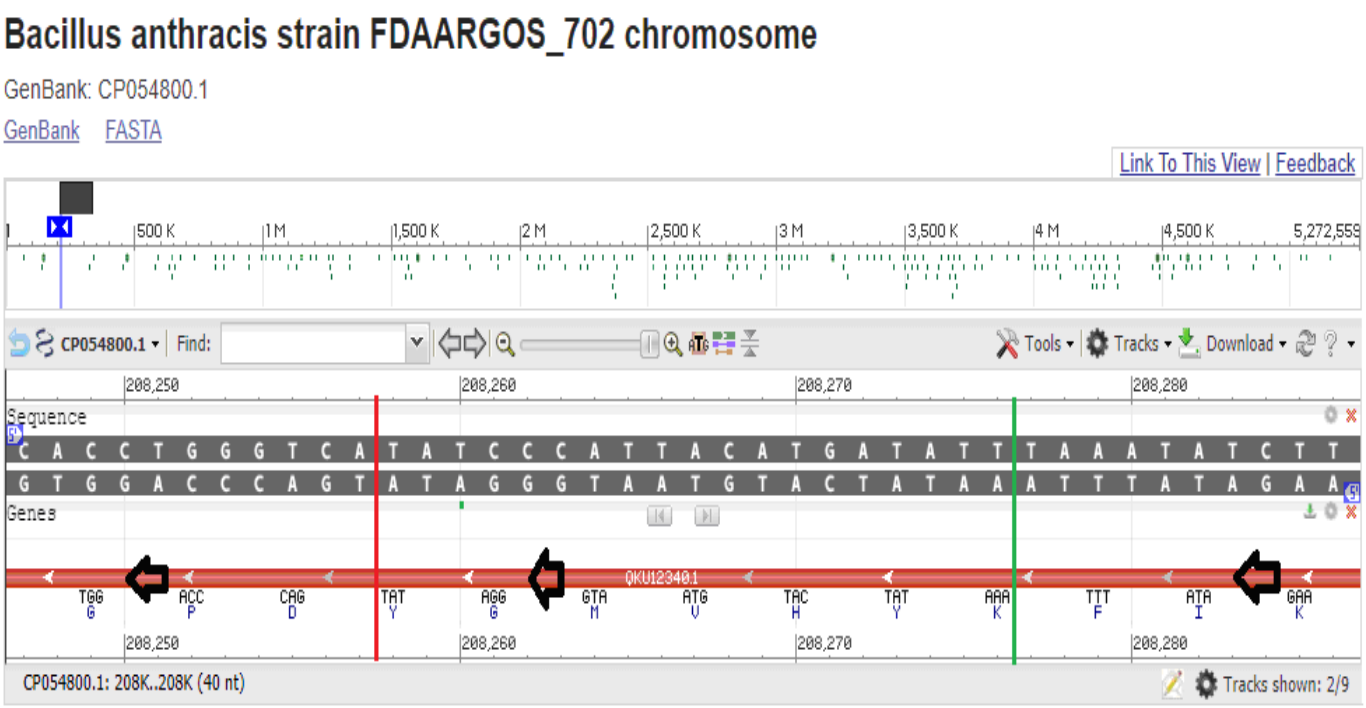
11056417

BioProject SRA Nucleotide

Bacillus anthracis strain FDAARGOS_702 chromosome
Sequence ID: CP054800.1 **Length: 5272559 Number of Matches: 1**
Range 1: 208258 to 208276 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
38.2 bits(19)	2.2	19/19(100%)	0/19(0%)	Plus/Plus

```
CDS: Putative 1       5          G  M  V  H  Y
Query                 1       TATCCCATTACATGATATT  19
                              ||||||||||||||||||
Sbjct                 208258  TATCCCATTACATGATATT  208276
CDS:serine hydrolase  301        Y  G  M  V  H  Y  K
```



Looking at the graph just shown, we notice **two amino acids** common respectively to the sequences (already shown in the previous chapters and that I propose again here below) of the "**Bat SARS-like coronavirus"** and "**Severe acute respiratory syndrome coronavirus 2**".
The amino acids in question are **Valine (V)** and **Methionine (M)**, codified respectively by codons (read according to the direction established by the black arrow) **GTA** and **ATG** [in the graph in the sequence called "*Genes*" and corresponding to **TAC** and **CAT** in the sequence called "*Sequence*"].
The difference that can be seen in the **Bacillus anthracis** sequence compared to the "**Bat SARS-like coronavirus**" and "**Severe acute respiratory syndrome coronavirus 2**" sequences is in the **GTA** codon of **Valine (V)**, in which the base "A" replaces the base "C" in the corresponding codon **GTC** (see the two sequences shown below).

Bat SARS-like coronavirus isolate bat-SL-CoVZC45, complete genome
Sequence ID: MG772933.1 **Length: 29802 Number of Matches: 1**
Range 1: 15418 to 15439 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
44.1 bits(22)	0.036	22/22(100%)	0/22(0%)	Plus/Plus

```
CDS: Putative 1       1          E   M   V   M   C   G   G
Query                 1       GTGAGATGGTCATGTGTGGCGG   22
                              ||||||||||||||||||||||
Sbjct                 15418   GTGAGATGGTCATGTGTGGCGG   15439
CDS:non-structural p  5052    S   E   M   V   M   C   G   G
```

Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/CA-CZB-11679/2020 ORF1ab polyprotein (ORF1ab), ORF1a polyprotein (ORF1ab), surface glycoprotein (S), ORF3a protein (ORF3a), envelope protein (E), membrane glycoprotein (M), ORF6 protein (ORF6), ORF7a protein (ORF7a), ORF7b (ORF7b), ORF8 protein (ORF8), nucleocapsid phosphoprotein (N), and ORF10 protein (ORF10) genes, complete cds
Sequence ID: MW276628.1 Length: 29741 Number of Matches: 1

Range 1: 15377 to 15398 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
40.1 bits(20)	0.59	21/22(95%)	0/22(0%)	Plus/Plus

```
CDS: Putative 1       1          E  M  V  M  C  G  G
Query                 1      GTGAGATGGTCATGTGTGGCGG  22
                             |||| ||||||||||||||||||
Sbjct                 15377  GTGARATGGTCATGTGTGGCGG  15398
CDS:ORF1ab polyprote  5056   S  E  M  V  M  C  G  G
ATTENZIONE: R = A o G
```

In light of what has already been discussed, especially in the previous chapter, the correspondence of these two amino acids (between the three sequences considered above) acquires a certain importance, also because the "**product**" of the **Bacillus anthracis** sequence identified by the BLAST research is **Serine** (the **Serine Hydrolases**), already widely considered in Chapter IX.

Serine hydrolase

Serine hydrolases are one of the largest known enzyme classes comprising approximately ~200 enzymes or 1% of the genes in the human proteome. A defining characteristic of these enzymes is the presence of a **nucleophilic serine** in their *active site*, which is used for the **hydrolysis of substrates**. *Catalysis* proceeds by the formation of an acyl-enzyme intermediate through this **serine**, followed by water/hydroxide-induced saponification of the intermediate and regeneration of the enzyme. Unlike other *non-catalytic serines*, the **nucleophilic serine** of *these hydrolases* is typically activated by a **proton relay** involving a **catalytic triad** consisting of the **serine**, an acidic residue (e.g. *aspartate* or *glutamate*) and a **basic residue** (usually *histidine*), although variations on this mechanism exist. Superfamilies of **serine hydrolases** includes: *Serine proteases, Extracellular lipases, Intracellular lipases, Cholinesterases, Small molecule thioesterases, Some phospholipases, Protein and glycan hydrolases, Some amidases, Some peptidases* (https://en.wikipedia.org/wiki/Serine_hydrolase). As shown on page 282, the **Serine Hydrolase** region name (protein_id="**QKU12340.1**") is "**AmpC**", */note="CubicO group peptidase, beta-lactamase class C family [Defense mechanisms]; **COG1680***" (https://www.ncbi.nlm.nih.gov/Structure/cdd/cddsrv.cgi?uid=COG1680). **Beta Lactamase** (*β-lattamasi, βL*), are a *large family of hydrolases* comprising more than 850 identified members expressed in *Gram-positive* and *Gram-negative bacteria*. βLs can be classified according to their substrate or inhibitor specificity. *These enzymes are capable of hydrolyzing four atom rings known as β-lactams* (https://www.abcam.com/beta-lactamase-activity-assay-kit-colorimetric-ab197008.html?gclsrc=aw.ds|aw.ds&gclid=EAIaIQobChMIiLTdx6Oo7wIVAuqyCh2G6AsAEAAYASAAEgJjrPD_BwE; see also https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi?seqinput=NP_845386.1). **COG1680** is member of the

Transpeptidase Superfamily cl21491, ***Penicillin binding protein transpeptidase domain*. The active site serine** is conserved in all members of this family (https://www.ncbi.nlm.nih.gov/Structure/cdd/cl21491).
This **signature** identifies a *large group of proteins*, which include (https://www.ebi.ac.uk/interpro/entry/InterPro/IPR001460/):
- ***Beta-lactamase precursor, penicillinase*** (A group of enzymes of varying specificity hydrolysing β-lactams; some act more rapidly on *penicillins*, some more rapidly on *cephalosporins*. The latter were formerly listed as EC 3.5.2.8, *cephalosporinase*) (https://www.ebi.ac.uk/intenz/query?cmd=SearchEC&ec=3.5.2.6);
- ***Peptidoglycan synthetase ftsI*** (*peptidoglycan glycosyltransferase 3*) is a enzyme also works when the *lysine* residue is replaced by meso-2,6-diaminoheptanedioate (meso-2,6-diaminopimelate, A2pm) combined with adjacent residues through its L-centre, as it is in *Gram-negative* and some *Gram-positive* organisms. The *undecaprenol* involved is *ditrans, octacis-undecaprenol*. *Involved in the synthesis of cell-wall peptidoglycan* (https://www.ebi.ac.uk/intenz/query?cmd=SearchEC&ec=2.4.1.129).
- ***Methicillin resistance mecR1 protein*** (see below);
- ***Methicillin resistance blaR1 protein*** (see below).
The large number of **penicillin binding proteins**, which are represented in this group of sequences, *are responsible for the final stages of peptidoglycan biosynthesis for cell wall formation*. The proteins synthesise *cross-linked peptidoglycan* from lipid intermediates, and *contain a penicillin-sensitive transpeptidase carboxy-terminal domain*. The **active site serine** (**residue 337 in P14677**) **is conserved in all members of this family**. **MecR1** and **BlaR1** are *metallopeptidases* belonging to MEROPS *peptidase family M56*, *clan M-*. *BlaR1* and *MecR1* cleave their cognate *transcriptional repressors BlaI and MecI*, respectively, activating the *synthesis of MecA*. **MecR1** is present in *Staphylococcus aureus* and *Staphylococcus sciuri*, whereas **BlaR1** (also known as *BlaR*, *PenR1*, or *PenJ*) has been found in *Bacillus licheniformis*, *Staphylococcus epidermidis*, *Staphylococcus haemolyticus*, and several *S. aureus strains*. These proteins are either plasmid-encoded, chromosomal, or transposon-mediated. ***MecR1/BlaR1*** proteins are made up by *homologous N-terminal 330-residue transmembrane metallopeptidase domains linked to extracellular 260-residue homologous PBP-like penicillin sensor moieties* (https://www.ebi.ac.uk/interpro/entry/InterPro/IPR001460/).

Bacillus anthracis strain FDAARGOS_703 chromosome

GenBank: CP054797.1

FASTA Graphics

Go to:

```
LOCUS       CP054797                 19 bp    DNA     linear   BCT 17-JUN-2020
DEFINITION  Bacillus anthracis strain FDAARGOS_703 chromosome.
ACCESSION   CP054797 REGION: 65981..65999
VERSION     CP054797.1
DBLINK      BioProject: PRJNA231221
            BioSample: SAMN11056418
KEYWORDS    .
SOURCE      Bacillus anthracis
  ORGANISM  Bacillus anthracis
            Bacteria; Firmicutes; Bacilli; Bacillales; Bacillaceae; Bacillus;
```

Bacillus cereus group.

REFERENCE 1 (bases 1 to 19)

AUTHORS Sozhamannan,S., Rosenzweig,N., Tallon,L., Sadzewicz,L., Vavikolanu,K., Mehta,A., Aluvathingal,J., Nadendla,S., Yan,Y. and Sichtig,H.

TITLE FDA dAtabase for Regulatory Grade micrObial Sequences (FDA-ARGOS): Supporting development and validation of Infectious Disease Dx tests

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 19)

AUTHORS Sozhamannan,S., Rosenzweig,N., Tallon,L., Sadzewicz,L., Vavikolanu,K., Mehta,A., Aluvathingal,J., Nadendla,S., Yan,Y. and Sichtig,H.

TITLE Direct Submission

JOURNAL **Submitted (25-NOV-2019) Center for Devices and Radiological Health, US Food and Drug Administration, 10903 New Hampshire Avenue, Silver Spring, MD 20993-0002, USA**

COMMENT Assembly not circularized.
The annotation was added by the NCBI Prokaryotic Genome Annotation Pipeline (PGAP). Information about PGAP can be found here: https://www.ncbi.nlm.nih.gov/genome/annotation_prok/

##Genome-Assembly-Data-START##
Assembly Method :: SMRT v. 2.3.0, HGAP v. 3, SPAdes v. 3.6.0
Genome Coverage :: 942.4x
Sequencing Technology :: Pacbio; Illumina
##Genome-Assembly-Data-END##

##Genome-Annotation-Data-START##
Annotation Provider :: NCBI
Annotation Date :: 06/10/2020 00:12:20
Annotation Pipeline :: NCBI Prokaryotic Genome Annotation Pipeline (PGAP)
Annotation Method :: Best-placed reference protein set; GeneMarkS-2+
Annotation Software revision :: 4.11
Features Annotated :: Gene; CDS; rRNA; tRNA; ncRNA; repeat_region
Genes (total) :: 5,514
CDSs (total) :: 5,381
Genes (coding) :: 5,095
CDSs (with protein) :: 5,095
Genes (RNA) :: 133
rRNAs :: 11, 11, 11 (5S, 16S, 23S)

```
            complete rRNAs                    :: 11, 11, 11 (5S, 16S, 23S)
            tRNAs                             :: 95
            ncRNAs                            :: 5
            Pseudo Genes (total)              :: 286
            CDSs (without protein)            :: 286
            Pseudo Genes (ambiguous residues) :: 0 of 286
            Pseudo Genes (frameshifted)       :: 201 of 286
            Pseudo Genes (incomplete)         :: 36 of 286
            Pseudo Genes (internal stop)      :: 71 of 286
            Pseudo Genes (multiple problems)  :: 19 of 286
            ##Genome-Annotation-Data-END##

FEATURES             Location/Qualifiers
     source          1..19
                     /organism="Bacillus anthracis"
                     /mol_type="genomic DNA"
                     /strain="FDAARGOS_703"
                     /culture_collection="FDA:FDAARGOS_703"
                     /db_xref="taxon:1392"
                     /country="USA: MD"
                     /collected_by="U.S. Army Edgewood Chemical Biological
                       Center"
                     /note="genetically modified strain"
     gene            <1..>19
                     /locus_tag="FOC11_00315"
     CDS             <1..>19
                     /locus_tag="FOC11_00315"
                     /inference="COORDINATES: similar to AA
                     sequence:RefSeq:WP_000849936.1"
                     /note="Derived by automated computational analysis using
                     gene prediction method: Protein Homology."
                     /codon_start=3
                     /transl_table=11
                     /product="serine hydrolase"
                     /protein_id="QKU07095.1"
                     /translation="MKTKDQIDNIVKETYRHIDFSGVILVKEKKGIVYEEVFGYANRN
                     ECINNTLQTRFGIASGCKIFTAIGICQLVENGVITFQTKLKECLSVNFPNFNEDITIH
                     HLLTHSSGIPDYFDESIMDNFEDLWKQTPMYLLKSLKDFLPLFQNRDMMYAPGSKFHY
                     NNAGFIILGLIIEEQTGLTFTEYIEKNIFNPIGMNHSGYFSLDRLPRQTALGYIKDEI
                     SQTWRTNAYSIPIKGGSDGGAFITASDMLKFWEALFNYEIISQEYTKILLTPHIQVNN
                     NQSYGYGIWIETRENKIFKYHVMGYDPGVSFRSAVYPDLGITLVIPSNKGAGAEKLMI
                     EIEGSF"
ORIGIN
            1 aatatcatgt aatgggata
//
```

Bacillus anthracis strain FDAARGOS_703 chromosome
Sequence ID: CP054797.1 **Length: 5261520 Number of Matches: 1**
Range 1: 65981 to 65999 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
38.2 bits(19)	2.2	19/19(100%)	0/19(0%)	Plus/Minus

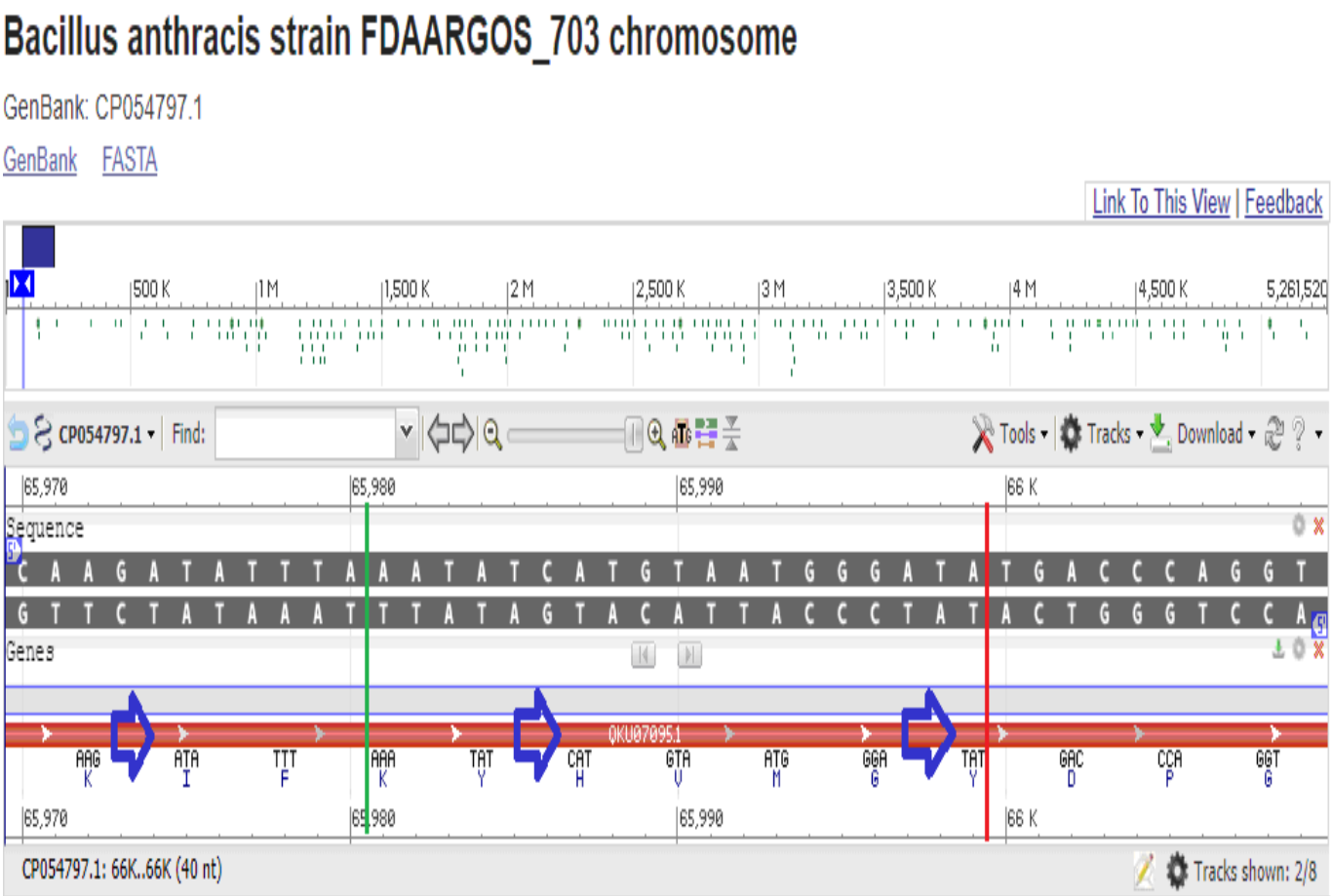
```
CDS: Putative 1        5          G  M  V  H  Y
Query                  1      TATCCCATTACATGATATT  19
                              ||||||||||||||||||
Sbjct                  65999  TATCCCATTACATGATATT  65981
CDS:serine hydrolase   301      Y  G  M  V  H  Y  K
```

ATTENTION:
The sequence "**aatatcatgt aatgggata**" shown above is the *reverse and mirror sequence* of the sequence "**TATCCCATTACATGATATT**" (partial sequence of **Sequence n° 3/1**, "**TATCCCATTACATGATATTATT**"):

ATAGGGTAATGTACTATAA

||||||||||||||||||||

TATCCCATTACATGATATT



Bacillus anthracis strain FDAARGOS_699 chromosome

GenBank: CP050973.1

FASTA Graphics

Go to:

LOCUS CP050973 19 bp DNA linear BCT **12-APR-2020**

DEFINITION Bacillus anthracis strain FDAARGOS_699 chromosome.

ACCESSION CP050973 REGION: 1083080..1083098

VERSION CP050973.1

DBLINK BioProject: PRJNA231221
BioSample: SAMN11056414

KEYWORDS .

SOURCE Bacillus anthracis

ORGANISM Bacillus anthracis
Bacteria; Firmicutes; Bacilli; Bacillales; Bacillaceae; Bacillus;
Bacillus cereus group.

REFERENCE 1 (bases 1 to 19)

AUTHORS Sozhamannan,S., Rosenzweig,N., Tallon,L., Sadzewicz,L.,
Vavikolanu,K., Mehta,A., Aluvathingal,J., Nadendla,S., Yan,Y. and
Sichtig,H.

TITLE FDA dAtabase for Regulatory Grade micrObial Sequences (FDA-ARGOS): Supporting development and validation of Infectious Disease Dx tests

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 19)

AUTHORS Sozhamannan,S., Rosenzweig,N., Tallon,L., Sadzewicz,L.,
Vavikolanu,K., Mehta,A., Aluvathingal,J., Nadendla,S., Yan,Y. and
Sichtig,H.

TITLE Direct Submission

JOURNAL **Submitted (22-NOV-2019) Center for Devices and Radiological Health, US Food and Drug Administration, 10903 New Hampshire Avenue, Silver Spring, MD 20993-0002, USA**

COMMENT The annotation was added by the NCBI Prokaryotic Genome Annotation Pipeline (PGAP). Information about PGAP can be found here: https://www.ncbi.nlm.nih.gov/genome/annotation_prok/

##Genome-Assembly-Data-START##
Assembly Method :: SMRT v. 2.3.0, HGAP v. 3
Genome Coverage :: 823.80x
Sequencing Technology :: Pacbio; Illumina
##Genome-Assembly-Data-END##
##Genome-Annotation-Data-START##
Annotation Provider :: NCBI
Annotation Date :: 04/06/2020 10:47:45
Annotation Pipeline :: NCBI Prokaryotic Genome

```
                                          Annotation Pipeline (PGAP)
            Annotation Method             :: Best-placed reference protein
                                          set; GeneMarkS-2+
            Annotation Software revision  :: 4.11
            Features Annotated            :: Gene; CDS; rRNA; tRNA; ncRNA;
                                          repeat_region
            Genes (total)                 :: 5,481
            CDSs (total)                  :: 5,348
            Genes (coding)                :: 5,064
            CDSs (with protein)           :: 5,064
            Genes (RNA)                   :: 133
            rRNAs                         :: 11, 11, 11 (5S, 16S, 23S)
            complete rRNAs                :: 11, 11, 11 (5S, 16S, 23S)
            tRNAs                         :: 95
            ncRNAs                        :: 5
            Pseudo Genes (total)          :: 284
            CDSs (without protein)        :: 284
            Pseudo Genes (ambiguous residues) :: 0 of 284
            Pseudo Genes (frameshifted)   :: 202 of 284
            Pseudo Genes (incomplete)     :: 35 of 284
            Pseudo Genes (internal stop)  :: 65 of 284
            Pseudo Genes (multiple problems) :: 15 of 284
            ##Genome-Annotation-Data-END##

FEATURES             Location/Qualifiers
     source          1..19
                     /organism="Bacillus anthracis"
                     /mol_type="genomic DNA"
                     /strain="FDAARGOS_699"
                     /isolation_source="synthetic"
                     /culture_collection="FDA:FDAARGOS_699"
                     /db_xref="taxon:1392"
                     /country="USA: MD"
                     /collected_by="U.S. Army Edgewood Chemical Biological
                       Center"
     gene            complement(<1..>19)
                     /locus_tag="FOC07_05665"
     CDS             complement(<1..>19)
                     /locus_tag="FOC07_05665"
                     /inference="COORDINATES: similar to AA
                     sequence:RefSeq:WP_000849936.1"
                     /note="Derived by automated computational analysis using
                     gene prediction method: Protein Homology."
                     /codon_start=3
                     /transl_table=11
```

/product="**serine hydrolase**"
/protein_id="QIY26953.1"
/translation="MKTKDQIDNIVKETYRHIDFSGVILVKEKKGIVYEEVFGYANRN
ECINNTLQTRFGIASGCKIFTAIGICQLVENGVITFQTKLKECLSVNFPNFNEDITIH
HLLTHSSGIPDYFDESIMDNFEDLWKQTPMYLLKSLKDFLPLFQNRDMMYAPGSKFHY
NNAGFIILGLIIEEQTGLTFTEYIEKNIFNPIGMNHSGYFSLDRLPRQTALGYIKDEI
SQTWRTNAYSIPIKGGSDGGAFITASDMLKFWEALFNYEIISQEYTKILLTPHIQVNN
NQSYGYGIWIETRENKIF**KYHVMGY**DPGVSFRSAVYPDLGITLVIPSNKGAGAEKLMI
EIEGSF"

ORIGIN

1 **tatcccatta catgatatt**

//

Bacillus anthracis strain FDAARGOS_699 chromosome
Sequence ID: CP050973.1 **Length: 5227345 Number of Matches: 1**
Range 1: 1083080 to 1083098 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
38.2 bits(19)	2.2	19/19(100%)	0/19(0%)	Plus/Plus

```
CDS: Putative 1       5           G  M  V  H  Y
Query                 1        TATCCCATTACATGATATT  19
                               ||||||||||||||||||
Sbjct                 1083080  TATCCCATTACATGATATT  1083098
CDS:serine hydrolase  301      Y  G  M  V  H  Y  K
```

Bacillus anthracis strain FDAARGOS_697 chromosome

GenBank: CP050972.1

FASTA Graphics

Go to:

LOCUS CP050972 19 bp DNA linear BCT **12-APR-2020**
DEFINITION Bacillus anthracis strain FDAARGOS_697 chromosome.
ACCESSION CP050972 REGION: 3047578..3047596
VERSION CP050972.1
DBLINK BioProject: PRJNA231221
BioSample: SAMN11056412
KEYWORDS .
SOURCE Bacillus anthracis
ORGANISM Bacillus anthracis
Bacteria; Firmicutes; Bacilli; Bacillales; Bacillaceae; Bacillus;
Bacillus cereus group.

REFERENCE 1 (bases 1 to 19)
AUTHORS Sozhamannan,S., Rosenzweig,N., Tallon,L., Sadzewicz,L., Vavikolanu,K., Mehta,A., Aluvathingal,J., Nadendla,S., Yan,Y. and Sichtig,H.
TITLE FDA dAtabase for Regulatory Grade micrObial Sequences (FDA-ARGOS): Supporting development and validation of Infectious Disease Dx tests
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 19)
AUTHORS Sozhamannan,S., Rosenzweig,N., Tallon,L., Sadzewicz,L., Vavikolanu,K., Mehta,A., Aluvathingal,J., Nadendla,S., Yan,Y. and Sichtig,H.
TITLE Direct Submission
JOURNAL **Submitted (22-NOV-2019) Center for Devices and Radiological Health, US Food and Drug Administration, 10903 New Hampshire Avenue, Silver Spring, MD 20993-0002, USA**
COMMENT The annotation was added by the NCBI Prokaryotic Genome Annotation Pipeline (PGAP). Information about PGAP can be found here: https://www.ncbi.nlm.nih.gov/genome/annotation_prok/

##Genome-Assembly-Data-START##
Assembly Method :: SMRT v. 2.3.0, HGAP v. 3
Genome Coverage :: 591.14x
Sequencing Technology :: Pacbio; Illumina
##Genome-Assembly-Data-END##

##Genome-Annotation-Data-START##
Annotation Provider :: NCBI
Annotation Date :: 04/06/2020 10:44:45
Annotation Pipeline :: NCBI Prokaryotic Genome Annotation Pipeline (PGAP)
Annotation Method :: Best-placed reference protein set; GeneMarkS-2+
Annotation Software revision :: 4.11
Features Annotated :: Gene; CDS; rRNA; tRNA; ncRNA; repeat_region
Genes (total) :: 5,499
CDSs (total) :: 5,362
Genes (coding) :: 5,066
CDSs (with protein) :: 5,066
Genes (RNA) :: 137
rRNAs :: 12, 13, 12 (5S, 16S, 23S)
complete rRNAs :: 12, 13, 12 (5S, 16S, 23S)
tRNAs :: 95
ncRNAs :: 5

```
            Pseudo Genes (total)              :: 296
            CDSs (without protein)            :: 296
            Pseudo Genes (ambiguous residues) :: 0 of 296
            Pseudo Genes (frameshifted)       :: 215 of 296
            Pseudo Genes (incomplete)         :: 34 of 296
            Pseudo Genes (internal stop)      :: 65 of 296
            Pseudo Genes (multiple problems)  :: 15 of 296
            ##Genome-Annotation-Data-END##
FEATURES             Location/Qualifiers
     source          1..19
                     /organism="Bacillus anthracis"
                     /mol_type="genomic DNA"
                     /strain="FDAARGOS_697"
                     /isolation_source="synthetic"
                     /culture_collection="FDA:FDAARGOS_697"
                     /db_xref="taxon:1392"
                     /country="USA: MD"
                     /collected_by="U.S. Army Edgewood Chemical Biological
                     Center"
     gene            complement(<1..>19)
                     /locus_tag="FOC05_16190"
     CDS             complement(<1..>19)
                     /locus_tag="FOC05_16190"
                     /inference="COORDINATES: similar to AA
                     sequence:RefSeq:WP_000849936.1"
                     /note="Derived by automated computational analysis using
                     gene prediction method: Protein Homology."
                     /codon_start=3
                     /transl_table=11
                     /product="serine hydrolase"
                     /protein_id="QIY23737.1"
                     /translation="MKTKDQIDNIVKETYRHIDFSGVILVKEKKGIVYEEVFGYANRN
                     ECINNTLQTRFGIASGCKIFTAIGICQLVENGVITFQTKLKECLSVNFPNFNEDITIH
                     HLLTHSSGIPDYFDESIMDNFEDLWKQTPMYLLKSLKDFLPLFQNRDMMYAPGSKFHY
                     NNAGFIILGLIIEEQTGLTFTEYIEKNIFNPIGMNHSGYFSLDRLPRQTALGYIKDEI
                     SQTWRTNAYSIPIKGGSDGGAFITASDMLKFWEALFNYEIISQEYTKILLTPHIQVNN
                     NQSYGYGIWIETRENKIFKYHVMGYDPGVSFRSAVYPDLGITLVIPSNKGAGAEKLMI
                     EIEGSF"
ORIGIN
        1 tatcccatta catgatatt
//
```

Bacillus anthracis strain FDAARGOS_697 chromosome
Sequence ID: CP050972.1 **Length: 5240765 Number of Matches: 1**
Range 1: 3047578 to 3047596 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
38.2 bits(19)	2.2	19/19(100%)	0/19(0%)	Plus/Plus

```
CDS: Putative 1       5           G  M  V  H  Y
Query                 1        TATCCCATTACATGATATT  19
                               ||||||||||||||||||
Sbjct                 3047578  TATCCCATTACATGATATT  3047596
CDS:serine hydrolase  301      Y  G  M  V  H  Y  K
```

Bacillus anthracis strain FDAARGOS_696 chromosome

GenBank: CP050971.1

FASTA Graphics

Go to:

```
LOCUS       CP050971                  19 bp    DNA     linear   BCT 12-APR-2020
DEFINITION  Bacillus anthracis strain FDAARGOS_696 chromosome.
ACCESSION   CP050971 REGION: 60987..61005
VERSION     CP050971.1
DBLINK      BioProject: PRJNA231221
            BioSample: SAMN11056411
KEYWORDS    .
SOURCE      Bacillus anthracis
  ORGANISM  Bacillus anthracis
            Bacteria; Firmicutes; Bacilli; Bacillales; Bacillaceae; Bacillus;
            Bacillus cereus group.
REFERENCE   1  (bases 1 to 19)
  AUTHORS   Sozhamannan,S., Rosenzweig,N., Tallon,L., Sadzewicz,L.,
            Vavikolanu,K., Mehta,A., Aluvathingal,J., Nadendla,S., Yan,Y. and
            Sichtig,H.
  TITLE     FDA dAtabase for Regulatory Grade micrObial Sequences (FDA-ARGOS):
            Supporting development and validation of Infectious Disease Dx
            tests
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 19)
  AUTHORS   Sozhamannan,S., Rosenzweig,N., Tallon,L., Sadzewicz,L.,
            Vavikolanu,K., Mehta,A., Aluvathingal,J., Nadendla,S., Yan,Y. and
            Sichtig,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-NOV-2019) Center for Devices and Radiological Health,
            US Food and Drug Administration, 10903 New Hampshire Avenue, Silver
            Spring, MD 20993-0002, USA
```

```
COMMENT     The annotation was added by the NCBI Prokaryotic Genome Annotation
            Pipeline (PGAP). Information about PGAP can be found here:
            https://www.ncbi.nlm.nih.gov/genome/annotation_prok/


            ##Genome-Assembly-Data-START##
            Assembly Method       :: SMRT v. 2.3.0, HGAP v. 3
            Genome Coverage       :: 246.96x
            Sequencing Technology :: Pacbio; Illumina
            ##Genome-Assembly-Data-END##

            ##Genome-Annotation-Data-START##
            Annotation Provider               :: NCBI
            Annotation Date                   :: 04/06/2020 10:41:57
            Annotation Pipeline               :: NCBI Prokaryotic Genome
                                                 Annotation Pipeline (PGAP)
            Annotation Method                 :: Best-placed reference protein
                                                 set; GeneMarkS-2+
            Annotation Software revision      :: 4.11
            Features Annotated                :: Gene; CDS; rRNA; tRNA; ncRNA;
                                                 repeat_region
            Genes (total)                     :: 5,519
            CDSs (total)                      :: 5,386
            Genes (coding)                    :: 5,097
            CDSs (with protein)               :: 5,097
            Genes (RNA)                       :: 133
            rRNAs                             :: 11, 11, 11 (5S, 16S, 23S)
            complete rRNAs                    :: 11, 11, 11 (5S, 16S, 23S)
            tRNAs                             :: 95
            ncRNAs                            :: 5
            Pseudo Genes (total)              :: 289
            CDSs (without protein)            :: 289
            Pseudo Genes (ambiguous residues) :: 0 of 289
            Pseudo Genes (frameshifted)       :: 206 of 289
            Pseudo Genes (incomplete)         :: 34 of 289
            Pseudo Genes (internal stop)      :: 69 of 289
            Pseudo Genes (multiple problems)  :: 17 of 289
            ##Genome-Annotation-Data-END##
FEATURES             Location/Qualifiers
     source          1..19
                     /organism="Bacillus anthracis"
                     /mol_type="genomic DNA"
                     /strain="FDAARGOS_696"
                     /isolation_source="synthetic"
                     /culture_collection="FDA:FDAARGOS_696"
```

/db_xref="taxon:1392"
/country="USA: MD"
/collected_by="U.S. Army Edgewood Chemical Biological Center"

gene <1..>19
/locus_tag="FOC04_00300"

CDS <1..>19
/locus_tag="FOC04_00300"
/inference="COORDINATES: similar to AA sequence:RefSeq:WP_000849936.1"
/note="Derived by automated computational analysis using gene prediction method: Protein Homology."
/codon_start=3
/transl_table=11
/product="**serine hydrolase**"
/protein_id="QIY15820.1"
/translation="MKTKDQIDNIVKETYRHIDFSGVILVKEKKGIVYEEVFGYANRN
ECINNTLQTRFGIASGCKIFTAIGICQLVENGVITFQTKLKECLSVNFPNFNEDITIH
HLLTHSSGIPDYFDESIMDNFEDLWKQTPMYLLKSLKDFLPLFQNRDMMYAPGSKFHY
NNAGFIILGLIIEEQTGLTFTEYIEKNIFNPIGMNHSGYFSLDRLPRQTALGYIKDEI
SQTWRTNAYSIPIKGGSDGGAFITASDMLKFWEALFNYEIISQEYTKILLTPHIQVNN
NQSYGYGIWIETRENKIF**KYHVMGY**DPGVSFRSAVYPDLGITLVIPSNKGAGAEKLMI
EIEGSF"

ORIGIN
1 **aatatcatgt aatgggata**
//

Bacillus anthracis strain FDAARGOS_696 chromosome
Sequence ID: CP050971.1 **Length: 5255974 Number of Matches: 1**
Range 1: 60987 to 61005 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
38.2 bits(19)	2.2	19/19(100%)	0/19(0%)	Plus/Minus

```
CDS: Putative 1       5         G  M  V  H  Y
Query                 1      TATCCCATTACATGATATT  19
                             ||||||||||||||||||
Sbjct                 61005  TATCCCATTACATGATATT  60987
CDS:serine hydrolase  301     Y  G  M  V  H  Y  K
```

Bacillus anthracis strain FDAARGOS_694 chromosome

GenBank: CP050970.1

FASTA Graphics

Go to:

LOCUS CP050970 19 bp DNA linear BCT **12-APR-2020**

DEFINITION Bacillus anthracis strain FDAARGOS_694 chromosome.

ACCESSION CP050970 REGION: 4915742..4915760

VERSION CP050970.1

DBLINK BioProject: PRJNA231221
BioSample: SAMN11056409

KEYWORDS .

SOURCE Bacillus anthracis

ORGANISM Bacillus anthracis
Bacteria; Firmicutes; Bacilli; Bacillales; Bacillaceae; Bacillus;
Bacillus cereus group.

REFERENCE 1 (bases 1 to 19)

AUTHORS Sozhamannan,S., Rosenzweig,N., Tallon,L., Sadzewicz,L.,
Vavikolanu,K., Mehta,A., Aluvathingal,J., Nadendla,S., Yan,Y. and
Sichtig,H.

TITLE FDA dAtabase for Regulatory Grade micrObial Sequences (FDA-ARGOS):
Supporting development and validation of Infectious Disease Dx
tests

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 19)

AUTHORS Sozhamannan,S., Rosenzweig,N., Tallon,L., Sadzewicz,L.,
Vavikolanu,K., Mehta,A., Aluvathingal,J., Nadendla,S., Yan,Y. and
Sichtig,H.

TITLE Direct Submission

JOURNAL **Submitted (22-NOV-2019) Center for Devices and Radiological Health, US Food and Drug Administration, 10903 New Hampshire Avenue, Silver Spring, MD 20993-0002, USA**

COMMENT The annotation was added by the NCBI Prokaryotic Genome Annotation
Pipeline (PGAP). Information about PGAP can be found here:
https://www.ncbi.nlm.nih.gov/genome/annotation_prok/

##Genome-Assembly-Data-START##
Assembly Method :: SMRT v. 2.3.0, HGAP v. 3
Genome Coverage :: 562.13x
Sequencing Technology :: Pacbio; Illumina
##Genome-Assembly-Data-END##

##Genome-Annotation-Data-START##
Annotation Provider :: NCBI
Annotation Date :: 04/06/2020 10:31:52

```
            Annotation Pipeline                 :: NCBI Prokaryotic Genome
                                                   Annotation Pipeline (PGAP)
            Annotation Method                   :: Best-placed reference protein
                                                   set; GeneMarkS-2+
            Annotation Software revision        :: 4.11
            Features Annotated                  :: Gene; CDS; rRNA; tRNA; ncRNA;
                                                   repeat_region
            Genes (total)                       :: 5,503
            CDSs (total)                        :: 5,370
            Genes (coding)                      :: 5,085
            CDSs (with protein)                 :: 5,085
            Genes (RNA)                         :: 133
            rRNAs                               :: 11, 11, 11 (5S, 16S, 23S)
            complete rRNAs                      :: 11, 11, 11 (5S, 16S, 23S)
            tRNAs                               :: 95
            ncRNAs                              :: 5
            Pseudo Genes (total)                :: 285
            CDSs (without protein)              :: 285
            Pseudo Genes (ambiguous residues)   :: 0 of 285
            Pseudo Genes (frameshifted)         :: 204 of 285
            Pseudo Genes (incomplete)           :: 34 of 285
            Pseudo Genes (internal stop)        :: 70 of 285
            Pseudo Genes (multiple problems)    :: 20 of 285
            ##Genome-Annotation-Data-END##
FEATURES             Location/Qualifiers
     source          1..19
                     /organism="Bacillus anthracis"
                     /mol_type="genomic DNA"
                     /strain="FDAARGOS_694"
                     /isolation_source="synthetic"
                     /culture_collection="FDA:FDAARGOS_694"
                     /db_xref="taxon:1392"
                     /country="USA: MD"
                     /collected_by="U.S. Army Edgewood Chemical Biological
                     Center"
     gene            <1..>19
                     /locus_tag="FOC02_25700"
     CDS             <1..>19
                     /locus_tag="FOC02_25700"
                     /inference="COORDINATES: similar to AA
                     sequence:RefSeq:WP_000849936.1"
                     /note="Derived by automated computational analysis using
                     gene prediction method: Protein Homology."
                     /codon_start=3
```

/transl_table=11
/product="**serine hydrolase**"
/protein_id="QIY15283.1"
/translation="MKTKDQIDNIVKETYRHIDFSGVILVKEKKGIVYEEVFGYANRN
ECINNTLQTRFGIASGCKIFTAIGICQLVENGVITFQTKLKECLSVNFPNFNEDITIH
HLLTHSSGIPDYFDESIMDNFEDLWKQTPMYLLKSLKDFLPLFQNRDMMYAPGSKFHY
NNAGFIILGLIIEEQTGLTFTEYIEKNIFNPIGMNHSGYFSLDRLPRQTALGYIKDEI
SQTWRTNAYSIPIKGGSDGGAFITASDMLKFWEALFNYEIISQEYTKILLTPHIQVNN
NQSYGYGIWIETRENKIF**KYHVMGY**DPGVSFRSAVYPDLGITLVIPSNKGAGAEKLMI
EIEGSF"

ORIGIN

1 **aatatcatgt aatgggata**

//

Bacillus anthracis strain FDAARGOS_694 chromosome
Sequence ID: CP050970.1 **Length: 5251490 Number of Matches: 1**
Range 1: 4915742 to 4915760 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
38.2 bits(19)	2.2	19/19(100%)	0/19(0%)	Plus/Minus

```
CDS: Putative 1       5           G  M  V  H  Y
Query                 1        TATCCCATTACATGATATT  19
                               ||||||||||||||||||
Sbjct                 4915760  TATCCCATTACATGATATT  4915742
CDS:serine hydrolase  301      Y  G  M  V  H  Y  K
```

Bacillus paranthracis strain MN1F chromosome, complete genome

GenBank: CP048687.1

FASTA Graphics

Go to:

```
LOCUS       CP048687                 19 bp    DNA     linear   BCT 17-FEB-2020
DEFINITION  Bacillus paranthracis strain MN1F chromosome, complete genome.
ACCESSION   CP048687 REGION: 5037812..5037830
VERSION     CP048687.1
DBLINK      BioProject: PRJNA605734
            BioSample: SAMN14074934
KEYWORDS    .
SOURCE      Bacillus paranthracis
  ORGANISM  Bacillus paranthracis
```

```
            Bacteria; Firmicutes; Bacilli; Bacillales; Bacillaceae; Bacillus;
            Bacillus cereus group.
REFERENCE   1  (bases 1 to 19)
  AUTHORS   Sousa,L.
  TITLE     Bacillus paranthracis BCCL
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 19)
  AUTHORS   Sousa,L.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-FEB-2020) Genetic Department, UNICAMP, Rua Bertrand
            Russel, s/n, Campinas 6109, Brazil
COMMENT     ##Genome-Assembly-Data-START##
            Assembly Method        :: Velvet v. 1.2.10
            Genome Representation  :: Full
            Expected Final Version :: Yes
            Genome Coverage        :: 80.0x
            Sequencing Technology  :: PacBio
            ##Genome-Assembly-Data-END##
FEATURES             Location/Qualifiers
     source          1..19
                     /organism="Bacillus paranthracis"
                     /mol_type="genomic DNA"
                     /strain="MN1F"
                     /isolation_source="book surface"
                     /db_xref="taxon:2026186"
                     /country="Brazil: Campinas"
                     /lat_lon="22.54 S 47.3 W"
                     /collection_date="Aug-2019"
                     /collected_by="Leandro Pio de Sousa"
ORIGIN
        1 tatcccatta catgatatt
//
```

Bacillus paranthracis strain MN1F chromosome, complete genome

Sequence ID: CP048687.1 **Length: 5687871 Number of Matches: 2**

Range 1: 5037812 to 5037830 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
38.2 bits(19)	2.2	19/19(100%)	0/19(0%)	Plus/Plus

```
Query  1         TATCCCATTACATGATATT  19
                 ||||||||||||||||||
Sbjct  5037812   TATCCCATTACATGATATT  5037830
```

Bacillus paranthracis strain MN1F chromosome, complete genome
GenBank: CP048687.1

FASTA Graphics

Go to:

LOCUS CP048687 16 bp DNA linear BCT **17-FEB-2020**
DEFINITION Bacillus paranthracis strain MN1F chromosome, complete genome.
ACCESSION CP048687 REGION: 430378..430393
VERSION CP048687.1
DBLINK BioProject: PRJNA605734
BioSample: SAMN14074934
KEYWORDS .
SOURCE Bacillus paranthracis
ORGANISM Bacillus paranthracis
Bacteria; Firmicutes; Bacilli; Bacillales; Bacillaceae; Bacillus;
Bacillus cereus group.
REFERENCE 1 (bases 1 to 16)
AUTHORS Sousa,L.
TITLE Bacillus paranthracis BCCL
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 16)
AUTHORS Sousa,L.
TITLE **Direct Submission**
JOURNAL **Submitted (10-FEB-2020) Genetic Department, UNICAMP, Rua Bertrand Russel, s/n, Campinas 6109, Brazil**
COMMENT ##Genome-Assembly-Data-START##
Assembly Method :: Velvet v. 1.2.10
Genome Representation :: Full
Expected Final Version :: Yes
Genome Coverage :: 80.0x
Sequencing Technology :: PacBio
##Genome-Assembly-Data-END##
FEATURES Location/Qualifiers
source 1..16
/organism="Bacillus paranthracis"
/mol_type="genomic DNA"
/strain="MN1F"
/isolation_source="book surface"
/db_xref="taxon:2026186"
/country="Brazil: Campinas"
/lat_lon="22.54 S 47.3 W"
/collection_date="Aug-2019"
/collected_by="Leandro Pio de Sousa"
ORIGIN
1 **aatatcatgt aatggg** //

Range 2: 430378 to 430393 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
32.2 bits(16)	138	16/16(100%)	0/16(0%)	Plus/Minus

```
Query  4       CCCATTACATGATATT  19
               ||||||||||||||||
Sbjct  430393  CCCATTACATGATATT  430378
```

Bacillus paranthracis strain NCCP 15910 chromosome, complete genome

GenBank: CP041981.1

FASTA Graphics

Go to:

LOCUS CP041981 19 bp DNA linear BCT **15-JAN-2020**

DEFINITION Bacillus paranthracis strain NCCP 15910 chromosome, complete genome.

ACCESSION CP041981 REGION: 2708037..2708055

VERSION CP041981.1

DBLINK BioProject: PRJNA556101
BioSample: SAMN12340779

KEYWORDS .

SOURCE Bacillus paranthracis

ORGANISM Bacillus paranthracis
Bacteria; Firmicutes; Bacilli; Bacillales; Bacillaceae; Bacillus; Bacillus cereus group.

REFERENCE 1 (bases 1 to 19)

AUTHORS Yu,W.S., Cheong,H.-M., Choi,Y., Hwang,K.J., Jung,K., Lee,S. and Choi,C.

TITLE **Direct Submission**

JOURNAL **Submitted (23-JUL-2019) Pathogen Resource TF, Center for Infectious Diseases Research, Korea National Institute of Health, Korea Centers for Disease Control & Prevention (KCDC), Osong Health Technology Administration Complex, 187, Osongsaengmyeong 2-ro, Osong-eup, Heungdeok-gu, Cheongju-si, Chungcheongbuk-do 28159, South Korea**

COMMENT The annotation was added by the NCBI Prokaryotic Genome Annotation Pipeline (PGAP). Information about PGAP can be found here: https://www.ncbi.nlm.nih.gov/genome/annotation_prok/

##Genome-Assembly-Data-START##
Assembly Date :: JUL-2017
Assembly Method :: SPAdes v. 3.1
Genome Representation :: Full

```
            Expected Final Version :: Yes
            Genome Coverage        :: 207.0x
            Sequencing Technology  :: IonTorrent; PacBio
            ##Genome-Assembly-Data-END##


            ##Genome-Annotation-Data-START##
            Annotation Provider               :: NCBI
            Annotation Date                   :: 07/24/2019 15:40:42
            Annotation Pipeline               :: NCBI Prokaryotic Genome
                                                 Annotation Pipeline (PGAP)
            Annotation Method                 :: Best-placed reference protein
                                                 set; GeneMarkS-2+
            Annotation Software revision      :: 4.8
            Features Annotated                :: Gene; CDS; rRNA; tRNA; ncRNA;
                                                 repeat_region
            Genes (total)                     :: 5,737
            CDSs (total)                      :: 5,589
            Genes (coding)                    :: 4,975
            CDSs (with protein)               :: 4,975
            Genes (RNA)                       :: 148
            rRNAs                             :: 13, 13, 13 (5S, 16S, 23S)
            complete rRNAs                    :: 13, 13, 13 (5S, 16S, 23S)
            tRNAs                             :: 104
            ncRNAs                            :: 5
            Pseudo Genes (total)              :: 614
            CDSs (without protein)            :: 614
            Pseudo Genes (ambiguous residues) :: 0 of 614
            Pseudo Genes (frameshifted)       :: 497 of 614
            Pseudo Genes (incomplete)         :: 114 of 614
            Pseudo Genes (internal stop)      :: 65 of 614
            Pseudo Genes (multiple problems)  :: 53 of 614
            ##Genome-Annotation-Data-END##
FEATURES             Location/Qualifiers
     source          1..19
                     /organism="Bacillus paranthracis"
                     /mol_type="genomic DNA"
                     /strain="NCCP 15910"
                     /isolation_source="stool"
                     /host="Homo sapiens"
                     /culture_collection="NCCP:15910"
                     /db_xref="taxon:2026186"
                     /country="South Korea: Seoul"
                     /lat_lon="37.547889 N 126.941889 E"
                     /collection_date="2013-05-20"
```

```
                     /collected_by="Korea National Institute of Health"
     gene            <1..>19
                     /locus_tag="FPL02_15370"
                     /pseudo
     CDS             <1..>19
                     /locus_tag="FPL02_15370"
                     /inference="COORDINATES: similar to AA
                     sequence:RefSeq:WP_000849936.1"
                     /note="incomplete; partial on complete genome; missing
                     C-terminus; Derived by automated computational analysis
                     using gene prediction method: Protein Homology."
                     /pseudo
                     /codon_start=3
                     /transl_table=11
                     /product="serine hydrolase"
ORIGIN
        1 aatatcatgt aatgggata
//
```

Bacillus paranthracis strain NCCP 15910 chromosome, complete genome
Sequence ID: CP041981.1 **Length: 5208566 Number of Matches: 2**
Range 1: 2708037 to 2708055 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
38.2 bits(19)	2.2	19/19(100%)	0/19(0%)	Plus/Minus

```
CDS: Putative 1   5           G  M  V  H  Y
Query             1        TATCCCATTACATGATATT  19
                           ||||||||||||||||||
Sbjct             2708055  TATCCCATTACATGATATT  2708037
CDS: serine hydr  301      Y  G  M  V  H  Y  K
```

Bacillus paranthracis strain NCCP 15910 chromosome, complete genome

GenBank: CP041981.1

FASTA Graphics

Go to:

LOCUS CP041981 16 bp DNA linear BCT **15-JAN-2020**

DEFINITION Bacillus paranthracis strain NCCP 15910 chromosome, complete genome.

ACCESSION CP041981 REGION: 1645460..1645475

VERSION CP041981.1

DBLINK BioProject: PRJNA556101
BioSample: SAMN12340779

KEYWORDS .

SOURCE Bacillus paranthracis

ORGANISM Bacillus paranthracis
Bacteria; Firmicutes; Bacilli; Bacillales; Bacillaceae; Bacillus; Bacillus cereus group.

REFERENCE 1 (bases 1 to 16)

AUTHORS Yu,W.S., Cheong,H.-M., Choi,Y., Hwang,K.J., Jung,K., Lee,S. and Choi,C.

TITLE Direct Submission

JOURNAL **Submitted (23-JUL-2019) Pathogen Resource TF, Center for Infectious Diseases Research, Korea National Institute of Health, Korea Centers for Disease Control & Prevention (KCDC), Osong Health Technology Administration Complex, 187, Osongsaengmyeong 2-ro, Osong-eup, Heungdeok-gu, Cheongju-si, Chungcheongbuk-do 28159, South Korea**

COMMENT The annotation was added by the NCBI Prokaryotic Genome Annotation Pipeline (PGAP). Information about PGAP can be found here: https://www.ncbi.nlm.nih.gov/genome/annotation_prok/

##Genome-Assembly-Data-START##
Assembly Date :: JUL-2017
Assembly Method :: SPAdes v. 3.1
Genome Representation :: Full
Expected Final Version :: Yes
Genome Coverage :: 207.0x
Sequencing Technology :: IonTorrent; PacBio
##Genome-Assembly-Data-END##

##Genome-Annotation-Data-START##
Annotation Provider :: NCBI
Annotation Date :: 07/24/2019 15:40:42
Annotation Pipeline :: NCBI Prokaryotic Genome Annotation Pipeline (PGAP)

```
            Annotation Method                 :: Best-placed reference protein
                                                 set; GeneMarkS-2+
            Annotation Software revision      :: 4.8
            Features Annotated                :: Gene; CDS; rRNA; tRNA; ncRNA;
                                                 repeat_region
            Genes (total)                     :: 5,737
            CDSs (total)                      :: 5,589
            Genes (coding)                    :: 4,975
            CDSs (with protein)               :: 4,975
            Genes (RNA)                       :: 148
            rRNAs                             :: 13, 13, 13 (5S, 16S, 23S)
            complete rRNAs                    :: 13, 13, 13 (5S, 16S, 23S)
            tRNAs                             :: 104
            ncRNAs                            :: 5
            Pseudo Genes (total)              :: 614
            CDSs (without protein)            :: 614
            Pseudo Genes (ambiguous residues) :: 0 of 614
            Pseudo Genes (frameshifted)       :: 497 of 614
            Pseudo Genes (incomplete)         :: 114 of 614
            Pseudo Genes (internal stop)      :: 65 of 614
            Pseudo Genes (multiple problems)  :: 53 of 614
            ##Genome-Annotation-Data-END##

FEATURES             Location/Qualifiers
     source          1..16
                     /organism="Bacillus paranthracis"
                     /mol_type="genomic DNA"
                     /strain="NCCP 15910"
                     /isolation_source="stool"
                     /host="Homo sapiens"
                     /culture_collection="NCCP:15910"
                     /db_xref="taxon:2026186"
                     /country="South Korea: Seoul"
                     /lat_lon="37.547889 N 126.941889 E"
                     /collection_date="2013-05-20"
                     /collected_by="Korea National Institute of Health"
     gene            <1..>16
                     /gene="cysH"
                     /locus_tag="FPL02_09960"
     CDS             <1..>16
                     /gene="cysH"
                     /locus_tag="FPL02_09960"
                     /EC_number="1.8.4.8"
                     /inference="COORDINATES: similar to AA
                     sequence:RefSeq:WP_000958995.1"
```

/note="Derived by automated computational analysis using
gene prediction method: Protein Homology."
/codon_start=2
/transl_table=11
/product="**phosphoadenylyl-sulfate reductase**"
/protein_id="QHH84108.1"
/translation="MLTYETWEENDVLFSEEDETKGALSVLSWAYKEYKSEIVYACSF
GVEGMVLLHLINQVNPSAKVVFLDTNVHFQETYELIQKVRERFPSLNIIEKQPKLTLD
EQAKLHGNKLWESNPNLCCKIRKISPLEESLANEKAWISGLRREQSETRKHTKFINQD
HRFQSIKVCPLIHWTWKEVWRYVYKHSLPYN**PLHDI**GYPSIGCEKCTLPVGEGGDSRD
GRWAGKVKTECGLHYQ"

ORIGIN

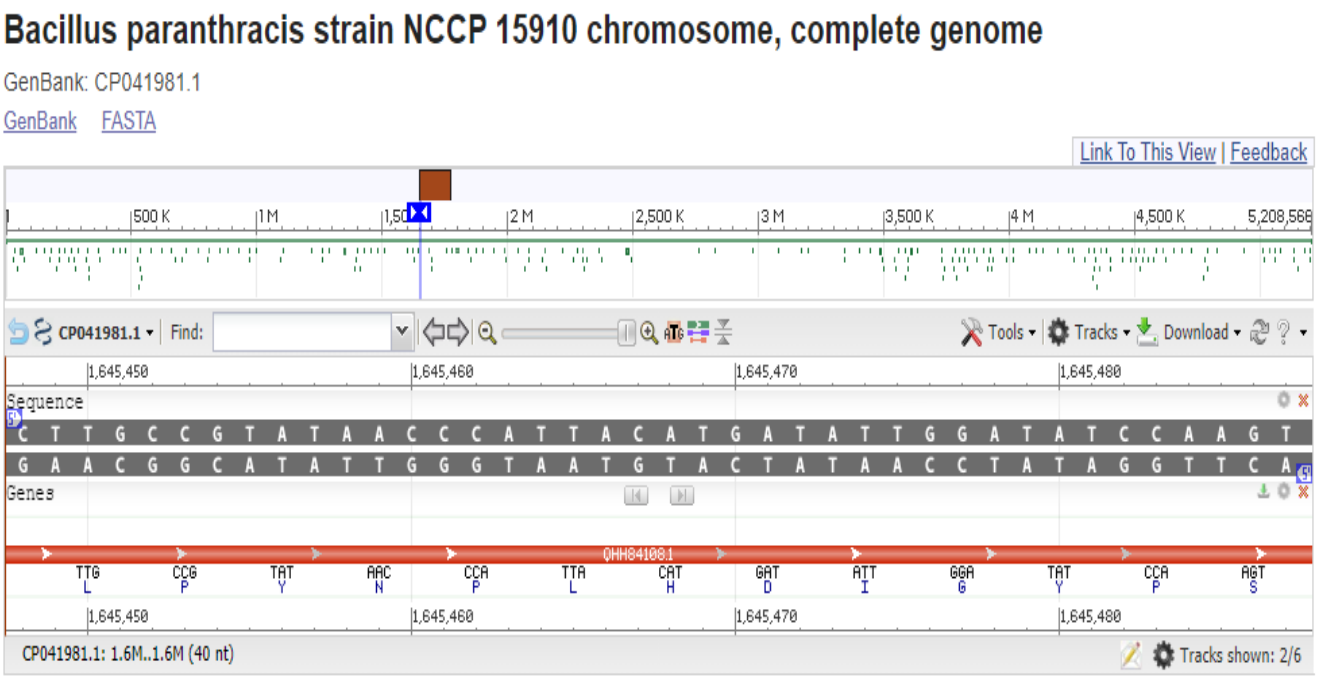
1 **cccattacat gatatt**

//

Range 2: 1645460 to 1645475 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
32.2 bits(16)	136	16/16(100%)	0/16(0%)	Plus/Plus

```
CDS: Putative 1        1          P  L  H  D  I
Query                  4          CCCATTACATGATATT  19
                                  ||||||||||||||||
Sbjct                  1645460    CCCATTACATGATATT  1645475
CDS:phosphoadenylyl-   192          P  L  H  D  I
```



Phosphoadenylyl-sulfate reductase (thioredoxin)

Phosphoadenylyl-sulfate reductase (thioredoxin) (EC 1.8.4.8) is an enzyme that catalyzes the chemical reaction

adenosine 3',5'-bisphosphate + sulfite + thioredoxin disulfide ⇌ 3'-phosphoadenylyl sulfate + thioredoxin

The 3 substrates of this enzyme are adenosine 3', 5'-bisphosphate, sulfite, and thioredoxin disulfide, whereas its two products are ***3'-phosphoadenylyl sulfate*** and ***thioredoxin***. This enzyme belongs to the family of *oxidoreductases*, specifically those acting on a *sulfur group of donors* with a disulfide as acceptor. This enzyme participates in **sulfur metabolism** (https://en.wikipedia.org/wiki/Phosphoadenylyl-sulfate_reductase_(thioredoxin)).

3'-Phosphoadenosine-5'-phosphosulfate (**PAPS**) is a derivative of *adenosine monophosphate* that is phosphorylated at the 3' position and has a sulfate group attached to the 5' phosphate. It is the most common coenzyme in *sulfotransferase reactions*. It is endogenously synthesized by organisms via the phosphorylation of *adenosine 5'-phosphosulfate* (APS), an intermediary metabolite. In humans such reaction is performed by bifunctional *3'-phosphoadenosine 5'-phosphosulfate synthases* (PAPSS1 and PAPSS2) using ATP as the phosphate donor (https://en.wikipedia.org/wiki/3%27-Phosphoadenosine-5%27-phosphosulfate).

Thioredoxin (Trx) is a class of *small redox proteins* known to be present in all organisms. *Thioredoxin* plays a role in many important biological processes, including *redox signaling*. These proteins have also recently been found to play a role in *cell-to-cell communication* (https://en.wikipedia.org/wiki/Thioredoxin). The thioredoxins are kept in the reduced state by the *flavoenzyme thioredoxin reductase*, in a *NADPH-dependent reaction*. The primary function of Thioredoxin is the *reduction of oxidized cysteine residues* and the *cleavage of disulfide bonds*.

Thioredoxin reductases (**TR**, **TrxR**) (EC 1.8.1.9) are the only known enzymes to reduce ***thioredoxin*** (Trx). *Two classes of thioredoxin reductase* have been identified: one class in *bacteria* and some *eukaryotes* and *one in animals*. In ***bacteria TrxR*** also catalyzes the *reduction of glutaredoxin like proteins known as NrdH*. Both classes are ***flavoproteins*** which function as *homodimers*. Each *monomer* contains a *FAD prosthetic group*, a *NADPH binding domain*, and an *active site* containing a *redox-active disulfide bond* (https://en.wikipedia.org/wiki/Thioredoxin_reductase). The activity of this enzyme is essential for cell growth and survival, it is a good target for **anti-tumor therapy**. Furthermore, **the enzyme is upregulated in several types of cancer**. Here has recently been some research to show that low *molecular weight thioredoxin reductase* could be a target for novel antibiotics (such as *auranofin* or *Ebselen*) and could be used for **antibiotic resistant bacteria**.

Bacillus paranthracis strain NCCP 14796 chromosome, complete genome

GenBank: CP041750.1

FASTA Graphics

Go to:

LOCUS CP041750 19 bp DNA linear BCT **14-JAN-2020**

DEFINITION Bacillus paranthracis strain NCCP 14796 chromosome, complete genome.

ACCESSION CP041750 REGION: 2498260..2498278

VERSION CP041750.1

DBLINK BioProject: PRJNA555180
BioSample: SAMN12302793

KEYWORDS .

SOURCE Bacillus paranthracis

ORGANISM Bacillus paranthracis
Bacteria; Firmicutes; Bacilli; Bacillales; Bacillaceae; Bacillus; Bacillus cereus group.

REFERENCE 1 (bases 1 to 19)

AUTHORS Yu,W.S., Cheong,H.-M., Choi,Y., Hwang,K.J., Kim,D.-W. and Yun,M.-R.

TITLE Direct Submission

JOURNAL **Submitted (18-JUL-2019) Pathogen Resource TF, Center for Infectious Diseases Research, Korea National Institute of Health, Korea Centers for Disease Control and Prevention, #200 Osongsaengmyeong 2-ro, Osong-eup, Heungdeok-gu, Cheongju-si, Chungcheongbuk-do 28160, South Korea**

COMMENT The annotation was added by the NCBI Prokaryotic Genome Annotation Pipeline (PGAP). Information about PGAP can be found here: https://www.ncbi.nlm.nih.gov/genome/annotation_prok/

```
##Genome-Assembly-Data-START##
Assembly Method        :: SPAdes v. 3.1
Genome Representation  :: Full
Expected Final Version :: Yes
Genome Coverage        :: 241.0x
Sequencing Technology  :: PacBio; IonTorrent
##Genome-Assembly-Data-END##

##Genome-Annotation-Data-START##
Annotation Provider               :: NCBI
Annotation Date                   :: 07/19/2019 12:30:27
Annotation Pipeline               :: NCBI Prokaryotic Genome
                                     Annotation Pipeline (PGAP)
Annotation Method                 :: Best-placed reference protein
                                     set; GeneMarkS-2+
Annotation Software revision      :: 4.8
```

Features Annotated :: Gene; CDS; rRNA; tRNA; ncRNA; repeat_region
Genes (total) :: 5,401
CDSs (total) :: 5,254
Genes (coding) :: 4,717
CDSs (with protein) :: 4,717
Genes (RNA) :: 147
rRNAs :: 14, 14, 14 (5S, 16S, 23S)
complete rRNAs :: 14, 14, 14 (5S, 16S, 23S)
tRNAs :: 100
ncRNAs :: 5
Pseudo Genes (total) :: 537
CDSs (without protein) :: 537
Pseudo Genes (ambiguous residues) :: 0 of 537
Pseudo Genes (frameshifted) :: 404 of 537
Pseudo Genes (incomplete) :: 119 of 537
Pseudo Genes (internal stop) :: 53 of 537
Pseudo Genes (multiple problems) :: 35 of 537
##Genome-Annotation-Data-END##

FEATURES Location/Qualifiers
source 1..19
/organism="Bacillus paranthracis"
/mol_type="genomic DNA"
/strain="NCCP 14796"
/isolation_source="stool"
/host="**Homo sapiens**"
/db_xref="taxon:2026186"
/country="South Korea: Ulsan"
/lat_lon="35.56 N 129.12 E"
/collection_date="01-Jan-2008"
/collected_by="Korea National Institute of Health"
gene <1..>19
/locus_tag="FOW48_12885"
CDS <1..>19
/locus_tag="FOW48_12885"
/inference="COORDINATES: similar to AA sequence:RefSeq:WP_000849936.1"
/note="Derived by automated computational analysis using gene prediction method: Protein Homology."
/codon_start=3
/transl_table=11
/product="serine hydrolase"
/protein_id="QHG35229.1"

/translation="MKTKNQIDNIVKEIYRHIDFSGVILVKEKKGIVYEEAFGYANRN
EYINNTLQTRFGIASGCKIFTAIGICQLVERGVITFHTKLKDCLSIKFSNFNEDITIH
HLLTHSSGIPDYFDESILDNFEDLWKETPMYLLKNLKDFLPLFQNRDMMFAPGSKFHY
NNAGFIILGLIIEEQTGLKFTEYIETNIFKLIGMNDSGYFSLDRLPKQTALGYIKDEI
SQTWRTNAYSIPIKGGSDGGAFITAPDMMKFWEALFNYEIISQEYTKILLTPHIQVNS
YQSYGYGIWIEIRENKIF**KYHVMGY**DPGVSFRSAVYPDLGITLVIPSNKGAGPEKLMI
EIEGAF"

ORIGIN

1 **aatatcatgt aatgggata**

//

Bacillus paranthracis strain NCCP 14796 chromosome, complete genome
Sequence ID: CP041750.1 **Length: 5017580 Number of Matches: 2**
Range 1: 2498260 to 2498278 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
38.2 bits(19)	2.2	19/19(100%)	0/19(0%)	Plus/Minus

```
CDS: Putative 1       5           G  M  V  H  Y
Query                 1        TATCCCATTACATGATATT  19
                               ||||||||||||||||||
Sbjct                 2498278  TATCCCATTACATGATATT  2498260
CDS:serine hydrolase  301      Y  G  M  V  H  Y  K
```

Bacillus paranthracis strain NCCP 14796 chromosome, complete genome

GenBank: CP041750.1

FASTA Graphics

Go to:

```
LOCUS       CP041750                 16 bp    DNA     linear   BCT 14-JAN-2020
DEFINITION  Bacillus paranthracis strain NCCP 14796 chromosome, complete
            genome.
ACCESSION   CP041750 REGION: 1427940..1427955
VERSION     CP041750.1
DBLINK      BioProject: PRJNA555180
            BioSample: SAMN12302793
KEYWORDS    .
SOURCE      Bacillus paranthracis
  ORGANISM  Bacillus paranthracis
            Bacteria; Firmicutes; Bacilli; Bacillales; Bacillaceae; Bacillus;
            Bacillus cereus group.
REFERENCE   1  (bases 1 to 16)
  AUTHORS   Yu,W.S., Cheong,H.-M., Choi,Y., Hwang,K.J., Kim,D.-W. and Yun,M.-R.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-JUL-2019) Pathogen Resource TF, Center for Infectious
            Diseases Research, Korea National Institute of Health, Korea
```

Centers for Disease Control and Prevention, #200 Osongsaengmyeong 2-ro, Osong-eup, Heungdeok-gu, Cheongju-si, Chungcheongbuk-do 28160, South Korea

COMMENT The annotation was added by the NCBI Prokaryotic Genome Annotation Pipeline (PGAP). Information about PGAP can be found here: https://www.ncbi.nlm.nih.gov/genome/annotation_prok/

```
##Genome-Assembly-Data-START##
Assembly Method        :: SPAdes v. 3.1
Genome Representation  :: Full
Expected Final Version :: Yes
Genome Coverage        :: 241.0x
Sequencing Technology  :: PacBio; IonTorrent
##Genome-Assembly-Data-END##


##Genome-Annotation-Data-START##
Annotation Provider               :: NCBI
Annotation Date                   :: 07/19/2019 12:30:27
Annotation Pipeline               :: NCBI Prokaryotic Genome
                                     Annotation Pipeline (PGAP)
Annotation Method                 :: Best-placed reference protein
                                     set; GeneMarkS-2+
Annotation Software revision      :: 4.8
Features Annotated                :: Gene; CDS; rRNA; tRNA; ncRNA;
                                     repeat_region
Genes (total)                     :: 5,401
CDSs (total)                      :: 5,254
Genes (coding)                    :: 4,717
CDSs (with protein)               :: 4,717
Genes (RNA)                       :: 147
rRNAs                             :: 14, 14, 14 (5S, 16S, 23S)
complete rRNAs                    :: 14, 14, 14 (5S, 16S, 23S)
tRNAs                             :: 100
ncRNAs                            :: 5
Pseudo Genes (total)              :: 537
CDSs (without protein)            :: 537
Pseudo Genes (ambiguous residues) :: 0 of 537
Pseudo Genes (frameshifted)       :: 404 of 537
Pseudo Genes (incomplete)         :: 119 of 537
Pseudo Genes (internal stop)      :: 53 of 537
Pseudo Genes (multiple problems)  :: 35 of 537
##Genome-Annotation-Data-END##
FEATURES             Location/Qualifiers
     source          1..16
```

/organism="Bacillus paranthracis"
/mol_type="genomic DNA"
/strain="NCCP 14796"
/isolation_source="stool"
/host="Homo sapiens"
/db_xref="taxon:2026186"
/country="South Korea: Ulsan"
/lat_lon="35.56 N 129.12 E"
/collection_date="01-Jan-2008"
/collected_by="Korea National Institute of Health"

gene <1..>16
/gene="cysH"
/locus_tag="FOW48_07370"

CDS <1..>16
/gene="cysH"
/locus_tag="FOW48_07370"
/EC_number="1.8.4.8"
/inference="COORDINATES: similar to AA sequence:RefSeq:WP_000958995.1"
/note="Derived by automated computational analysis using gene prediction method: Protein Homology."
/codon_start=2
/transl_table=11
/product="**phosphoadenylyl-sulfate reductase**"
/protein_id="QHG34288.1"
/translation="MLTYETWEENDVSFSEEDETKGALSVLSWAYKEYENEIVYACSF
GVEGMVLLHLINQVNPSAKVVFLDTNVHFQETYELIQKVRERFPSLNIIEKQPKLTLD
EQAKLHGNKLWESNPNLCCKIRKILPLEESLANEKAWISGLRREQSETRKHTKFINQD
HRFQSIKVCPLIHWTWKEVWRYVYKHSLPYN**PLHDI**GYPSIGCEKCTLPVGEGGDSRD
GRWAGKVKTECGLHYQ"

ORIGIN

1 **cccattacat gatatt**

//

Range 2: 1427940 to 1427955 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
32.2 bits(16)	136	16/16(100%)	0/16(0%)	Plus/Plus

```
CDS: Putative 1       1          P  L  H  D  I
Query                 4        CCCATTACATGATATT  19
                               ||||||||||||||||
Sbjct                 1427940  CCCATTACATGATATT  1427955
CDS:phosphoadenylyl-  192        P  L  H  D  I
```

Bacillus anthracis str. BF1 chromosome, complete genome

GenBank: CP047131.1

FASTA Graphics

Go to:

LOCUS CP047131 19 bp DNA linear BCT **08-JAN-2020**
DEFINITION Bacillus anthracis str. BF1 chromosome, complete genome.
ACCESSION CP047131 REGION: 2458681..2458699
VERSION CP047131.1
DBLINK BioProject: PRJNA171093
BioSample: SAMN02469407

KEYWORDS .
SOURCE Bacillus anthracis str. BF1
ORGANISM Bacillus anthracis str. BF1
Bacteria; Firmicutes; Bacilli; Bacillales; Bacillaceae; Bacillus;
Bacillus cereus group.
REFERENCE 1 (bases 1 to 19)
AUTHORS Antwerpen,M., Proenca,D.N., Ruckert,C., Licht,K., Kalinowski,J.,
Hanczaruk,M., Tiemann,C. and Grass,G.
TITLE Draft genome sequence of Bacillus anthracis BF-1, isolated from
Bavarian cattle
JOURNAL J Bacteriol 194 (22), 6360-6361 (2012)
PUBMED 23105087
REFERENCE 2 (bases 1 to 19)
AUTHORS Antwerpen,M., Proenca,D.N., Ruckert,C., Licht,K., Kalinowski,J.,
Hanczaruk,M., Tiemann,C. and Grass,G.
TITLE Direct Submission
JOURNAL **Submitted (10-AUG-2012) Bundeswehr Institute of Microbiology,
Neuherbergstrasse 11, Munich 80937, Germany**
REFERENCE 3 (bases 1 to 19)
AUTHORS Antwerpen,M.H., Walter,M.C. and Grass,G.
TITLE Direct Submission
JOURNAL **Submitted (20-DEC-2019) Microbial Genomics, Bundeswehr, Institute
of Microbiology, Neuherbergstr. 11, Munich 80937, Germany**
COMMENT Bacteria and source DNA available from Bundeswehr Institute of
Microbiology.
The annotation was added by the NCBI Prokaryotic Genome Annotation
Pipeline (PGAP). Information about PGAP can be found here:
https://www.ncbi.nlm.nih.gov/genome/annotation_prok/

##Genome-Assembly-Data-START##
Assembly Date :: 2019
Assembly Method :: SPAdes v. 3.13.1; Unicycler v. 0.4.8
Genome Representation :: Full
Expected Final Version :: Yes

```
Genome Coverage        :: 287.0x
Sequencing Technology  :: Illumina MiSeq; Oxford Nanopore GridION
##Genome-Assembly-Data-END##

##Genome-Annotation-Data-START##
Annotation Provider               :: NCBI
Annotation Date                   :: 12/31/2019 11:49:50
Annotation Pipeline               :: NCBI Prokaryotic Genome
                                     Annotation Pipeline (PGAP)
Annotation Method                 :: Best-placed reference protein
                                     set; GeneMarkS-2+
Annotation Software revision      :: 4.10
Features Annotated                :: Gene; CDS; rRNA; tRNA; ncRNA;
                                     repeat_region
Genes (total)                     :: 5,655
CDSs (total)                      :: 5,524
Genes (coding)                    :: 5,248
CDSs (with protein)               :: 5,248
Genes (RNA)                       :: 131
rRNAs                             :: 10, 10, 10 (5S, 16S, 23S)
complete rRNAs                    :: 10, 10, 10 (5S, 16S, 23S)
tRNAs                             :: 96
ncRNAs                            :: 5
Pseudo Genes (total)              :: 276
CDSs (without protein)            :: 276
Pseudo Genes (ambiguous residues) :: 0 of 276
Pseudo Genes (frameshifted)       :: 186 of 276
Pseudo Genes (incomplete)         :: 41 of 276
Pseudo Genes (internal stop)      :: 74 of 276
Pseudo Genes (multiple problems)  :: 21 of 276
##Genome-Annotation-Data-END##

FEATURES             Location/Qualifiers
     source          1..19
                     /organism="Bacillus anthracis str. BF1"
                     /mol_type="genomic DNA"
                     /strain="BF1"
                     /isolation_source="cow carcass"
                     /host="cow"
                     /db_xref="taxon:1213182"
                     /country="Germany"
                     /collection_date="Jul-2009"
     gene            <1..>19
                     /locus_tag="BABF1_012635"
     CDS             <1..>19
```

/locus_tag="BABF1_012635"
/inference="COORDINATES: similar to AA
sequence:RefSeq:WP_000849936.1"
/note="Derived by automated computational analysis using
gene prediction method: Protein Homology."
/codon_start=3
/transl_table=11
/product="**serine hydrolase**"
/protein_id="QHD22348.1"
/translation="MKTKDQIDNIVKETYRHIDFSGVILVKEKKGIVYEEVFGYANRN
ECINNTLQTRFGIASGCKIFTAIGICQLVENGVITFQTKLKECLSVNFPNFNEDITIH
HLLTHSSGIPDYFDESIMDNFEDLWKQTPMYLLKSLKDFLPLFQNRDMMYAPGSKFHY
NNAGFIILGLIIEEQTGLTFTEYIEKNIFNPIGMNHSGYFSLDRLPRQTALGYIKDEI
SQTWRTNAYSIPIKGGSDGGAFITASDMLKFWEALFNYEIISQEYTKILLTPHIQVNN
NQSYGYGIWIETRENKIF**KYHVMGY**DPGVSFRSAVYPDLGITLVIPSNKGAGAEKLMI
EIEGSF"

ORIGIN

1 **aatatcatgt aatgggata**

//

Bacillus anthracis str. BF1 chromosome, complete genome
Sequence ID: CP047131.1 **Length: 5213532 Number of Matches: 1**
Range 1: 2458681 to 2458699 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
38.2 bits(19)	2.2	19/19(100%)	0/19(0%)	Plus/Minus

```
CDS: Putative 1       5           G  M  V  H  Y
Query                 1        TATCCCATTACATGATATT  19
                               ||||||||||||||||||
Sbjct                 2458699  TATCCCATTACATGATATT  2458681
CDS:serine hydrolase  301      Y  G  M  V  H  Y  K
```

Bacillus anthracis strain FDAARGOS_704 chromosome, complete genome

GenBank: CP047111.1

FASTA Graphics

Go to:

```
LOCUS       CP047111                 19 bp    DNA     linear   BCT 28-DEC-2019
DEFINITION  Bacillus anthracis strain FDAARGOS_704 chromosome, complete genome.
ACCESSION   CP047111 REGION: 400614..400632
VERSION     CP047111.1
DBLINK      BioProject: PRJNA231221
            BioSample: SAMN11056419
KEYWORDS    .
SOURCE      Bacillus anthracis
```

ORGANISM Bacillus anthracis
Bacteria; Firmicutes; Bacilli; Bacillales; Bacillaceae; Bacillus; Bacillus cereus group.
REFERENCE 1 (bases 1 to 19)
AUTHORS Sozhamannan,S., Rosenzweig,N., Tallon,L., Sadzewicz,L., Vavikolanu,K., Mehta,A., Aluvathingal,J., Nadendla,S., Yan,Y. and Sichtig,H.
TITLE FDA dAtabase for Regulatory Grade micrObial Sequences (FDA-ARGOS): Supporting development and validation of Infectious Disease Dx tests
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 19)
AUTHORS Sozhamannan,S., Rosenzweig,N., Tallon,L., Sadzewicz,L., Vavikolanu,K., Mehta,A., Aluvathingal,J., Nadendla,S., Yan,Y. and Sichtig,H.
TITLE Direct Submission
JOURNAL **Submitted (21-NOV-2019) Center for Devices and Radiological Health, US Food and Drug Administration, 10903 New Hampshire Avenue, Silver Spring, MD 20993-0002, USA**
COMMENT The annotation was added by the NCBI Prokaryotic Genome Annotation Pipeline (PGAP). Information about PGAP can be found here: https://www.ncbi.nlm.nih.gov/genome/annotation_prok/

##Genome-Assembly-Data-START##
Assembly Method :: SMRT v. 2.3.0, HGAP v. 3, SPAdes v. 3.6.0
Genome Coverage :: 1225.21x
Sequencing Technology :: Pacbio; Illumina
##Genome-Assembly-Data-END##
##Genome-Annotation-Data-START##

Annotation Provider	:: NCBI
Annotation Date	:: 12/20/2019 16:14:56
Annotation Pipeline	:: NCBI Prokaryotic Genome Annotation Pipeline (PGAP)
Annotation Method	:: Best-placed reference protein set; GeneMarkS-2+
Annotation Software revision	:: 4.10
Features Annotated	:: Gene; CDS; rRNA; tRNA; ncRNA; repeat_region
Genes (total)	:: 5,416
CDSs (total)	:: 5,283
Genes (coding)	:: 5,008
CDSs (with protein)	:: 5,008
Genes (RNA)	:: 133
rRNAs	:: 11, 11, 11 (5S, 16S, 23S)

```
            complete rRNAs                    :: 11, 11, 11 (5S, 16S, 23S)
            tRNAs                             :: 95
            ncRNAs                            :: 5
            Pseudo Genes (total)              :: 275
            CDSs (without protein)            :: 275
            Pseudo Genes (ambiguous residues) :: 0 of 275
            Pseudo Genes (frameshifted)       :: 194 of 275
            Pseudo Genes (incomplete)         :: 35 of 275
            Pseudo Genes (internal stop)      :: 65 of 275
            Pseudo Genes (multiple problems)  :: 16 of 275
            ##Genome-Annotation-Data-END##
FEATURES             Location/Qualifiers
     source          1..19
                     /organism="Bacillus anthracis"
                     /mol_type="genomic DNA"
                     /strain="FDAARGOS_704"
                     /culture_collection="FDA:FDAARGOS_704"
                     /db_xref="taxon:1392"
                     /country="USA: MD"
                     /collected_by="U.S. Army Edgewood Chemical Biological
                     Center"
                     /note="ECBC strain; genetically modified strain"
     gene            complement(<1..>19)
                     /locus_tag="FOC12_02160"
     CDS             complement(<1..>19)
                     /locus_tag="FOC12_02160"
                     /inference="COORDINATES: similar to AA
                     sequence:RefSeq:WP_000849936.1"
                     /note="Derived by automated computational analysis using
                     gene prediction method: Protein Homology."
                     /codon_start=3
                     /transl_table=11
                     /product="serine hydrolase"
                     /protein_id="QHA64246.1"
                     /translation="MKTKDQIDNIVKETYRHIDFSGVILVKEKKGIVYEEVFGYANRN
                     ECINNTLQTRFGIASGCKIFTAIGICQLVENGVITFQTKLKECLSVNFPNFNEDITIH
                     HLLTHSSGIPDYFDESIMDNFEDLWKQTPMYLLKSLKDFLPLFQNRDMMYAPGSKFHY
                     NNAGFIILGLIIEEQTGLTFTEYIEKNIFNPIGMNHSGYFSLDRLPRQTALGYIKDEI
                     SQTWRTNAYSIPIKGGSDGGAFITASDMLKFWEALFNYEIISQEYTKILLTPHIQVNN
                     NQSYGYGIWIETRENKIFKYHVMGYDPGVSFRSAVYPDLGITLVIPSNKGAGAEKLMI
                     EIEGSF"
ORIGIN
        1 tatcccatta catgatatt
//
```

Bacillus anthracis strain FDAARGOS_704 chromosome, complete genome
Sequence ID: CP047111.1 **Length: 5227450 Number of Matches: 1**
Range 1: 400614 to 400632 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
38.2 bits(19)	2.2	19/19(100%)	0/19(0%)	Plus/Plus

```
CDS: Putative 1       5          G  M  V  H  Y
Query                 1       TATCCCATTACATGATATT  19
                              ||||||||||||||||||
Sbjct                 400614  TATCCCATTACATGATATT  400632
CDS:serine hydrolase  301     Y  G  M  V  H  Y  K
```

Bacillus anthracis strain FDAARGOS_705 chromosome, complete genome

GenBank: CP047107.1

FASTA Graphics

Go to:

```
LOCUS       CP047107                  19 bp    DNA     linear   BCT 28-DEC-2019
DEFINITION  Bacillus anthracis strain FDAARGOS_705 chromosome, complete genome.
ACCESSION   CP047107 REGION: 4514812..4514830
VERSION     CP047107.1
DBLINK      BioProject: PRJNA231221
            BioSample: SAMN11056420
KEYWORDS    .
SOURCE      Bacillus anthracis
  ORGANISM  Bacillus anthracis
            Bacteria; Firmicutes; Bacilli; Bacillales; Bacillaceae; Bacillus;
            Bacillus cereus group.
REFERENCE   1  (bases 1 to 19)
  AUTHORS   Sozhamannan,S., Rosenzweig,N., Tallon,L., Sadzewicz,L.,
            Vavikolanu,K., Mehta,A., Aluvathingal,J., Nadendla,S., Yan,Y. and
            Sichtig,H.
  TITLE     FDA dAtabase for Regulatory Grade micrObial Sequences (FDA-ARGOS):
            Supporting development and validation of Infectious Disease Dx
            tests
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 19)
  AUTHORS   Sozhamannan,S., Rosenzweig,N., Tallon,L., Sadzewicz,L.,
            Vavikolanu,K., Mehta,A., Aluvathingal,J., Nadendla,S., Yan,Y. and
            Sichtig,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-NOV-2019) Center for Devices and Radiological Health,
            US Food and Drug Administration, 10903 New Hampshire Avenue, Silver
            Spring, MD 20993-0002, USA
```

```
COMMENT     The annotation was added by the NCBI Prokaryotic Genome Annotation
            Pipeline (PGAP). Information about PGAP can be found here:
            https://www.ncbi.nlm.nih.gov/genome/annotation_prok/
            ##Genome-Assembly-Data-START##
            Assembly Method       :: canu v. 1.4, SPAdes v. 3.6.0
            Genome Coverage       :: 1067.97x
            Sequencing Technology :: Pacbio; Illumina
            ##Genome-Assembly-Data-END##

            ##Genome-Annotation-Data-START##
            Annotation Provider               :: NCBI
            Annotation Date                   :: 12/20/2019 16:10:42
            Annotation Pipeline               :: NCBI Prokaryotic Genome
                                                 Annotation Pipeline (PGAP)
            Annotation Method                 :: Best-placed reference protein
                                                 set; GeneMarkS-2+
            Annotation Software revision      :: 4.10
            Features Annotated                :: Gene; CDS; rRNA; tRNA; ncRNA;
                                                 repeat_region
            Genes (total)                     :: 5,421
            CDSs (total)                      :: 5,288
            Genes (coding)                    :: 5,013
            CDSs (with protein)               :: 5,013
            Genes (RNA)                       :: 133
            rRNAs                             :: 11, 11, 11 (5S, 16S, 23S)
            complete rRNAs                    :: 11, 11, 11 (5S, 16S, 23S)
            tRNAs                             :: 95
            ncRNAs                            :: 5
            Pseudo Genes (total)              :: 275
            CDSs (without protein)            :: 275
            Pseudo Genes (ambiguous residues) :: 0 of 275
            Pseudo Genes (frameshifted)       :: 194 of 275
            Pseudo Genes (incomplete)         :: 35 of 275
            Pseudo Genes (internal stop)      :: 65 of 275
            Pseudo Genes (multiple problems)  :: 16 of 275
            ##Genome-Annotation-Data-END##

FEATURES             Location/Qualifiers
     source          1..19
                     /organism="Bacillus anthracis"
                     /mol_type="genomic DNA"
                     /strain="FDAARGOS_705"
                     /culture_collection="FDA:FDAARGOS_705"
                     /db_xref="taxon:1392"
                     /country="USA: MD"
```

/collected_by="U.S. Army Edgewood Chemical Biological Center"
/note="ECBC strain; genetically modified strain"

gene complement(<1..>19)
/locus_tag="FOC13_23510"

CDS complement(<1..>19)
/locus_tag="FOC13_23510"
/inference="COORDINATES: similar to AA sequence:RefSeq:WP_000849936.1"
/note="Derived by automated computational analysis using gene prediction method: Protein Homology."
/codon_start=3
/transl_table=11
/product="**serine hydrolase**"
/protein_id="QHA63078.1"
/translation="MKTKDQIDNIVKETYRHIDFSGVILVKEKKGIVYEEVFGYANRN
ECINNTLQTRFGIASGCKIFTAIGICQLVENGVITFQTKLKECLSVNFPNFNEDITIH
HLLTHSSGIPDYFDESIMDNFEDLWKQTPMYLLKSLKDFLPLFQNRDMMYAPGSKFHY
NNAGFIILGLIIEEQTGLTFTEYIEKNIFNPIGMNHSGYFSLDRLPRQTALGYIKDEI
SQTWRTNAYSIPIKGGSDGGAFITASDMLKFWEALFNYEIISQEYTKILLTPHIQVNN
NQSYGYGIWIETRENKIF**KYHVMGY**DPGVSFRSAVYPDLGITLVIPSNKGAGAEKLMI
EIEGSF"

ORIGIN

1 **tatcccatta catgatatt**

//

Bacillus anthracis strain FDAARGOS_705 chromosome, complete genome
Sequence ID: CP047107.1**Length: 5227452Number of Matches: 1**
Range 1: 4514812 to 4514830 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
38.2 bits(19)	2.2	19/19(100%)	0/19(0%)	Plus/Plus

```
CDS: Putative 1       5           G  M  V  H  Y
Query                 1        TATCCCATTACATGATATT  19
                               ||||||||||||||||||
Sbjct                 4514812  TATCCCATTACATGATATT  4514830
CDS:serine hydrolase  301      Y  G  M  V  H  Y  K
```

Bacillus anthracis strain FDAARGOS_706 chromosome, complete genome

GenBank: CP047104.1

FASTA Graphics

Go to:

LOCUS CP047104 19 bp DNA linear BCT **28-DEC-2019**

DEFINITION Bacillus anthracis strain FDAARGOS_706 chromosome, complete genome.

ACCESSION CP047104 REGION: 231048..231066

VERSION CP047104.1

DBLINK BioProject: PRJNA231221
BioSample: SAMN11056421

KEYWORDS .

SOURCE Bacillus anthracis

ORGANISM Bacillus anthracis
Bacteria; Firmicutes; Bacilli; Bacillales; Bacillaceae; Bacillus;
Bacillus cereus group.

REFERENCE 1 (bases 1 to 19)

AUTHORS Sozhamannan,S., Rosenzweig,N., Tallon,L., Sadzewicz,L.,
Vavikolanu,K., Mehta,A., Aluvathingal,J., Nadendla,S., Yan,Y. and
Sichtig,H.

TITLE FDA dAtabase for Regulatory Grade micrObial Sequences (FDA-ARGOS):
Supporting development and validation of Infectious Disease Dx
tests

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 19)

AUTHORS Sozhamannan,S., Rosenzweig,N., Tallon,L., Sadzewicz,L.,
Vavikolanu,K., Mehta,A., Aluvathingal,J., Nadendla,S., Yan,Y. and
Sichtig,H.

TITLE Direct Submission

JOURNAL **Submitted (21-NOV-2019) Center for Devices and Radiological Health,
US Food and Drug Administration, 10903 New Hampshire Avenue, Silver
Spring, MD 20993-0002, USA**

COMMENT The annotation was added by the NCBI Prokaryotic Genome Annotation
Pipeline (PGAP). Information about PGAP can be found here:
https://www.ncbi.nlm.nih.gov/genome/annotation_prok/

##Genome-Assembly-Data-START##
Assembly Method :: canu v. 1.4, SPAdes v. 3.6.0
Genome Coverage :: 879.71x
Sequencing Technology :: Pacbio; Illumina
##Genome-Assembly-Data-END##

##Genome-Annotation-Data-START##
Annotation Provider :: NCBI
Annotation Date :: 12/20/2019 16:08:05

```
            Annotation Pipeline                 :: NCBI Prokaryotic Genome
                                                   Annotation Pipeline (PGAP)
            Annotation Method                   :: Best-placed reference protein
                                                   set; GeneMarkS-2+
            Annotation Software revision        :: 4.10
            Features Annotated                  :: Gene; CDS; rRNA; tRNA; ncRNA;
                                                   repeat_region
            Genes (total)                       :: 5,426
            CDSs (total)                        :: 5,293
            Genes (coding)                      :: 5,017
            CDSs (with protein)                 :: 5,017
            Genes (RNA)                         :: 133
            rRNAs                               :: 11, 11, 11 (5S, 16S, 23S)
            complete rRNAs                      :: 11, 11, 11 (5S, 16S, 23S)
            tRNAs                               :: 95
            ncRNAs                              :: 5
            Pseudo Genes (total)                :: 276
            CDSs (without protein)              :: 276
            Pseudo Genes (ambiguous residues)   :: 0 of 276
            Pseudo Genes (frameshifted)         :: 193 of 276
            Pseudo Genes (incomplete)           :: 34 of 276
            Pseudo Genes (internal stop)        :: 68 of 276
            Pseudo Genes (multiple problems)    :: 16 of 276
            ##Genome-Annotation-Data-END##

FEATURES             Location/Qualifiers
     source          1..19
                     /organism="Bacillus anthracis"
                     /mol_type="genomic DNA"
                     /strain="FDAARGOS_706"
                     /culture_collection="FDA:FDAARGOS_706"
                     /db_xref="taxon:1392"
                     /country="USA: MD"
                     /collected_by="U.S. Army Edgewood Chemical Biological
                     Center"
                     /note="ECBC strain; genetically modified strain"
     gene            <1..>19
                     /locus_tag="FOC14_01155"
     CDS             <1..>19
                     /locus_tag="FOC14_01155"
                     /inference="COORDINATES: similar to AA
                     sequence:RefSeq:WP_000849936.1"
                     /note="Derived by automated computational analysis using
                     gene prediction method: Protein Homology."
                     /codon_start=3
```

/transl_table=11
/product="**serine hydrolase**"
/protein_id="QHA54030.1"
/translation="MKTKDQIDNIVKETYRHIDFSGVILVKEKKGIVYEEVFGYANRN
ECINNTLQTRFGIASGCKIFTAIGICQLVENGVITFQTKLKECLSVNFPNFNEDITIH
HLLTHSSGIPDYFDESIMDNFEDLWKQTPMYLLKSLKDFLPLFQNRDMMYAPGSKFHY
NNAGFIILGLIIEEQTGLTFTEYIEKNIFNPIGMNHSGYFSLDRLPRQTALGYIKDEI
SQTWRTNAYSIPIKGGSDGGAFITASDMLKFWEALFNYEIISQEYTKILLTPHIQVNN
NQSYGYGIWIETRENKIF**KYHVMGY**DPGVSFRSAVYPDLGITLVIPSNKGAGAEKLMI
EIEGSF"

ORIGIN

1 **aatatcatgt aatgggata**

//

Bacillus anthracis strain FDAARGOS_706 chromosome, complete genome
Sequence ID: CP047104.1 **Length: 5227454 Number of Matches: 1**
Range 1: 231048 to 231066 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
38.2 bits(19)	2.2	19/19(100%)	0/19(0%)	Plus/Minus

```
CDS: Putative 1       5          G  M  V  H  Y
Query                 1       TATCCCATTACATGATATT  19
                              ||||||||||||||||||
Sbjct                 231066  TATCCCATTACATGATATT  231048
CDS:serine hydrolase  301     Y  G  M  V  H  Y  K
```

Bacillus anthracis strain FDAARGOS_700 chromosome, complete genome

GenBank: CP047099.1

FASTA Graphics

Go to:

```
LOCUS       CP047099                  19 bp    DNA     linear   BCT 28-DEC-2019
DEFINITION  Bacillus anthracis strain FDAARGOS_700 chromosome, complete genome.
ACCESSION   CP047099 REGION: 2849908..2849926
VERSION     CP047099.1
DBLINK      BioProject: PRJNA231221
            BioSample: SAMN11056415
KEYWORDS    .
SOURCE      Bacillus anthracis
  ORGANISM  Bacillus anthracis
```

Bacteria; Firmicutes; Bacilli; Bacillales; Bacillaceae; Bacillus; Bacillus cereus group.

REFERENCE 1 (bases 1 to 19)

AUTHORS Sozhamannan,S., Rosenzweig,N., Tallon,L., Sadzewicz,L., Vavikolanu,K., Mehta,A., Aluvathingal,J., Nadendla,S., Yan,Y. and Sichtig,H.

TITLE FDA dAtabase for Regulatory Grade micrObial Sequences (FDA-ARGOS): Supporting development and validation of Infectious Disease Dx tests

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 19)

AUTHORS Sozhamannan,S., Rosenzweig,N., Tallon,L., Sadzewicz,L., Vavikolanu,K., Mehta,A., Aluvathingal,J., Nadendla,S., Yan,Y. and Sichtig,H.

TITLE Direct Submission

JOURNAL **Submitted (21-NOV-2019) Center for Devices and Radiological Health, US Food and Drug Administration, 10903 New Hampshire Avenue, Silver Spring, MD 20993-0002, USA**

COMMENT The annotation was added by the NCBI Prokaryotic Genome Annotation Pipeline (PGAP). Information about PGAP can be found here: https://www.ncbi.nlm.nih.gov/genome/annotation_prok/

##Genome-Assembly-Data-START##
Assembly Method :: canu v. 1.4, SPAdes v. 3.6.0
Genome Coverage :: 801.78x
Sequencing Technology :: Pacbio; Illumina
##Genome-Assembly-Data-END##

##Genome-Annotation-Data-START##
Annotation Provider :: NCBI
Annotation Date :: 12/20/2019 16:07:15
Annotation Pipeline :: NCBI Prokaryotic Genome Annotation Pipeline (PGAP)
Annotation Method :: Best-placed reference protein set; GeneMarkS-2+
Annotation Software revision :: 4.10
Features Annotated :: Gene; CDS; rRNA; tRNA; ncRNA; repeat_region
Genes (total) :: 5,420
CDSs (total) :: 5,287
Genes (coding) :: 5,019
CDSs (with protein) :: 5,019
Genes (RNA) :: 133
rRNAs :: 11, 11, 11 (5S, 16S, 23S)

```
            complete rRNAs                    :: 11, 11, 11 (5S, 16S, 23S)
            tRNAs                             :: 95
            ncRNAs                            :: 5
            Pseudo Genes (total)              :: 268
            CDSs (without protein)            :: 268
            Pseudo Genes (ambiguous residues) :: 0 of 268
            Pseudo Genes (frameshifted)       :: 185 of 268
            Pseudo Genes (incomplete)         :: 35 of 268
            Pseudo Genes (internal stop)      :: 67 of 268
            Pseudo Genes (multiple problems)  :: 16 of 268
            ##Genome-Annotation-Data-END##
FEATURES             Location/Qualifiers
     source          1..19
                     /organism="Bacillus anthracis"
                     /mol_type="genomic DNA"
                     /strain="FDAARGOS_700"
                     /culture_collection="FDA:FDAARGOS_700"
                     /db_xref="taxon:1392"
                     /country="USA: MD"
                     /collected_by="U.S. Army Edgewood Chemical Biological
                     Center"
                     /note="ECBC strain; genetically modified strain"
     gene            <1..>19
                     /locus_tag="FOC08_14485"
     CDS             <1..>19
                     /locus_tag="FOC08_14485"
                     /inference="COORDINATES: similar to AA
                     sequence:RefSeq:WP_000849936.1"
                     /note="Derived by automated computational analysis using
                     gene prediction method: Protein Homology."
                     /codon_start=3
                     /transl_table=11
                     /product="serine hydrolase"
                     /protein_id="QHA51366.1"
                     /translation="MKTKDQIDNIVKETYRHIDFSGVILVKEKKGIVYEEVFGYANRN
                     ECINNTLQTRFGIASGCKIFTAIGICQLVENGVITFQTKLKECLSVNFPNFNEDITIH
                     HLLTHSSGIPDYFDESIMDNFEDLWKQTPMYLLKSLKDFLPLFQNRDMMYAPGSKFHY
                     NNAGFIILGLIIEEQTGLTFTEYIEKNIFNPIGMNHSGYFSLDRLPRQTALGYIKDEI
                     SQTWRTNAYSIPIKGGSDGGAFITASDMLKFWEALFNYEIISQEYTKILLTPHIQVNN
                     NQSYGYGIWIETRENKIFKYHVMGYDPGVSFRSAVYPDLGITLVIPSNKGAGAEKLMI
                     EIEGSF"
ORIGIN
        1 aatatcatgt aatgggata
//
```

Bacillus anthracis strain FDAARGOS_700 chromosome, complete genome
Sequence ID: CP047099.1 **Length: 5227419 Number of Matches: 1**
Range 1: 2849908 to 2849926 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
38.2 bits(19)	2.2	19/19(100%)	0/19(0%)	Plus/Minus

```
CDS: Putative 1       5           G  M  V  H  Y
Query                 1        TATCCCATTACATGATATT  19
                               ||||||||||||||||||
Sbjct                 2849926  TATCCCATTACATGATATT  2849908
CDS:serine hydrolase  301      Y  G  M  V  H  Y  K
```

Bacillus anthracis strain FDAARGOS_701 chromosome, complete genome

GenBank: CP047098.1

FASTA Graphics

Go to:

```
LOCUS       CP047098                 19 bp    DNA     linear   BCT 28-DEC-2019
DEFINITION  Bacillus anthracis strain FDAARGOS_701 chromosome, complete genome.
ACCESSION   CP047098 REGION: 5146998..5147016
VERSION     CP047098.1
DBLINK      BioProject: PRJNA231221
            BioSample: SAMN11056416
KEYWORDS    .
SOURCE      Bacillus anthracis
  ORGANISM  Bacillus anthracis
            Bacteria; Firmicutes; Bacilli; Bacillales; Bacillaceae; Bacillus;
            Bacillus cereus group.
REFERENCE   1  (bases 1 to 19)
  AUTHORS   Sozhamannan,S., Rosenzweig,N., Tallon,L., Sadzewicz,L.,
            Vavikolanu,K., Mehta,A., Aluvathingal,J., Nadendla,S., Yan,Y. and
            Sichtig,H.
  TITLE     FDA dAtabase for Regulatory Grade micrObial Sequences (FDA-ARGOS):
            Supporting development and validation of Infectious Disease Dx
            tests
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 19)
  AUTHORS   Sozhamannan,S., Rosenzweig,N., Tallon,L., Sadzewicz,L.,
            Vavikolanu,K., Mehta,A., Aluvathingal,J., Nadendla,S., Yan,Y. and
            Sichtig,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-NOV-2019) Center for Devices and Radiological Health,
            US Food and Drug Administration, 10903 New Hampshire Avenue, Silver
            Spring, MD 20993-0002, USA
```

```
COMMENT     The annotation was added by the NCBI Prokaryotic Genome Annotation
            Pipeline (PGAP). Information about PGAP can be found here:
            https://www.ncbi.nlm.nih.gov/genome/annotation_prok/

            ##Genome-Assembly-Data-START##
            Assembly Method       :: SMRT v. 2.3.0, HGAP v. 3
            Genome Coverage       :: 924.47x
            Sequencing Technology :: Pacbio; Illumina
            ##Genome-Assembly-Data-END##


            ##Genome-Annotation-Data-START##
            Annotation Provider               :: NCBI
            Annotation Date                   :: 12/20/2019 16:30:49
            Annotation Pipeline               :: NCBI Prokaryotic Genome
                                                 Annotation Pipeline (PGAP)
            Annotation Method                 :: Best-placed reference protein
                                                 set; GeneMarkS-2+
            Annotation Software revision      :: 4.10
            Features Annotated                :: Gene; CDS; rRNA; tRNA; ncRNA;
                                                 repeat_region
            Genes (total)                     :: 5,423
            CDSs (total)                      :: 5,290
            Genes (coding)                    :: 5,023
            CDSs (with protein)               :: 5,023
            Genes (RNA)                       :: 133
            rRNAs                             :: 11, 11, 11 (5S, 16S, 23S)
            complete rRNAs                    :: 11, 11, 11 (5S, 16S, 23S)
            tRNAs                             :: 95
            ncRNAs                            :: 5
            Pseudo Genes (total)              :: 267
            CDSs (without protein)            :: 267
            Pseudo Genes (ambiguous residues) :: 0 of 267
            Pseudo Genes (frameshifted)       :: 186 of 267
            Pseudo Genes (incomplete)         :: 35 of 267
            Pseudo Genes (internal stop)      :: 65 of 267
            Pseudo Genes (multiple problems)  :: 16 of 267
            ##Genome-Annotation-Data-END##
FEATURES             Location/Qualifiers
     source          1..19
                     /organism="Bacillus anthracis"
                     /mol_type="genomic DNA"
                     /strain="FDAARGOS_701"
                     /culture_collection="FDA:FDAARGOS_701"
                     /db_xref="taxon:1392"
```

/country="USA: MD"
/collected_by="U.S. Army Edgewood Chemical Biological Center"
/note="ECBC strain; genetically modified strain"

gene complement(<1..>19)
/locus_tag="FOC09_26745"

CDS complement(<1..>19)
/locus_tag="FOC09_26745"
/inference="COORDINATES: similar to AA sequence:RefSeq:WP_000849936.1"
/note="Derived by automated computational analysis using gene prediction method: Protein Homology."
/codon_start=3
/transl_table=11
/product="**serine hydrolase**"
/protein_id="QHA48551.1"
/translation="MKTKDQIDNIVKETYRHIDFSGVILVKEKKGIVYEEVFGYANRN
ECINNTLQTRFGIASGCKIFTAIGICQLVENGVITFQTKLKECLSVNFPNFNEDITIH
HLLTHSSGIPDYFDESIMDNFEDLWKQTPMYLLKSLKDFLPLFQNRDMMYAPGSKFHY
NNAGFIILGLIIEEQTGLTFTEYIEKNIFNPIGMNHSGYFSLDRLPRQTALGYIKDEI
SQTWRTNAYSIPIKGGSDGGAFITASDMLKFWEALFNYEIISQEYTKILLTPHIQVNN
NQSYGYGIWIETRENKIF**KYHVMGY**DPGVSFRSAVYPDLGITLVIPSNKGAGAEKLMI
EIEGSF"

ORIGIN
1 **tatcccatta catgatatt**
//

Bacillus anthracis strain FDAARGOS_701 chromosome, complete genome
Sequence ID: CP047098.1 **Length: 5227419 Number of Matches: 1**
Range 1: 5146998 to 5147016 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
38.2 bits(19)	2.2	19/19(100%)	0/19(0%)	Plus/Plus

```
CDS: Putative 1       5           G  M  V  H  Y
Query                 1        TATCCCATTACATGATATT  19
                               ||||||||||||||||||
Sbjct                 5146998  TATCCCATTACATGATATT  5147016
CDS:serine hydrolase  301      Y  G  M  V  H  Y  K
```

Bacillus anthracis strain FDAARGOS_698 chromosome, complete genome

GenBank: CP047097.1

FASTA Graphics

Go to:

LOCUS CP047097 19 bp DNA linear BCT **28-DEC-2019**

DEFINITION Bacillus anthracis strain FDAARGOS_698 chromosome, complete genome.

ACCESSION CP047097 REGION: 1622488..1622506

VERSION CP047097.1

DBLINK BioProject: PRJNA231221

BioSample: SAMN11056413

KEYWORDS .

SOURCE Bacillus anthracis

ORGANISM Bacillus anthracis

Bacteria; Firmicutes; Bacilli; Bacillales; Bacillaceae; Bacillus; Bacillus cereus group.

REFERENCE 1 (bases 1 to 19)

AUTHORS Sozhamannan,S., Rosenzweig,N., Tallon,L., Sadzewicz,L., Vavikolanu,K., Mehta,A., Aluvathingal,J., Nadendla,S., Yan,Y. and Sichtig,H.

TITLE FDA dAtabase for Regulatory Grade micrObial Sequences (FDA-ARGOS): Supporting development and validation of Infectious Disease Dx tests

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 19)

AUTHORS Sozhamannan,S., Rosenzweig,N., Tallon,L., Sadzewicz,L., Vavikolanu,K., Mehta,A., Aluvathingal,J., Nadendla,S., Yan,Y. and Sichtig,H.

TITLE Direct Submission

JOURNAL **Submitted (21-NOV-2019) Center for Devices and Radiological Health, US Food and Drug Administration, 10903 New Hampshire Avenue, Silver Spring, MD 20993-0002, USA**

COMMENT The annotation was added by the NCBI Prokaryotic Genome Annotation Pipeline (PGAP). Information about PGAP can be found here: https://www.ncbi.nlm.nih.gov/genome/annotation_prok/

```
##Genome-Assembly-Data-START##
Assembly Method       :: CA v. 8.2
Genome Coverage       :: 947.26x
Sequencing Technology :: Pacbio; Illumina
##Genome-Assembly-Data-END##

##Genome-Annotation-Data-START##
Annotation Provider               :: NCBI
Annotation Date                   :: 12/20/2019 16:21:08
```

```
            Annotation Pipeline                 :: NCBI Prokaryotic Genome
                                                   Annotation Pipeline (PGAP)
            Annotation Method                   :: Best-placed reference protein
                                                   set; GeneMarkS-2+
            Annotation Software revision        :: 4.10
            Features Annotated                  :: Gene; CDS; rRNA; tRNA; ncRNA;
                                                   repeat_region
            Genes (total)                       :: 5,422
            CDSs (total)                        :: 5,289
            Genes (coding)                      :: 5,007
            CDSs (with protein)                 :: 5,007
            Genes (RNA)                         :: 133
            rRNAs                               :: 11, 11, 11 (5S, 16S, 23S)
            complete rRNAs                      :: 11, 11, 11 (5S, 16S, 23S)
            tRNAs                               :: 95
            ncRNAs                              :: 5
            Pseudo Genes (total)                :: 282
            CDSs (without protein)              :: 282
            Pseudo Genes (ambiguous residues)   :: 0 of 282
            Pseudo Genes (frameshifted)         :: 198 of 282
            Pseudo Genes (incomplete)           :: 36 of 282
            Pseudo Genes (internal stop)        :: 67 of 282
            Pseudo Genes (multiple problems)    :: 16 of 282
            ##Genome-Annotation-Data-END##

FEATURES             Location/Qualifiers
     source          1..19
                     /organism="Bacillus anthracis"
                     /mol_type="genomic DNA"
                     /strain="FDAARGOS_698"
                     /culture_collection="FDA:FDAARGOS_698"
                     /db_xref="taxon:1392"
                     /country="USA: MD"
                     /collected_by="U.S. Army Edgewood Chemical Biological
                     Center"
                     /note="ECBC strain; genetically modified strain"
     gene            <1..>19
                     /locus_tag="FOC06_08230"
     CDS             <1..>19
                     /locus_tag="FOC06_08230"
                     /inference="COORDINATES: similar to AA
                     sequence:RefSeq:WP_000849936.1"
                     /note="Derived by automated computational analysis using
                     gene prediction method: Protein Homology."
                     /codon_start=3
```

/transl_table=11
/product="**serine hydrolase**"
/protein_id="QHA40231.1"
/translation="MKTKDQIDNIVKETYRHIDFSGVILVKEKKGIVYEEVFGYANRN
ECINNTLQTRFGIASGCKIFTAIGICQLVENGVITFQTKLKECLSVNFPNFNEDITIH
HLLTHSSGIPDYFDESIMDNFEDLWKQTPMYLLKSLKDFLPLFQNRDMMYAPGSKFHY
NNAGFIILGLIIEEQTGLTFTEYIEKNIFNPIGMNHSGYFSLDRLPRQTALGYIKDEI
SQTWRTNAYSIPIKGGSDGGAFITASDMLKFWEALFNYEIISQEYTKILLTPHIQVNN
NQSYGYGIWIETRENKIF**KYHVMGY**DPGVSFRSAVYPDLGITLVIPSNKGAGAEKLMI
EIEGSF"

ORIGIN

1 **aatatcatgt aatgggata**

//

Bacillus anthracis strain FDAARGOS_698 chromosome, complete genome
Sequence ID: CP047097.1 **Length: 5225746 Number of Matches: 1**
Range 1: 1622488 to 1622506 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
38.2 bits(19)	2.2	19/19(100%)	0/19(0%)	Plus/Minus

```
CDS: Putative 1       5           G  M  V  H  Y
Query                 1        TATCCCATTACATGATATT  19
                               ||||||||||||||||||
Sbjct                 1622506  TATCCCATTACATGATATT  1622488
CDS:serine hydrolase  301      Y  G  M  V  H  Y  K
```

Bacillus paranthracis strain BC307 chromosome, complete genome

GenBank: CP047085.1

FASTA Graphics

Go to:

```
LOCUS       CP047085                 19 bp    DNA     linear   BCT 26-DEC-2019
DEFINITION  Bacillus paranthracis strain BC307 chromosome, complete genome.
ACCESSION   CP047085 REGION: 1791999..1792017
VERSION     CP047085.1
DBLINK      BioProject: PRJNA596370
            BioSample: SAMN13620036
KEYWORDS    .
SOURCE      Bacillus paranthracis
  ORGANISM  Bacillus paranthracis
            Bacteria; Firmicutes; Bacilli; Bacillales; Bacillaceae; Bacillus;
            Bacillus cereus group.
REFERENCE   1  (bases 1 to 19)
```

AUTHORS Wang,D., Liu,X., Liu,X., Lyu,Y., Tong,Y. and Wang,H.
TITLE Complete genome sequence of Bacillus cereus BC307, a typical strain with high production of emetic toxin cereulide
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 19)
AUTHORS Wang,D., Liu,X., Liu,X., Lyu,Y., Tong,Y. and Wang,H.
TITLE Direct Submission
JOURNAL **Submitted (18-DEC-2019) 1.tState Key Laboratory of Pathogen and Biosecurity, Beijing Institute of Biotechnology, 20 Dongdajie Street, Beijing 100071, China**
COMMENT Bacteria and source DNA available from http://www.cmccb.org.cn/cmccbnew/
The annotation was added by the NCBI Prokaryotic Genome Annotation Pipeline (PGAP). Information about PGAP can be found here: https://www.ncbi.nlm.nih.gov/genome/annotation_prok/

##Genome-Assembly-Data-START##
Assembly Method :: HGAP v. SMRT Analysis 2.3
Genome Representation :: Full
Expected Final Version :: Yes
Genome Coverage :: 20.0x
Sequencing Technology :: PacBio RSII; Illumina MiniSeq
##Genome-Assembly-Data-END##

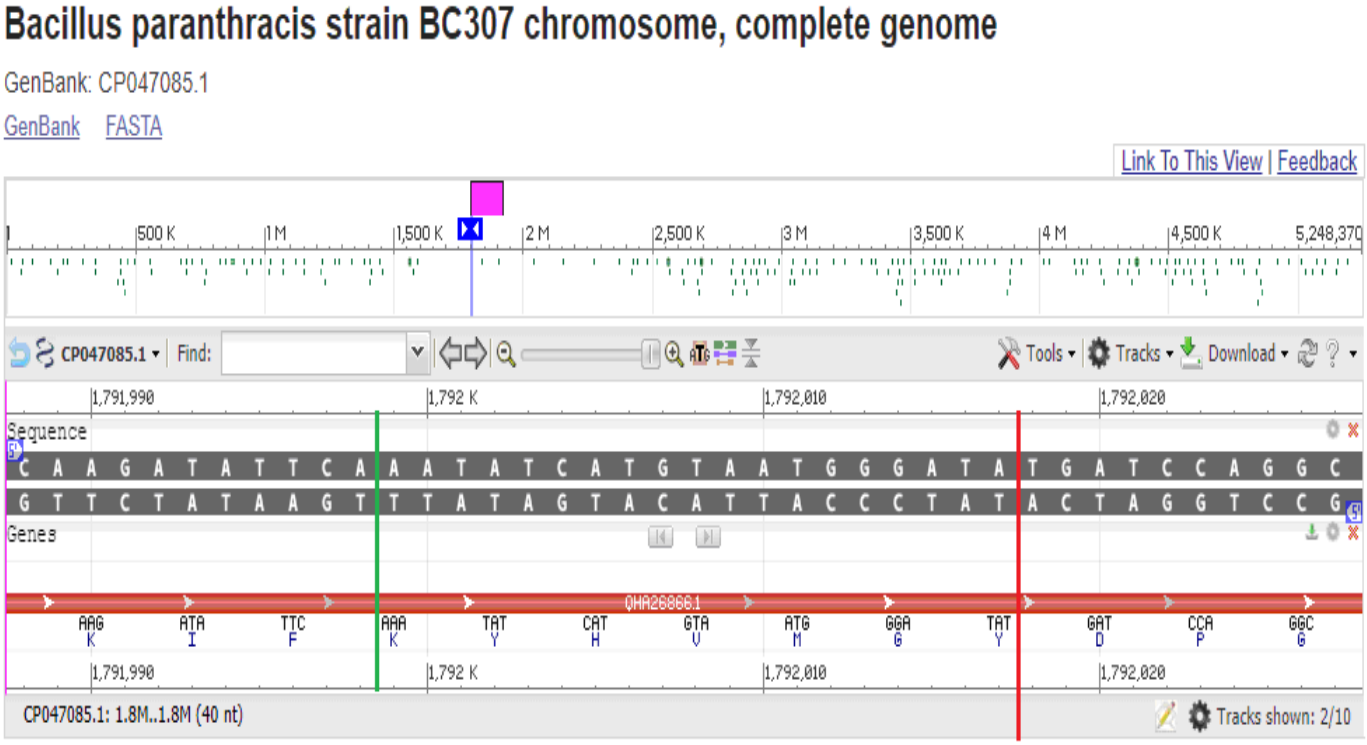
##Genome-Annotation-Data-START##
Annotation Provider :: NCBI
Annotation Date :: 12/19/2019 16:23:41
Annotation Pipeline :: NCBI Prokaryotic Genome Annotation Pipeline (PGAP)
Annotation Method :: Best-placed reference protein set; GeneMarkS-2+
Annotation Software revision :: 4.10
Features Annotated :: Gene; CDS; rRNA; tRNA; ncRNA; repeat_region
Genes (total) :: 5,736
CDSs (total) :: 5,585
Genes (coding) :: 5,253
CDSs (with protein) :: 5,253
Genes (RNA) :: 151
rRNAs :: 14, 14, 14 (5S, 16S, 23S)
complete rRNAs :: 14, 14, 14 (5S, 16S, 23S)
tRNAs :: 104
ncRNAs :: 5
Pseudo Genes (total) :: 332

```
            CDSs (without protein)             :: 332
            Pseudo Genes (ambiguous residues)  :: 0 of 332
            Pseudo Genes (frameshifted)        :: 261 of 332
            Pseudo Genes (incomplete)          :: 58 of 332
            Pseudo Genes (internal stop)       :: 30 of 332
            Pseudo Genes (multiple problems)   :: 15 of 332
            ##Genome-Annotation-Data-END##
FEATURES             Location/Qualifiers
     source          1..19
                     /organism="Bacillus paranthracis"
                     /mol_type="genomic DNA"
                     /strain="BC307"
                     /isolation_source="vomit from a food poisoning case"
                     /culture_collection="CMCC:63317"
                     /db_xref="taxon:2026186"
                     /country="China: Beijing"
                     /collection_date="2015-08-12"
                     /collected_by="Liu Xingming"
     gene            <1..>19
                     /locus_tag="GQZ26_09170"
     CDS             <1..>19
                     /locus_tag="GQZ26_09170"
                     /inference="COORDINATES: similar to AA
                     sequence:RefSeq:WP_000849936.1"
                     /note="Derived by automated computational analysis using
                     gene prediction method: Protein Homology."
                     /codon_start=3
                     /transl_table=11
                     /product="serine hydrolase"
                     /protein_id="QHA26866.1"
                     /translation="MKTKNQIDNIVKEIYRHIDFSGVILVKEKKGIVYEEAFGYANRN
                     EYINNTLQTRFGIASGCKIFTAIGICQLVERGVITFHTKLKDCLSIKFSNFNEDITIH
                     HLLTHSSGIPDYFDESILDNFEDLWKETPMYLLKNLKDFLPLFQNRDMMFAPGSKFHY
                     NNAGFIILGLIIEEQTGLKFTEYIETNIFKLIGMNDSGYFSLDRLPKQTALGYIKDEI
                     SQTWRTNAYSIPIKGGSDGGAFITAPDMMKFWEALFNYEIISQEYTKILLTPHIQVNS
                     YQSYGYGIWIEIRENKIFKYHVMGYDPGVSFRSAVYPDLGITLVIPSNKGAGPEKLMI
                     EIEGAF"
ORIGIN
        1 aatatcatgt aatgggata
//
```

Bacillus paranthracis strain BC307 chromosome, complete genome
Sequence ID: CP047085.1 **Length: 5248370 Number of Matches: 2**
Range 1: 1791999 to 1792017 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
38.2 bits(19)	2.2	19/19(100%)	0/19(0%)	Plus/Minus

```
CDS: Putative 1        5            G  M  V  H  Y
Query                  1        TATCCCATTACATGATATT  19
                                |||||||||||||||||||
Sbjct                  1792017  TATCCCATTACATGATATT  1791999
CDS:serine hydrolase   301         Y  G  M  V  H  Y  K
```



Bacillus paranthracis strain BC307 chromosome, complete genome

GenBank: CP047085.1

FASTA Graphics

Go to:

```
LOCUS       CP047085                  16 bp    DNA     linear   BCT 26-DEC-2019
DEFINITION  Bacillus paranthracis strain BC307 chromosome, complete genome.
ACCESSION   CP047085 REGION: 717567..717582
VERSION     CP047085.1
DBLINK      BioProject: PRJNA596370
            BioSample: SAMN13620036
KEYWORDS    .
```

SOURCE Bacillus paranthracis
  ORGANISM Bacillus paranthracis
Bacteria; Firmicutes; Bacilli; Bacillales; Bacillaceae; Bacillus;
Bacillus cereus group.
REFERENCE 1 (bases 1 to 16)
  AUTHORS Wang,D., Liu,X., Liu,X., Lyu,Y., Tong,Y. and Wang,H.
  TITLE Complete genome sequence of Bacillus cereus BC307, a typical strain with high production of emetic toxin cereulide
  JOURNAL Unpublished
REFERENCE 2 (bases 1 to 16)
  AUTHORS Wang,D., Liu,X., Liu,X., Lyu,Y., Tong,Y. and Wang,H.
  TITLE Direct Submission
  JOURNAL **Submitted (18-DEC-2019) 1.tState Key Laboratory of Pathogen and Biosecurity, Beijing Institute of Biotechnology, 20 Dongdajie Street, Beijing 100071, China**
COMMENT Bacteria and source DNA available from
http://www.cmccb.org.cn/cmccbnew/
The annotation was added by the NCBI Prokaryotic Genome Annotation Pipeline (PGAP). Information about PGAP can be found here:
https://www.ncbi.nlm.nih.gov/genome/annotation_prok/

##Genome-Assembly-Data-START##
Assembly Method :: HGAP v. SMRT Analysis 2.3
Genome Representation :: Full
Expected Final Version :: Yes
Genome Coverage :: 20.0x
Sequencing Technology :: PacBio RSII; Illumina MiniSeq
##Genome-Assembly-Data-END##

##Genome-Annotation-Data-START##
Annotation Provider :: NCBI
Annotation Date :: 12/19/2019 16:23:41
Annotation Pipeline :: NCBI Prokaryotic Genome Annotation Pipeline (PGAP)
Annotation Method :: Best-placed reference protein set; GeneMarkS-2+
Annotation Software revision :: 4.10
Features Annotated :: Gene; CDS; rRNA; tRNA; ncRNA; repeat_region
Genes (total) :: 5,736
CDSs (total) :: 5,585
Genes (coding) :: 5,253
CDSs (with protein) :: 5,253
Genes (RNA) :: 151

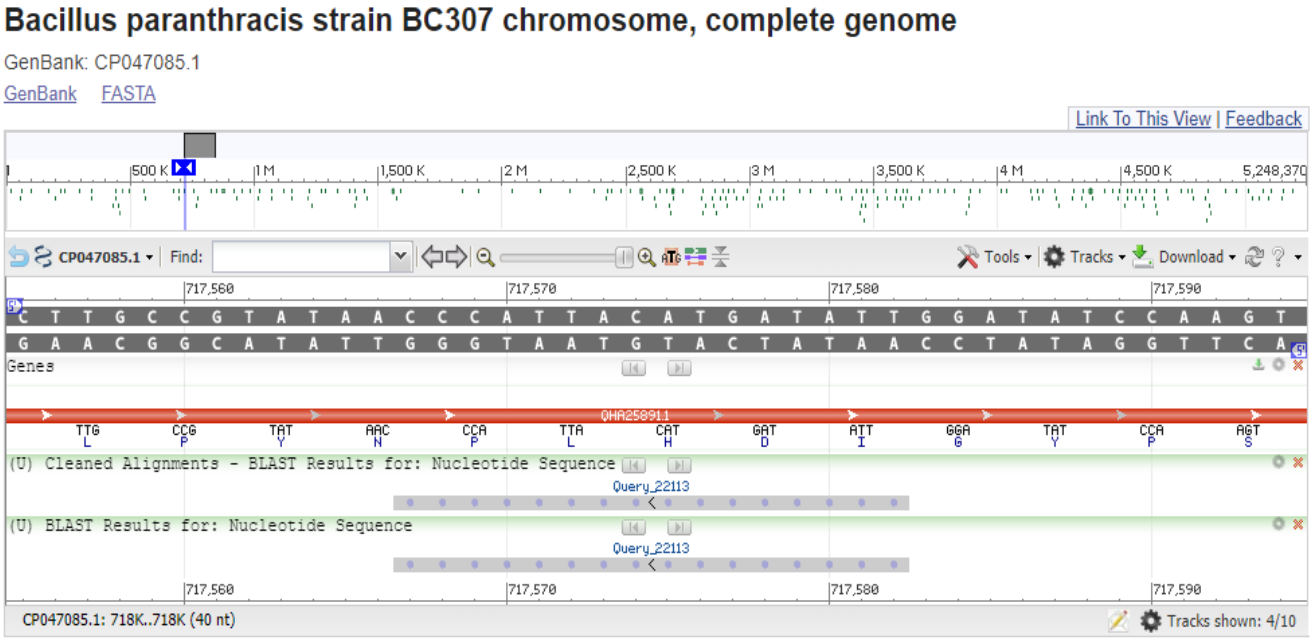
```
            rRNAs                              :: 14, 14, 14 (5S, 16S, 23S)
            complete rRNAs                     :: 14, 14, 14 (5S, 16S, 23S)
            tRNAs                              :: 104
            ncRNAs                             :: 5
            Pseudo Genes (total)               :: 332
            CDSs (without protein)             :: 332
            Pseudo Genes (ambiguous residues)  :: 0 of 332
            Pseudo Genes (frameshifted)        :: 261 of 332
            Pseudo Genes (incomplete)          :: 58 of 332
            Pseudo Genes (internal stop)       :: 30 of 332
            Pseudo Genes (multiple problems)   :: 15 of 332
            ##Genome-Annotation-Data-END##
FEATURES             Location/Qualifiers
     source          1..16
                     /organism="Bacillus paranthracis"
                     /mol_type="genomic DNA"
                     /strain="BC307"
                     /isolation_source="vomit from a food poisoning case"
                     /culture_collection="CMCC:63317"
                     /db_xref="taxon:2026186"
                     /country="China: Beijing"
                     /collection_date="2015-08-12"
                     /collected_by="Liu Xingming"
     gene            <1..>16
                     /gene="cysH"
                     /locus_tag="GQZ26_03700"
     CDS             <1..>16
                     /gene="cysH"
                     /locus_tag="GQZ26_03700"
                     /EC_number="1.8.4.8"
                     /inference="COORDINATES: similar to AA
                     sequence:RefSeq:WP_000958995.1"
                     /note="Derived by automated computational analysis using
                     gene prediction method: Protein Homology."
                     /codon_start=2
                     /transl_table=11
                     /product="phosphoadenylyl-sulfate reductase"
                     /protein_id="QHA25891.1"
                     /translation="MLTYETWEENDVSFSEEDETKGALSVLSWAYKEYENEIVYACSF
                     GVEGMVLLHLINQVNPSAKVVFLDTNVHFQETYELIQKVRERFPSLNIIEKQPKLTLD
                     EQAKLHGNKLWESNPNLCCKIRKILPLEESLANEKAWISGLRREQSETRKHTKFINQD
                     HRFQSIKVCPLIHWTWKEVWRYVYKHSLPYNPLHDIGYPSIGCEKCTLPVGEGGDSRD
                     GRWAGKVKTECGLHYQ"
ORIGIN
        1 cccattacat gatatt
```

//

Range 2: 717567 to 717582 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
32.2 bits(16)	136	16/16(100%)	0/16(0%)	Plus/Plus

```
CDS: Putative 1        1          P  L  H  D  I
Query                  4       CCCATTACATGATATT  19
                               ||||||||||||||||
Sbjct                  717567  CCCATTACATGATATT  717582
CDS:phosphoadenylyl-   192        P  L  H  D  I
```



Bacillus paranthracis strain CFSAN068816 chromosome, complete genome

GenBank: CP045777.1

FASTA Graphics

Go to:

```
LOCUS       CP045777                19 bp    DNA     linear   BCT 11-NOV-2019
DEFINITION  Bacillus paranthracis strain CFSAN068816 chromosome, complete
            genome.
ACCESSION   CP045777 REGION: 3106170..3106188
VERSION     CP045777.1
DBLINK      BioProject: PRJNA541977
            BioSample: SAMN10574728
KEYWORDS    .
SOURCE      Bacillus paranthracis
  ORGANISM  Bacillus paranthracis
```

Bacteria; Firmicutes; Bacilli; Bacillales; Bacillaceae; Bacillus; Bacillus cereus group.

REFERENCE 1 (bases 1 to 19)

AUTHORS Mafiz,A., Kastanis,G. and Gonzalez Escalona,N.

TITLE Closed genomes of twenty-eight foodborne pathogens from the CFSAN verification set by a combination of long and short-reads

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 19)

AUTHORS Mafiz,A., Kastanis,G. and Gonzalez Escalona,N.

TITLE Direct Submission

JOURNAL **Submitted (01-NOV-2019) ORS/DM/MMSB, Food and Drug Administration-College Park, 5001 Campus Drive, College Park, MD 20740-3835, USA**

COMMENT The annotation was added by the NCBI Prokaryotic Genome Annotation Pipeline (PGAP). Information about PGAP can be found here: https://www.ncbi.nlm.nih.gov/genome/annotation_prok/

##Genome-Assembly-Data-START##
Assembly Method :: SPAdes, CANU v. 3.12.0, 1.7
Genome Representation :: Full
Expected Final Version :: Yes
Genome Coverage :: 84, 43, 113, 105, 195, 264x
Sequencing Technology :: Oxford Nanopore GridION, Illumina MiSeq
##Genome-Assembly-Data-END##

##Genome-Annotation-Data-START##

Annotation Provider	:: NCBI
Annotation Date	:: 11/04/2019 10:11:02
Annotation Pipeline	:: NCBI Prokaryotic Genome Annotation Pipeline (PGAP)
Annotation Method	:: Best-placed reference protein set; GeneMarkS-2+
Annotation Software revision	:: 4.10
Features Annotated	:: Gene; CDS; rRNA; tRNA; ncRNA; repeat_region
Genes (total)	:: 5,736
CDSs (total)	:: 5,585
Genes (coding)	:: 5,463
CDSs (with protein)	:: 5,463
Genes (RNA)	:: 151
rRNAs	:: 14, 14, 14 (5S, 16S, 23S)
complete rRNAs	:: 14, 14, 14 (5S, 16S, 23S)
tRNAs	:: 104
ncRNAs	:: 5
Pseudo Genes (total)	:: 122

```
            CDSs (without protein)             :: 122
            Pseudo Genes (ambiguous residues)  :: 0 of 122
            Pseudo Genes (frameshifted)        :: 49 of 122
            Pseudo Genes (incomplete)          :: 63 of 122
            Pseudo Genes (internal stop)       :: 27 of 122
            Pseudo Genes (multiple problems)   :: 15 of 122
            ##Genome-Annotation-Data-END##
FEATURES             Location/Qualifiers
     source          1..19
                     /organism="Bacillus paranthracis"
                     /mol_type="genomic DNA"
                     /strain="CFSAN068816"
                     /isolation_source="Mac and Cheese"
                     /db_xref="taxon:2026186"
                     /country="USA:CT"
                     /collection_date="2009-06-11"
                     /collected_by="FDA"
     gene            complement(<1..>19)
                     /locus_tag="GH772_17370"
     CDS             complement(<1..>19)
                     /locus_tag="GH772_17370"
                     /inference="COORDINATES: similar to AA
                     sequence:RefSeq:WP_000849936.1"
                     /note="Derived by automated computational analysis using
                     gene prediction method: Protein Homology."
                     /codon_start=3
                     /transl_table=11
                     /product="serine hydrolase"
                     /protein_id="QGG19400.1"
                     /translation="MKTKNQIDNIVKEIYRHIDFSGVILVKEKKGIVYEEAFGYANRN
                     EYINNTLQTRFGIASGCKIFTAIGICQLVERGVITFHTKLKDCLSIKFSNFNEDITIH
                     HLLTHSSGIPDYFDESILDNFEDLWKETPMYLLKNLKDFLPLFQNRDMMFAPGSKFHY
                     NNAGFIILGLIIEEQTGLKFTEYIETNIFKLIGMNDSGYFSLDRLPKQTALGYIKDEI
                     SQTWRTNAYSIPIKGGSDGGAFITAPDMMKFWEALFNYEIISQEYTKILLTPHIQVNS
                     YQSYGYGIWIEIRENKIFKYHVMGYDPGVSFRSAVYPDLGITLVIPSNKGAGPEKLMI
                     EIEGAF"
ORIGIN
        1 tatcccatta catgatatt
//
```

Bacillus paranthracis strain CFSAN068816 chromosome, complete genome
Sequence ID: CP045777.1 **Length: 5351395 Number of Matches: 2**
Range 1: 3106170 to 3106188 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
38.2 bits(19)	2.2	19/19(100%)	0/19(0%)	Plus/Plus

```
CDS: Putative 1       5           G  M  V  H  Y
Query                 1        TATCCCATTACATGATATT  19
                               ||||||||||||||||||
Sbjct                 3106170  TATCCCATTACATGATATT  3106188
CDS:serine hydrolase  301      Y  G  M  V  H  Y  K
```

Bacillus paranthracis strain CFSAN068816 chromosome, complete genome

GenBank: CP045777.1

FASTA Graphics

Go to:

```
LOCUS       CP045777                 16 bp    DNA     linear   BCT 11-NOV-2019
DEFINITION  Bacillus paranthracis strain CFSAN068816 chromosome, complete
            genome.
ACCESSION   CP045777 REGION: 4177589..4177604
VERSION     CP045777.1
DBLINK      BioProject: PRJNA541977
            BioSample: SAMN10574728
KEYWORDS    .
SOURCE      Bacillus paranthracis
  ORGANISM  Bacillus paranthracis
            Bacteria; Firmicutes; Bacilli; Bacillales; Bacillaceae; Bacillus;
            Bacillus cereus group.
REFERENCE   1  (bases 1 to 16)
  AUTHORS   Mafiz,A., Kastanis,G. and Gonzalez Escalona,N.
  TITLE     Closed genomes of twenty-eight foodborne pathogens from the CFSAN
            verification set by a combination of long and short-reads
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 16)
  AUTHORS   Mafiz,A., Kastanis,G. and Gonzalez Escalona,N.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-NOV-2019) ORS/DM/MMSB, Food and Drug
            Administration-College Park, 5001 Campus Drive, College Park, MD
            20740-3835, USA
COMMENT     The annotation was added by the NCBI Prokaryotic Genome Annotation
            Pipeline (PGAP). Information about PGAP can be found here:
            https://www.ncbi.nlm.nih.gov/genome/annotation_prok/
```

```
##Genome-Assembly-Data-START##
Assembly Method        :: SPAdes, CANU v. 3.12.0, 1.7
Genome Representation  :: Full
Expected Final Version :: Yes
Genome Coverage        :: 84, 43, 113, 105, 195, 264x
Sequencing Technology  :: Oxford Nanopore GridION, Illumina MiSeq
##Genome-Assembly-Data-END##

##Genome-Annotation-Data-START##
Annotation Provider               :: NCBI
Annotation Date                   :: 11/04/2019 10:11:02
Annotation Pipeline               :: NCBI Prokaryotic Genome
                                     Annotation Pipeline (PGAP)
Annotation Method                 :: Best-placed reference protein
                                     set; GeneMarkS-2+
Annotation Software revision      :: 4.10
Features Annotated                :: Gene; CDS; rRNA; tRNA; ncRNA;
                                     repeat_region
Genes (total)                     :: 5,736
CDSs (total)                      :: 5,585
Genes (coding)                    :: 5,463
CDSs (with protein)               :: 5,463
Genes (RNA)                       :: 151
rRNAs                             :: 14, 14, 14 (5S, 16S, 23S)
complete rRNAs                    :: 14, 14, 14 (5S, 16S, 23S)
tRNAs                             :: 104
ncRNAs                            :: 5
Pseudo Genes (total)              :: 122
CDSs (without protein)            :: 122
Pseudo Genes (ambiguous residues) :: 0 of 122
Pseudo Genes (frameshifted)       :: 49 of 122
Pseudo Genes (incomplete)         :: 63 of 122
Pseudo Genes (internal stop)      :: 27 of 122
Pseudo Genes (multiple problems)  :: 15 of 122
##Genome-Annotation-Data-END##

FEATURES             Location/Qualifiers
     source          1..16
                     /organism="Bacillus paranthracis"
                     /mol_type="genomic DNA"
                     /strain="CFSAN068816"
                     /isolation_source="Mac and Cheese"
                     /db_xref="taxon:2026186"
                     /country="USA:CT"
```

```
                     /collection_date="2009-06-11"
                     /collected_by="FDA"
     gene            complement(<1..>16)
                     /gene="cysH"
                     /locus_tag="GH772_22815"
     CDS             complement(<1..>16)
                     /gene="cysH"
                     /locus_tag="GH772_22815"
                     /EC_number="1.8.4.8"
                     /inference="COORDINATES: similar to AA
                     sequence:RefSeq:WP_000958995.1"
                     /note="Derived by automated computational analysis using
                     gene prediction method: Protein Homology."
                     /codon_start=2
                     /transl_table=11
                     /product="phosphoadenylyl-sulfate reductase"
                     /protein_id="QGG20433.1"
                     /translation="MLTYETWEENDVSFSEEDETKGALSVLSWAYKEYENEIVYACSF
                     GVEGMVLLHLINQVNPSAKVVFLDTNVHFQETYELIQKVRERFPSLNIIEKQPKLTLD
                     EQAKLHGNKLWESNPNLCCKIRKILPLEESLANEKAWISGLRREQSETRKHTKFINQD
                     HRFQSIKVCPLIHWTWKEVWRYVYKHSLPYNPLHDIGYPSIGCEKCTLPVGEGGDSRD
                     GRWAGKVKTECGLHYQ"
ORIGIN
        1 aatatcatgt aatggg
//
```

Bacillus paranthracis strain CFSAN068816 chromosome, complete genome
Sequence ID: CP045777.1 Length: 5351395 Number of Matches: 2
Range 1: 3106173 to 3106188 GenBank Graphics

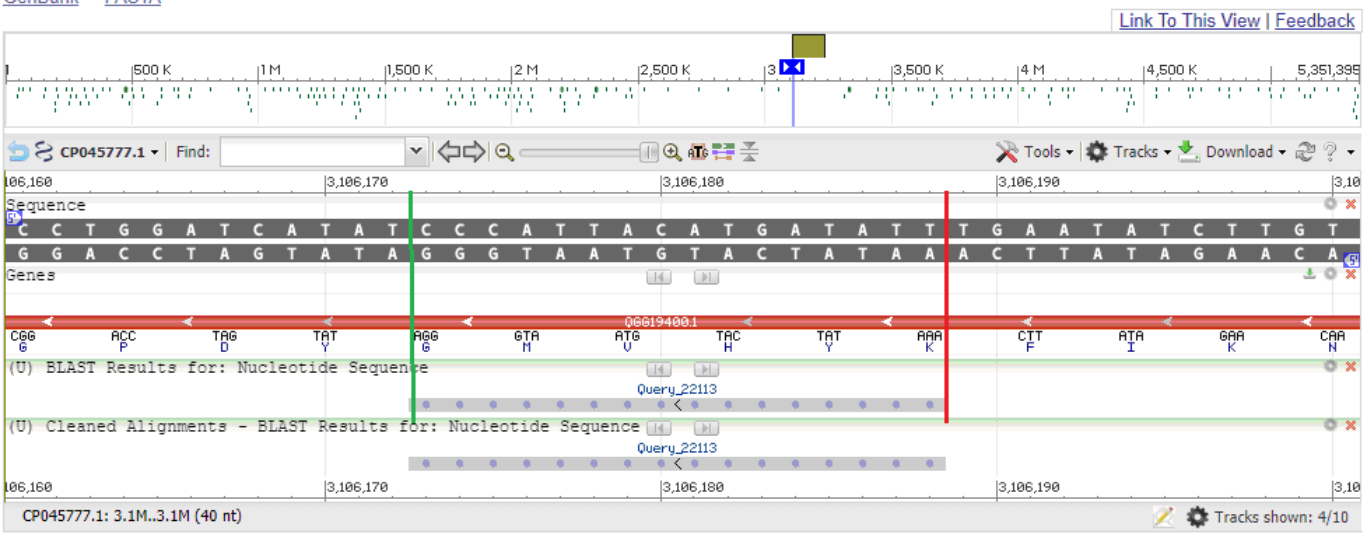
Score	Expect	Identities	Gaps	Strand
32.2 bits(16)	143	16/16(100%)	0/16(0%)	Plus/Minus

```
CDS: Putative 1       1           Y  H  V  M  G
Query                 1        AATATCATGTAATGGG  16
                               ||||||||||||||||
Sbjct                 3106188  AATATCATGTAATGGG  3106173
CDS:serine hydrolase  295      K  Y  H  V  M  G
```

Bacillus paranthracis strain CFSAN068816 chromosome, complete genome

GenBank: CP045777.1

GenBank FASTA



Range 2: 4177589 to 4177604 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
32.2 bits(16)	136	16/16(100%)	0/16(0%)	Plus/Minus

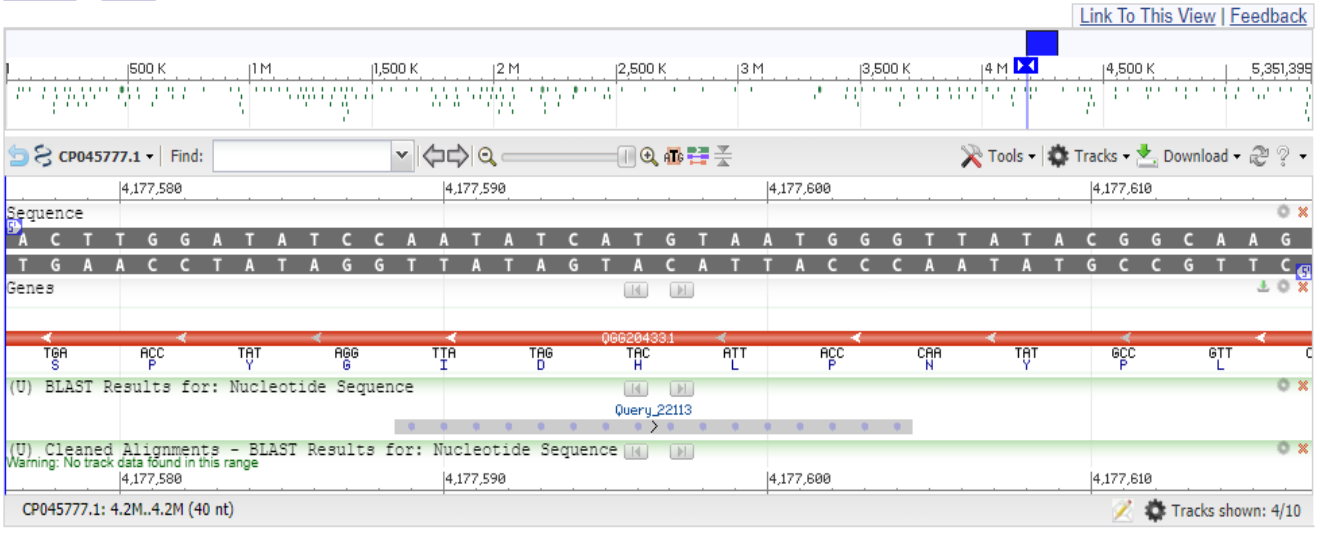
```
CDS: Putative 1       1           P  L  H  D  I

Query                 4         CCCATTACATGATATT  19
                                  ||||||||||||||||
Sbjct                 4177604   CCCATTACATGATATT  4177589
CDS:phosphoadenylyl-  192         P  L  H  D  I
```

Bacillus paranthracis strain CFSAN068816 chromosome, complete genome

GenBank: CP045777.1

GenBank FASTA



Bacillus anthracis PCr DNA, complete genome

GenBank: AP019731.1

FASTA Graphics

Go to:

```
LOCUS       AP019731                 19 bp    DNA     linear   BCT 13-JUN-2019
DEFINITION  Bacillus anthracis PCr DNA, complete genome.
ACCESSION   AP019731 REGION: 1368328..1368346
VERSION     AP019731.1
DBLINK      BioProject: PRJDB8367
            BioSample: SAMD00171951
KEYWORDS    .
SOURCE      Bacillus anthracis
  ORGANISM  Bacillus anthracis
            Bacteria; Firmicutes; Bacilli; Bacillales; Bacillaceae; Bacillus;
            Bacillus cereus group.
REFERENCE   1
  AUTHORS   Okutani,A., Inoue,S. and Morikawa,S.
  TITLE     Whole genome sequence of penicillin resistant Bacillus anthracis
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 19)
  AUTHORS   Okutani,A., Inoue,S. and Morikawa,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-JUN-2019) Contact:Akiko Okutani National Institute of
            Infectious Diseases, Department of Veterinary Science; Toyama
            1-23-1, Shinjuku, Tokyo 162-8640, Japan
COMMENT     ##Genome-Assembly-Data-START##
            Assembly Method       :: RS_HGAP_Assembly v. 3.0
            Genome Coverage       :: 188x
            Sequencing Technology :: PacBio RS II
            ##Genome-Assembly-Data-END##

FEATURES             Location/Qualifiers
     source          1..19
                     /organism="Bacillus anthracis"
                     /mol_type="genomic DNA"
                     /strain="PCr"
                     /isolation_source="bone powder"
                     /db_xref="taxon:1392"
                     /collection_date="1978"
     gene            <1..>19
                     /gene="pbpE_1"
                     /locus_tag="BAPCR_01478"
     CDS             <1..>19
                     /gene="pbpE_1"
                     /locus_tag="BAPCR_01478"
```

/note="beta-lactamase"
/codon_start=3
/transl_table=11
/product="**penicillin-binding protein 4**"
/protein_id="BBK95462.1"
/translation="MKTKDQIDNIVKETYRHIDFSGVILVKEKKGIVYEEVFGYANRN
ECINNTLQTRFGIASGCKIFTAIGICQLVENGVITFQTKLKECLSVNFPNFNEDITIH
HLLTHSSGIPDYFDESIMDNFEDLWKQTPMYLLKSLKDFLPLFQNRDMMYAPGSKFHY
NNAGFIILGLIIEEQTGLTFTEYIEKNIFNPIGMNHSGYFSLDRLPRQTALGYIKDEI
SQTWRTNAYSIPIKGGSDGGAFITASDMLKFWEALFNYEIISQEYTKILLTPHIQVNN
NQSYGYGIWIETRENKIF**KYHVMGY**DPGVSFRSAVYPDLGITLVIPSNKGAGAEKLMI
EIEGSF"

ORIGIN

1 **aatatcatgt aatgggata**

//

Bacillus anthracis PCr DNA, complete genome
Sequence ID: AP019731.1 **Length: 5209717 Number of Matches: 1**
Range 1: 1368328 to 1368346 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
38.2 bits(19)	2.2	19/19(100%)	0/19(0%)	Plus/Minus

```
CDS: Putative 1       5           G  M  V  H  Y
Query                 1        TATCCCATTACATGATATT  19
                               ||||||||||||||||||
Sbjct                 1368346  TATCCCATTACATGATATT  1368328
CDS:penicillin-bindi  301      Y  G  M  V  H  Y  K
```

Penicillin-binding protein 4

See when already discussed on page 286 on ***Serine hydrolase***.

Bacillus anthracis strain 17OD930 chromosome, complete genome

GenBank: CP029323.1

FASTA Graphics

Go to:

```
LOCUS       CP029323                 19 bp    DNA     linear   BCT 06-MAY-2020
DEFINITION  Bacillus anthracis strain 17OD930 chromosome, complete genome.
ACCESSION   CP029323 REGION: 2458980..2458998
VERSION     CP029323.1
DBLINK      BioProject: PRJNA454788
            BioSample: SAMN09041402
KEYWORDS    .
SOURCE      Bacillus anthracis
```

ORGANISM Bacillus anthracis
Bacteria; Firmicutes; Bacilli; Bacillales; Bacillaceae; Bacillus;
Bacillus cereus group.
REFERENCE 1 (bases 1 to 19)
AUTHORS Gobeli Brawand,S., Kittl,S., Dettwiler,M., Thomann,A., Feyer,S., Cachim,J., Theubet,G., Liechti,N., Wittwer,M., Schurch,N., Oberhansli,S., Heinimann,A. and Jores,J.
TITLE An unusual case of bovine anthrax in the canton of Jura, Switzerland in 2017
JOURNAL BMC Vet Res 15 (1), 265 (2019)
PUBMED 31357988
REMARK Publication Status: Online-Only
REFERENCE 2 (bases 1 to 19)
AUTHORS Kittl,S., Brodard,I. and Oberhaensli,S.
TITLE Direct Submission
JOURNAL **Submitted (03-MAY-2018) Department of Infectious Diseases and Pathobiology, Institute of Veterinary Bacteriology, University of Bern, Laengassstrasse 122, Bern 3012, Switzerland**
COMMENT Annotation was added by the NCBI Prokaryotic Genome Annotation Pipeline (released 2013). Information about the Pipeline can be found here: https://www.ncbi.nlm.nih.gov/genome/annotation_prok/

##Genome-Assembly-Data-START##
Assembly Method :: SPAdes v. 3.10.1
Genome Representation :: Full
Expected Final Version :: Yes
Genome Coverage :: 600.0x
Sequencing Technology :: Illumina MiSeq; Sanger
##Genome-Assembly-Data-END##
##Genome-Annotation-Data-START##
Annotation Provider :: NCBI
Annotation Date :: 05/10/2018 10:14:06
Annotation Pipeline :: NCBI Prokaryotic Genome Annotation Pipeline
Annotation Method :: Best-placed reference protein set; GeneMarkS+
Annotation Software revision :: 4.5
Features Annotated :: Gene; CDS; rRNA; tRNA; ncRNA; repeat_region
Genes (total) :: 6,083
CDS (total) :: 5,952
Genes (coding) :: 5,627
CDS (coding) :: 5,627
Genes (RNA) :: 131
rRNAs :: 10, 10, 10 (5S, 16S, 23S)

```
            complete rRNAs                    :: 10, 10, 10 (5S, 16S, 23S)
            tRNAs                             :: 96
            ncRNAs                            :: 5
            Pseudo Genes (total)              :: 325
            Pseudo Genes (ambiguous residues) :: 0 of 325
            Pseudo Genes (frameshifted)       :: 215 of 325
            Pseudo Genes (incomplete)         :: 67 of 325
            Pseudo Genes (internal stop)      :: 93 of 325
            Pseudo Genes (multiple problems)  :: 43 of 325
            ##Genome-Annotation-Data-END##
FEATURES             Location/Qualifiers
     source          1..19
                     /organism="Bacillus anthracis"
                     /mol_type="genomic DNA"
                     /strain="17OD930"
                     /isolation_source="placenta"
                     /host="Bos taurus"
                     /db_xref="taxon:1392"
                     /country="Switzerland: Kanton of Jura"
                     /lat_lon="47.42 N 7.08 E"
                     /collection_date="15-May-2017"
                     /collected_by="Institute of Veterinary Bacteriology,
                     University of Bern"
     gene            <1..>19
                     /locus_tag="DIR45_13585"
     CDS             <1..>19
                     /locus_tag="DIR45_13585"
                     /inference="COORDINATES: similar to AA
                     sequence:RefSeq:NP_845012.1"
                     /note="Derived by automated computational analysis using
                     gene prediction method: Protein Homology."
                     /codon_start=3
                     /transl_table=11
                     /product="penicillin-binding protein"
                     /protein_id="QCX56930.1"
                     /translation="MKTKDQIDNIVKETYRHIDFSGVILVKEKKGIVYEEVFGYANRN
                     ECINNTLQTRFGIASGCKIFTAIGICQLVENGVITFQTKLKECLSVNFPNFNEDITIH
                     HLLTHSSGIPDYFDESIMDNFEDLWKQTPMYLLKSLKDFLPLFQNRDMMYAPGSKFHY
                     NNAGFIILGLIIEEQTGLTFTEYIEKNIFNPIGMNHSGYFSLDRLPRQTALGYIKDEI
                     SQTWRTNAYSIPIKGGSDGGAFITASDMLKFWEALFNYEIISQEYTKILLTPHIQVNN
                     NQSYGYGIWIETRENKIFKYHVMGYDPGVSFRSAVYPDLGITLVIPSNKGAGAEKLMI
                     EIEGSF"
ORIGIN
        1 aatatcatgt aatgggata
//
```

Bacillus anthracis strain 17OD930 chromosome, complete genome
Sequence ID: CP029323.1 **Length: 5213297 Number of Matches: 1**
Range 1: 2458980 to 2458998 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
38.2 bits(19)	2.2	19/19(100%)	0/19(0%)	Plus/Minus

```
CDS: Putative 1       5           G  M  V  H  Y
Query                 1        TATCCCATTACATGATATT  19
                               ||||||||||||||||||
Sbjct                 2458998  TATCCCATTACATGATATT  2458980
CDS:penicillin-bindi  301      Y  G  M  V  H  Y  K
```

Bacillus anthracis strain HDZK-BYSB7 chromosome, complete genome

GenBank: CP026608.1

FASTA Graphics

Go to:

```
LOCUS       CP026608                  19 bp    DNA     linear   BCT 12-AUG-2020
DEFINITION  Bacillus anthracis strain HDZK-BYSB7 chromosome, complete genome.
ACCESSION   CP026608 REGION: 3594115..3594133
VERSION     CP026608.1
DBLINK      BioProject: PRJNA432651
            BioSample: SAMN08456585
KEYWORDS    .
SOURCE      Bacillus anthracis
  ORGANISM  Bacillus anthracis
            Bacteria; Firmicutes; Bacilli; Bacillales; Bacillaceae; Bacillus;
            Bacillus cereus group.
REFERENCE   1  (bases 1 to 19)
  AUTHORS   Duan,J. and Zhao,K.
  TITLE     Complete genome sequence of Bacillus anthracis HDZK-BYSB7
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 19)
  AUTHORS   Duan,J. and Zhao,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-FEB-2018) School of Life Sciences, Heilongjiang
            University, Xuefu Road, Haerbin 150080, China
COMMENT     Annotation was added by the NCBI Prokaryotic Genome Annotation
            Pipeline (released 2013). Information about the Pipeline can be
            found here: https://www.ncbi.nlm.nih.gov/genome/annotation_prok/

            ##Genome-Assembly-Data-START##
            Assembly Method       :: FALCON v. 4.0
            Genome Representation :: Full
```

```
            Expected Final Version :: Yes
            Genome Coverage        :: 100.0x
            Sequencing Technology  :: PacBio
            ##Genome-Assembly-Data-END##

            ##Genome-Annotation-Data-START##
            Annotation Provider               :: NCBI
            Annotation Date                   :: 02/07/2018 14:03:23
            Annotation Pipeline               :: NCBI Prokaryotic Genome
                                                 Annotation Pipeline
            Annotation Method                 :: Best-placed reference protein
                                                 set; GeneMarkS+
            Annotation Software revision      :: 4.4
            Features Annotated                :: Gene; CDS; rRNA; tRNA; ncRNA;
                                                 repeat_region
            Genes (total)                     :: 5,695
            CDS (total)                       :: 5,544
            Genes (coding)                    :: 5,315
            CDS (coding)                      :: 5,315
            Genes (RNA)                       :: 151
            rRNAs                             :: 14, 14, 14 (5S, 16S, 23S)
            complete rRNAs                    :: 14, 14, 14 (5S, 16S, 23S)
            tRNAs                             :: 104
            ncRNAs                            :: 5
            Pseudo Genes (total)              :: 229
            Pseudo Genes (ambiguous residues) :: 0 of 229
            Pseudo Genes (frameshifted)       :: 120 of 229
            Pseudo Genes (incomplete)         :: 85 of 229
            Pseudo Genes (internal stop)      :: 77 of 229
            Pseudo Genes (multiple problems)  :: 47 of 229
            ##Genome-Annotation-Data-END##
FEATURES             Location/Qualifiers
     source          1..19
                     /organism="Bacillus anthracis"
                     /mol_type="genomic DNA"
                     /strain="HDZK-BYSB7"
                     /isolation_source="bark of cypress"
                     /db_xref="taxon:1392"
                     /country="USA"
                     /collection_date="2016"
                     /note="Submitted as B. subtilis but found to be Bacillus
                     anthracis on the basis of ANI results"
     gene            <1..>19
                     /locus_tag="C3496_19285"
     CDS             <1..>19
```

/locus_tag="C3496_19285"
/inference="COORDINATES: similar to AA
sequence:RefSeq:YP_036759.1"
/note="Derived by automated computational analysis using
gene prediction method: Protein Homology."
/codon_start=3
/transl_table=11
/product="**penicillin-binding protein**"
/protein_id="QBJ70186.1"
/translation="MKTKDQIDNIVKETYRHIDFSGVILVKEKKGIVYEEAFGYANRN
ECINNTLQTRFGIASGCKIFTAIGICQLVENGVITFQTKLKECLSVNFPNFNEDITIH
HLLTHSSGIPDYFDESIMDNFEDLWKQTPMYLLKSLKDFLPLFQNRYMMYAPGSKFHY
NNAGFIILGLIIEEQTGLKFTEYIEKNIFNPIGMNHSGYFSLDRLPRQTALGYIKDEI
SQTWRTNAYSIPIKGGSDGGAFITASDMLKFWEALFNYEIISQEYTKILLTPHIQVNN
NQSYGYGIWIETRENKVF**KYHVMGY**DPGVSFRSAVYPDLGITLVIPSNKGAGPEKLMI
EIEGAF"

ORIGIN

1 **aatatcatgt aatgggata**

//

Bacillus anthracis strain HDZK-BYSB7 chromosome, complete genome
Sequence ID: CP026608.1 **Length: 5294779 Number of Matches: 2**
Range 1: 3594115 to 3594133 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
38.2 bits(19)	2.2	19/19(100%)	0/19(0%)	Plus/Minus

Features: single-stranded-DNA-specific exonuclease RecJpenicillin-binding protein

```
CDS: Putative 1       5           G  M  V  H  Y
Query                 1        TATCCCATTACATGATATT  19
                               ||||||||||||||||||
Sbjct                 3594133  TATCCCATTACATGATATT  3594115
CDS:penicillin-bindi  301      Y  G  M  V  H  Y  K
```

Bacillus anthracis strain HDZK-BYSB7 chromosome, complete genome

GenBank: CP026608.1

FASTA Graphics

Go to:

LOCUS CP026608 16 bp DNA linear **BCT** **12-AUG-2020**
DEFINITION Bacillus anthracis strain HDZK-BYSB7 chromosome, complete genome.
ACCESSION CP026608 REGION: 2431380..2431395

VERSION CP026608.1
DBLINK BioProject: PRJNA432651
BioSample: SAMN08456585
KEYWORDS .
SOURCE Bacillus anthracis
ORGANISM Bacillus anthracis
Bacteria; Firmicutes; Bacilli; Bacillales; Bacillaceae; Bacillus;
Bacillus cereus group.
REFERENCE 1 (bases 1 to 16)
AUTHORS Duan,J. and Zhao,K.
TITLE Complete genome sequence of Bacillus anthracis HDZK-BYSB7
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 16)
AUTHORS Duan,J. and Zhao,K.
TITLE Direct Submission
JOURNAL **Submitted (06-FEB-2018) School of Life Sciences, Heilongjiang University, Xuefu Road, Haerbin 150080, China**
COMMENT Annotation was added by the NCBI Prokaryotic Genome Annotation Pipeline (released 2013). Information about the Pipeline can be found here: https://www.ncbi.nlm.nih.gov/genome/annotation_prok/

##Genome-Assembly-Data-START##
Assembly Method :: FALCON v. 4.0
Genome Representation :: Full
Expected Final Version :: Yes
Genome Coverage :: 100.0x
Sequencing Technology :: PacBio
##Genome-Assembly-Data-END##

##Genome-Annotation-Data-START##
Annotation Provider :: NCBI
Annotation Date :: 02/07/2018 14:03:23
Annotation Pipeline :: NCBI Prokaryotic Genome Annotation Pipeline
Annotation Method :: Best-placed reference protein set; GeneMarkS+
Annotation Software revision :: 4.4
Features Annotated :: Gene; CDS; rRNA; tRNA; ncRNA; repeat_region
Genes (total) :: 5,695
CDS (total) :: 5,544
Genes (coding) :: 5,315
CDS (coding) :: 5,315
Genes (RNA) :: 151

```
            rRNAs                              :: 14, 14, 14 (5S, 16S, 23S)
            complete rRNAs                     :: 14, 14, 14 (5S, 16S, 23S)
            tRNAs                              :: 104
            ncRNAs                             :: 5
            Pseudo Genes (total)               :: 229
            Pseudo Genes (ambiguous residues)  :: 0 of 229
            Pseudo Genes (frameshifted)        :: 120 of 229
            Pseudo Genes (incomplete)          :: 85 of 229
            Pseudo Genes (internal stop)       :: 77 of 229
            Pseudo Genes (multiple problems)   :: 47 of 229
            ##Genome-Annotation-Data-END##
FEATURES             Location/Qualifiers
     source          1..16
                     /organism="Bacillus anthracis"
                     /mol_type="genomic DNA"
                     /strain="HDZK-BYSB7"
                     /isolation_source="bark of cypress"
                     /db_xref="taxon:1392"
                     /country="USA"
                     /collection_date="2016"
                     /note="Submitted as B. subtilis but found to be Bacillus
                     anthracis on the basis of ANI results"
     gene            <1..>16
                     /locus_tag="C3496_13100"
     CDS             <1..>16
                     /locus_tag="C3496_13100"
                     /inference="COORDINATES: similar to AA
                     sequence:RefSeq:YP_035637.1"
                     /note="Derived by automated computational analysis using
                     gene prediction method: Protein Homology."
                     /codon_start=2
                     /transl_table=11
                     /product="phosphoadenylyl-sulfate reductase"
                     /protein_id="QBJ67254.1"
                     /translation="MLTYETWEENNVSFSEEDETKGALSVLSWAYKEYEGEIVYACSF
                     GVEGMVLLHLINQVNPSAKVVFLDTNVHFQETYELIQKVRERFPSLNIIEKQPKLTLD
                     EQAKLHGNKLWESNPNLCCKIRKILPLEESLANEKAWISGLRREQSETRKHTKFINQD
                     HRFQSIKVCPLIHWTWKEVWRYVYKHSLPYNPLHDIGYPSIGCEKCTLPVGEGGDSRD
                     GRWAGKVKTECGLHYQ"
ORIGIN
        1 cccattacat gatatt
//
```

Range 2: 2431380 to 2431395 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
32.2 bits(16)	136	16/16(100%)	0/16(0%)	Plus/Plus

Features: single-stranded-DNA-specific exonuclease RecJphosphoadenylyl-sulfate reductase

```
CDS: Putative 1       1          P  L  H  D  I
Query                 4        CCCATTACATGATATT  19
                               ||||||||||||||||
Sbjct                 2431380  CCCATTACATGATATT  2431395
CDS:phosphoadenylyl-  192        P  L  H  D  I
```

Bacillus anthracis CZC5 DNA, complete genome

GenBank: AP018443.1

FASTA Graphics

Go to:

```
LOCUS       AP018443                  19 bp    DNA     linear   BCT 06-DEC-2018
DEFINITION  Bacillus anthracis CZC5 DNA, complete genome.
ACCESSION   AP018443 REGION: 2463006..2463024
VERSION     AP018443.1
DBLINK      BioProject: PRJDB1571
            BioSample: SAMD00002586
            Sequence Read Archive: DRR014735, DRR118320
KEYWORDS    .
SOURCE      Bacillus anthracis
  ORGANISM  Bacillus anthracis
            Bacteria; Firmicutes; Bacilli; Bacillales; Bacillaceae; Bacillus;
            Bacillus cereus group.
REFERENCE   1
  AUTHORS   Furuta,Y., Harima,H., Ito,E., Maruyama,F., Ohnishi,N., Osaki,K.,
            Ogawa,H., Squarre,D., Hang'ombe,B.M., Higashi,H. and Yang,Y.-L.
  TITLE     Loss of Bacitracin Resistance Due to a Large Genomic Deletion among
            Bacillus anthracis Strains
  JOURNAL   mSystems 3 (5), e00182-18 (2018)
   PUBMED   30417107
  REMARK    DOI:10.1128/mSystems.00182-18
            Publication Status: Online-Only
REFERENCE   2  (bases 1 to 19)
  AUTHORS   Furuta,Y., Ohnishi,N., Maruyama,F., Osaki,K., Squarre,D.,
            Hang'ombe,M.B. and Higashi,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-DEC-2017) Contact:Hideaki Higashi Hokkaido
            University, Research Center for Zoonosis Control; North 20, West
            10, Kita-ku, Sapporo, Hokkaido 001-0020, Japan
COMMENT     ##Genome-Assembly-Data-START##
```

```
            Assembly Method       :: Pilon v. 1.22
            Genome Coverage       :: 25x
            Sequencing Technology :: PacBio RS, Illumina MiSeq
            ##Genome-Assembly-Data-END##
FEATURES             Location/Qualifiers
     source          1..19
                     /organism="Bacillus anthracis"
                     /mol_type="genomic DNA"
                     /strain="CZC5"
                     /host="Hippopotamus amphibius"
                     /db_xref="taxon:1392"
                     /country="Zambia"
     gene            <1..>19
                     /gene="pbpE_1"
                     /locus_tag="BAZ_02616"
     CDS             <1..>19
                     /gene="pbpE_1"
                     /locus_tag="BAZ_02616"
                     /inference="ab initio prediction:Prodigal:2.6"
                     /codon_start=3
                     /transl_table=11
                     /product="penicillin-binding protein 4*"
                     /protein_id="BBB72799.1"
                     /translation="MKTKDQIDNIVKETYRHIDFSGVILVKEKKGIVYEEVFGYANRN
                     ECINNTLQTRFGIASGCKIFTAIGICQLVENGVITFQTKLKECLSVNFPNFNEDITIH
                     HLLTHSSGIPDYFDESIMDNFEDLWKQTPMYLLKSLKDFLPLFQNRDMMYAPGSKFHY
                     NNAGFIILGLIIEEQTGLTFTEYIEKNIFNPIGMNHSGYFSLDRLPRQTALGYIKDEI
                     SQTWRTNAYSIPIKGGSDGGAFITASDMLKFWEALFNYEIISQEYTKILLTPHIQVNN
                     NQSYGYGIWIETRENKIFKYHVMGYDPGVSFRSAVYPDLGITLVIPSNKGAGAEKLMI
                     EIEGSF"
ORIGIN
        1 aatatcatgt aatgggata
//
```

Bacillus anthracis CZC5 DNA, complete genome
Sequence ID: AP018443.1 **Length: 5219278 Number of Matches: 1**
Range 1: 2463006 to 2463024 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
38.2 bits(19)	2.2	19/19(100%)	0/19(0%)	Plus/Minus

```
CDS: Putative 1       5           G  M  V  H  Y
Query                 1        TATCCCATTACATGATATT  19
                               ||||||||||||||||||
Sbjct                 2463024  TATCCCATTACATGATATT  2463006
CDS:penicillin-bindi  301      Y  G  M  V  H  Y  K
```

Bacillus anthracis strain MCCC 1A02161 chromosome, complete genome

GenBank: CP031642.1

FASTA Graphics

Go to:

```
LOCUS       CP031642                 19 bp    DNA     linear   BCT 21-AUG-2018
DEFINITION  Bacillus anthracis strain MCCC 1A02161 chromosome, complete genome.
ACCESSION   CP031642 REGION: 2575637..2575655
VERSION     CP031642.1
DBLINK      BioProject: PRJNA485400
            BioSample: SAMN09789251
KEYWORDS    .
SOURCE      Bacillus anthracis
  ORGANISM  Bacillus anthracis
            Bacteria; Firmicutes; Bacilli; Bacillales; Bacillaceae; Bacillus;
            Bacillus cereus group.
REFERENCE   1  (bases 1 to 19)
  AUTHORS   Liu,Y.
  TITLE     The complete genome sequence of Bacillus anthracis MCCC 1A02161
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 19)
  AUTHORS   Liu,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-AUG-2018) State of Oceanic Administration, The Third
            Institute of Oceanography, Key Laboratory of Marine Biogenetic
            Resources, Room 116, Daxue Road 178#, Xiamen, Fujian 361005, China
COMMENT     Annotation was added by the NCBI Prokaryotic Genome Annotation
            Pipeline (released 2013). Information about the Pipeline can be
            found here: https://www.ncbi.nlm.nih.gov/genome/annotation_prok/
```

```
            ##Genome-Assembly-Data-START##
            Assembly Method        :: HGAP v. 2.3
            Genome Representation  :: Full
            Expected Final Version :: Yes
            Genome Coverage        :: 250.0x
            Sequencing Technology  :: PacBio RSII
            ##Genome-Assembly-Data-END##


            ##Genome-Annotation-Data-START##
            Annotation Provider               :: NCBI
            Annotation Date                   :: 08/15/2018 11:14:57
            Annotation Pipeline               :: NCBI Prokaryotic Genome
                                                 Annotation Pipeline
            Annotation Method                 :: Best-placed reference protein
                                                 set; GeneMarkS+
            Annotation Software revision      :: 4.6
            Features Annotated                :: Gene; CDS; rRNA; tRNA; ncRNA;
                                                 repeat_region
            Genes (total)                     :: 5,529
            CDS (total)                       :: 5,378
            Genes (coding)                    :: 5,175
            CDS (coding)                      :: 5,175
            Genes (RNA)                       :: 151
            rRNAs                             :: 14, 14, 14 (5S, 16S, 23S)
            complete rRNAs                    :: 14, 14, 14 (5S, 16S, 23S)
            tRNAs                             :: 104
            ncRNAs                            :: 5
            Pseudo Genes (total)              :: 203
            Pseudo Genes (ambiguous residues) :: 0 of 203
            Pseudo Genes (frameshifted)       :: 106 of 203
            Pseudo Genes (incomplete)         :: 80 of 203
            Pseudo Genes (internal stop)      :: 61 of 203
            Pseudo Genes (multiple problems)  :: 36 of 203
            ##Genome-Annotation-Data-END##
FEATURES             Location/Qualifiers
     source          1..19
                     /organism="Bacillus anthracis"
                     /mol_type="genomic DNA"
                     /strain="MCCC 1A02161"
                     /isolation_source="sediment"
                     /culture_collection="MCCC:1A02161"
                     /db_xref="taxon:1392"
                     /country="China: South China Sea"
                     /collection_date="2002-10"
```

```
     gene            <1..>19
                     /locus_tag="DY470_13615"
     CDS             <1..>19
                     /locus_tag="DY470_13615"
                     /inference="COORDINATES: similar to AA
                     sequence:RefSeq:WP_000849936.1"
                     /note="Derived by automated computational analysis using
                     gene prediction method: Protein Homology."
                     /codon_start=3
                     /transl_table=11
                     /product="class C beta-lactamase-related serine hydrolase"
                     /protein_id="AXO98692.1"
                     /translation="MKTKDQIDNIVKETYRHIDFSGVILVKEKKGIVYEEAFGYANRN
                     ECINNTLQTRFGIASGCKIFTAIGICQLVENGVITFQTKLKECLSVNFPNFNEDITIH
                     HLLTHSSGIPDYYDESIMDNFEDLWKQTPMYLLKSLKDFLPLFQNRDMMYAPGSKFHY
                     NNAGFIILGLIIEEQTGLTFTEYIEKNIFNPIGMNHSGYFSLDRLPRQTALGYIKDEI
                     TQTWRTNAYSIPIKGGSDGGAFITASDMLKFWEALFNYEIISQEYTKILLTPHIQVNN
                     NQSYGYGIWIETRENKIFKYHVMGYDPGVSFRSAVYPDLGITLVIPSNKEAGPEKLMI
                     EIEGDF"
ORIGIN
        1 aatatcatgt aatgggata
//
```

Bacillus anthracis strain MCCC 1A02161 chromosome, complete genome
Sequence ID: CP031642.1 **Length: 5231857 Number of Matches: 2**
Range 1: 2575637 to 2575655 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
38.2 bits(19)	2.2	19/19(100%)	0/19(0%)	Plus/Minus

```
CDS: Putative 1       5           G  M  V  H  Y
Query                 1        TATCCCATTACATGATATT  19
                               ||||||||||||||||||
Sbjct                 2575655  TATCCCATTACATGATATT  2575637
CDS:class C beta-lac  301      Y  G  M  V  H  Y  K
```

Class C beta-lactamase-related serine hydrolase

See when already discussed on page 286 on ***Serine hydrolase***.

Bacillus anthracis strain MCCC 1A02161 chromosome, complete genome

GenBank: CP031642.1

FASTA Graphics

Go to:

```
LOCUS       CP031642                16 bp    DNA     linear   BCT 21-AUG-2018
DEFINITION  Bacillus anthracis strain MCCC 1A02161 chromosome, complete genome.
ACCESSION   CP031642 REGION: 1399303..1399318
VERSION     CP031642.1
DBLINK      BioProject: PRJNA485400
            BioSample: SAMN09789251
KEYWORDS    .
SOURCE      Bacillus anthracis
  ORGANISM  Bacillus anthracis
            Bacteria; Firmicutes; Bacilli; Bacillales; Bacillaceae; Bacillus;
            Bacillus cereus group.
REFERENCE   1  (bases 1 to 16)
  AUTHORS   Liu,Y.
  TITLE     The complete genome sequence of Bacillus anthracis MCCC 1A02161
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 16)
  AUTHORS   Liu,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-AUG-2018) State of Oceanic Administration, The Third
            Institute of Oceanography, Key Laboratory of Marine Biogenetic
            Resources, Room 116, Daxue Road 178#, Xiamen, Fujian 361005, China
COMMENT     Annotation was added by the NCBI Prokaryotic Genome Annotation
            Pipeline (released 2013). Information about the Pipeline can be
            found here: https://www.ncbi.nlm.nih.gov/genome/annotation_prok/

            ##Genome-Assembly-Data-START##
            Assembly Method        :: HGAP v. 2.3
            Genome Representation  :: Full
            Expected Final Version :: Yes
            Genome Coverage        :: 250.0x
            Sequencing Technology  :: PacBio RSII
            ##Genome-Assembly-Data-END##

            ##Genome-Annotation-Data-START##
            Annotation Provider               :: NCBI
            Annotation Date                   :: 08/15/2018 11:14:57
            Annotation Pipeline               :: NCBI Prokaryotic Genome
                                                 Annotation Pipeline
            Annotation Method                 :: Best-placed reference protein
                                                 set; GeneMarkS+
            Annotation Software revision      :: 4.6
```

```
            Features Annotated                :: Gene; CDS; rRNA; tRNA; ncRNA;
                                                 repeat_region
            Genes (total)                     :: 5,529
            CDS (total)                       :: 5,378
            Genes (coding)                    :: 5,175
            CDS (coding)                      :: 5,175
            Genes (RNA)                       :: 151
            rRNAs                             :: 14, 14, 14 (5S, 16S, 23S)
            complete rRNAs                    :: 14, 14, 14 (5S, 16S, 23S)
            tRNAs                             :: 104
            ncRNAs                            :: 5
            Pseudo Genes (total)              :: 203
            Pseudo Genes (ambiguous residues) :: 0 of 203
            Pseudo Genes (frameshifted)       :: 106 of 203
            Pseudo Genes (incomplete)         :: 80 of 203
            Pseudo Genes (internal stop)      :: 61 of 203
            Pseudo Genes (multiple problems)  :: 36 of 203

            ##Genome-Annotation-Data-END##
FEATURES             Location/Qualifiers
     source          1..16
                     /organism="Bacillus anthracis"
                     /mol_type="genomic DNA"
                     /strain="MCCC 1A02161"
                     /isolation_source="sediment"
                     /culture_collection="MCCC:1A02161"
                     /db_xref="taxon:1392"
                     /country="China: South China Sea"
                     /collection_date="2002-10"
     gene            <1..>16
                     /locus_tag="DY470_07495"
     CDS             <1..>16
                     /locus_tag="DY470_07495"
                     /EC_number="1.8.4.8"
                     /inference="COORDINATES: similar to AA
                     sequence:RefSeq:WP_000958995.1"
                     /note="Derived by automated computational analysis using
                     gene prediction method: Protein Homology."
                     /codon_start=2
                     /transl_table=11
                     /product="phosphoadenylyl-sulfate reductase"
                     /protein_id="AXO97569.1"
                     /translation="MLTYETWEENDVSFSEEDETKGALSVLSWAYKEYKSEIVYACSF
                     GVEGMVLLHLINQVNPSAKVVFLDTNVHFQETYELIQKVRERFPSLNIIEKQPKLTLD
                     EQDKLHGDKLWESNPNLCCKIRKILPLEESLANEKAWISGLRREQSETRKHTKFINQD
```

HRFQSIKVCPLIHWTWKEVWRYVYKHSLPYN**PLHDI**GYPSIGCEKCTLPVGEGGDSRD

GRWAGKVKTECGLHYQ"

ORIGIN

1 **cccattacat gatatt**

//

Range 2: 1399303 to 1399318 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
32.2 bits(16)	136	16/16(100%)	0/16(0%)	Plus/Plus

```
CDS: Putative 1       1          P  L  H  D  I
Query                 4        CCCATTACATGATATT  19
                               ||||||||||||||||
Sbjct                 1399303  CCCATTACATGATATT  1399318
CDS:phosphoadenylyl-  192        P  L  H  D  I
```

Bacillus anthracis strain MCCC 1A01412 chromosome, complete genome

GenBank: CP031643.1

FASTA Graphics

Go to:

LOCUS CP031643 19 bp DNA linear BCT **21-AUG-2018**

DEFINITION Bacillus anthracis strain MCCC 1A01412 chromosome, complete genome.

ACCESSION CP031643 REGION: 2546553..2546571

VERSION CP031643.1

DBLINK BioProject: PRJNA485401

BioSample: SAMN09789252

KEYWORDS .

SOURCE Bacillus anthracis

ORGANISM Bacillus anthracis

Bacteria; Firmicutes; Bacilli; Bacillales; Bacillaceae; Bacillus;

Bacillus cereus group.

REFERENCE 1 (bases 1 to 19)

AUTHORS Liu,Y.

TITLE The complete genome sequence of Bacillus anthracis MCCC 1A01412

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 19)

AUTHORS Liu,Y.

TITLE Direct Submission

JOURNAL **Submitted (10-AUG-2018) State of Oceanic Administration, The Third Institute of Oceanography, Key Laboratory of Marine Biogenetic Resources, Room 116, Daxue Road 178#, Xiamen, Fujian 361005, China**

```
COMMENT     Annotation was added by the NCBI Prokaryotic Genome Annotation
            Pipeline (released 2013). Information about the Pipeline can be
            found here: https://www.ncbi.nlm.nih.gov/genome/annotation_prok/
            ##Genome-Assembly-Data-START##
            Assembly Method        :: HGAP v. 2.3
            Genome Representation  :: Full
            Expected Final Version :: Yes
            Genome Coverage        :: 250.0x
            Sequencing Technology  :: PacBio RSII
            ##Genome-Assembly-Data-END##
            ##Genome-Annotation-Data-START##
            Annotation Provider               :: NCBI
            Annotation Date                   :: 08/15/2018 12:21:11
            Annotation Pipeline               :: NCBI Prokaryotic Genome
                                                 Annotation Pipeline
            Annotation Method                 :: Best-placed reference protein
                                                 set; GeneMarkS+
            Annotation Software revision      :: 4.6
            Features Annotated                :: Gene; CDS; rRNA; tRNA; ncRNA;
                                                 repeat_region
            Genes (total)                     :: 5,677
            CDS (total)                       :: 5,526
            Genes (coding)                    :: 5,314
            CDS (coding)                      :: 5,314
            Genes (RNA)                       :: 151
            rRNAs                             :: 14, 14, 14 (5S, 16S, 23S)
            complete rRNAs                    :: 14, 14, 14 (5S, 16S, 23S)
            tRNAs                             :: 104
            ncRNAs                            :: 5
            Pseudo Genes (total)              :: 212
            Pseudo Genes (ambiguous residues) :: 0 of 212
            Pseudo Genes (frameshifted)       :: 109 of 212
            Pseudo Genes (incomplete)         :: 76 of 212
            Pseudo Genes (internal stop)      :: 79 of 212
            Pseudo Genes (multiple problems)  :: 44 of 212
            ##Genome-Annotation-Data-END##

FEATURES             Location/Qualifiers
     source          1..19
                     /organism="Bacillus anthracis"
                     /mol_type="genomic DNA"
                     /strain="MCCC 1A01412"
                     /isolation_source="sediment"
                     /culture_collection="MCCC:1A01412"
                     /db_xref="taxon:1392"
```

/country="China: South China Sea"
/collection_date="2002-10"

gene <1..>19
/locus_tag="DY471_13335"

CDS <1..>19
/locus_tag="DY471_13335"
/inference="COORDINATES: similar to AA sequence:RefSeq:WP_000849936.1"
/note="Derived by automated computational analysis using gene prediction method: Protein Homology."
/codon_start=3
/transl_table=11
/product="class C beta-lactamase-related serine hydrolase"
/protein_id="AXO96097.1"
/translation="MKTKDQIDNIVKETYRHIDFSGVILVKEKKGIVYEEAFGYANRN
ECINNTLQTRFGIASGCKIFTAIGICQLVENGVITFQTKLKECLSVNFPNFNEDITIH
HLLTHSSGIPDYFDESIMDNFEDLWKQTPMYLLKSLKDFLPLFQNRDMMYAPGSKFHY
NNAGFIILGLIIEEQTGLKFTEYIEKNIFNPIGMNHSGYFSLDRLPRQTALGYIKDEI
SQTWRTNAYSIPIKGGSDGGAFITAPDMLKFWEALFNYEIISQEYTKILLTPHIQVNN
NQSYGYGIWIETRENKVF**KYHVMGY**DPGVSFRSAVYPDLGITLVIPSNKGAGPEKLMI
EIEGAF"

ORIGIN
1 **aatatcatgt aatgggata**
//

Bacillus anthracis strain MCCC 1A01412 chromosome, complete genome
Sequence ID: CP031643.1 **Length: 5291783 Number of Matches: 2**
Range 1: 2546553 to 2546571 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
38.2 bits(19)	2.2	19/19(100%)	0/19(0%)	Plus/Minus

```
CDS: Putative 1       5           G  M  V  H  Y
Query                 1        TATCCCATTACATGATATT  19
                               ||||||||||||||||||
Sbjct                 2546571  TATCCCATTACATGATATT  2546553
CDS:class C beta-lac  301      Y  G  M  V  H  Y  K
```

Bacillus anthracis strain MCCC 1A01412 chromosome, complete genome

GenBank: CP031643.1

FASTA Graphics

Go to:

LOCUS CP031643 16 bp DNA linear BCT **21-AUG-2018**

DEFINITION Bacillus anthracis strain MCCC 1A01412 chromosome, complete genome.

ACCESSION CP031643 REGION: 1369983..1369998

VERSION CP031643.1

DBLINK BioProject: PRJNA485401
BioSample: SAMN09789252

KEYWORDS .

SOURCE Bacillus anthracis

ORGANISM Bacillus anthracis
Bacteria; Firmicutes; Bacilli; Bacillales; Bacillaceae; Bacillus;
Bacillus cereus group.

REFERENCE 1 (bases 1 to 16)

AUTHORS Liu,Y.

TITLE The complete genome sequence of Bacillus anthracis MCCC 1A01412

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 16)

AUTHORS Liu,Y.

TITLE Direct Submission

JOURNAL **Submitted (10-AUG-2018) State of Oceanic Administration, The Third Institute of Oceanography, Key Laboratory of Marine Biogenetic Resources, Room 116, Daxue Road 178#, Xiamen, Fujian 361005, China**

COMMENT Annotation was added by the NCBI Prokaryotic Genome Annotation Pipeline (released 2013). Information about the Pipeline can be found here: https://www.ncbi.nlm.nih.gov/genome/annotation_prok/

```
##Genome-Assembly-Data-START##
Assembly Method        :: HGAP v. 2.3
Genome Representation  :: Full
Expected Final Version :: Yes
Genome Coverage        :: 250.0x
Sequencing Technology  :: PacBio RSII
##Genome-Assembly-Data-END##
##Genome-Annotation-Data-START##
Annotation Provider               :: NCBI
Annotation Date                   :: 08/15/2018 12:21:11
Annotation Pipeline               :: NCBI Prokaryotic Genome
                                     Annotation Pipeline
Annotation Method                 :: Best-placed reference protein
                                     set; GeneMarkS+
Annotation Software revision      :: 4.6
```

```
            Features Annotated                :: Gene; CDS; rRNA; tRNA; ncRNA;
                                                 repeat_region
            Genes (total)                     :: 5,677
            CDS (total)                       :: 5,526
            Genes (coding)                    :: 5,314
            CDS (coding)                      :: 5,314
            Genes (RNA)                       :: 151
            rRNAs                             :: 14, 14, 14 (5S, 16S, 23S)
            complete rRNAs                    :: 14, 14, 14 (5S, 16S, 23S)
            tRNAs                             :: 104
            ncRNAs                            :: 5
            Pseudo Genes (total)              :: 212
            Pseudo Genes (ambiguous residues) :: 0 of 212
            Pseudo Genes (frameshifted)       :: 109 of 212
            Pseudo Genes (incomplete)         :: 76 of 212
            Pseudo Genes (internal stop)      :: 79 of 212
            Pseudo Genes (multiple problems)  :: 44 of 212
            ##Genome-Annotation-Data-END##

FEATURES             Location/Qualifiers
     source          1..16
                     /organism="Bacillus anthracis"
                     /mol_type="genomic DNA"
                     /strain="MCCC 1A01412"
                     /isolation_source="sediment"
                     /culture_collection="MCCC:1A01412"
                     /db_xref="taxon:1392"
                     /country="China: South China Sea"
                     /collection_date="2002-10"
     gene            <1..>16
                     /locus_tag="DY471_07220"
     CDS             <1..>16
                     /locus_tag="DY471_07220"
                     /EC_number="1.8.4.8"
                     /inference="COORDINATES: similar to AA
                     sequence:RefSeq:WP_000958995.1"
                     /note="Derived by automated computational analysis using
                     gene prediction method: Protein Homology."
                     /codon_start=2
                     /transl_table=11
                     /product="phosphoadenylyl-sulfate reductase"
                     /protein_id="AXO92209.1"
                     /translation="MLTYETWEENNVSFSEEDETKGALSVLSWAYKEYEGEIVYACSF
                     GVEGMVLLHLINQVNPSAKVVFLDTNVHFQETYELIQKVRERFPSLNIIEKQPKLTLD
                     EQAKLHGNKLWESNPNLCCKIRKILPLEESLANEKAWISGLRREQSETRKHTKFINQD
```

HRFQSIKVCPLIHWTWKEVWRYVYKHSLPYN**PLHDI**GYPSIGCEKCTLPVGEGGDSRD

GRWAGKVKTECGLHYQ"

ORIGIN

1 **cccattacat gatatt**

//

Range 2: 1369983 to 1369998 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
32.2 bits(16)	136	16/16(100%)	0/16(0%)	Plus/Plus

```
CDS: Putative 1       1          P  L  H  D  I
Query                 4        CCCATTACATGATATT  19
                               ||||||||||||||||
Sbjct                 1369983  CCCATTACATGATATT  1369998
CDS:phosphoadenylyl-  192        P  L  H  D  I
```

Bacillus anthracis strain London_499 chromosome, complete genome

GenBank: CP029805.1

FASTA Graphics

Go to:

LOCUS CP029805 19 bp DNA linear BCT **14-JUN-2018**

DEFINITION Bacillus anthracis strain London_499 chromosome, complete genome.

ACCESSION CP029805 REGION: 2472315..2472333

VERSION CP029805.1

DBLINK BioProject: PRJNA287512

BioSample: SAMN03790807

KEYWORDS .

SOURCE Bacillus anthracis

ORGANISM Bacillus anthracis

Bacteria; Firmicutes; Bacilli; Bacillales; Bacillaceae; Bacillus; Bacillus cereus group.

REFERENCE 1 (bases 1 to 19)

AUTHORS Pullan,S.T., Miles,R.W., Lewandowski,K. and Vipond,R.

TITLE Closed genome sequence using hybrid Nanopore/Illumina assembly of a Bacillus anthracis isolate from an animal-skin-drum-associated anthrax case in the UK

JOURNAL Microbiol Resour Announc (2018) In press

REFERENCE 2 (bases 1 to 19)

AUTHORS Pullan,S.T., Miles,R.W., Lewandowski,K. and Vipond,R.

TITLE Direct Submission

JOURNAL **Submitted (07-JUN-2018) Microbiology Services - Research, Public Health England, Porton Down, Salisbury SP4 0JG, United Kingdom**

```
COMMENT     Annotation was added by the NCBI Prokaryotic Genome Annotation
            Pipeline (released 2013). Information about the Pipeline can be
            found here: https://www.ncbi.nlm.nih.gov/genome/annotation_prok/
            ##Genome-Assembly-Data-START##
            Assembly Date          :: SEP-2017
            Assembly Method        :: Unicycler v. 2017-09
            Genome Representation  :: Full
            Expected Final Version :: Yes
            Genome Coverage        :: 50.0x
            Sequencing Technology  :: Illumina; Oxford nanopore
            ##Genome-Assembly-Data-END##

            ##Genome-Annotation-Data-START##
            Annotation Provider               :: NCBI
            Annotation Date                   :: 06/08/2018 12:09:11
            Annotation Pipeline               :: NCBI Prokaryotic Genome
                                                 Annotation Pipeline
            Annotation Method                 :: Best-placed reference protein
                                                 set; GeneMarkS+
            Annotation Software revision      :: 4.5
            Features Annotated                :: Gene; CDS; rRNA; tRNA; ncRNA;
                                                 repeat_region
            Genes (total)                     :: 6,105
            CDS (total)                       :: 5,971
            Genes (coding)                    :: 5,649
            CDS (coding)                      :: 5,649
            Genes (RNA)                       :: 134
            rRNAs                             :: 11, 11, 11 (5S, 16S, 23S)
            complete rRNAs                    :: 11, 11, 11 (5S, 16S, 23S)
            tRNAs                             :: 96
            ncRNAs                            :: 5
            Pseudo Genes (total)              :: 322
            Pseudo Genes (ambiguous residues) :: 0 of 322
            Pseudo Genes (frameshifted)       :: 214 of 322
            Pseudo Genes (incomplete)         :: 66 of 322
            Pseudo Genes (internal stop)      :: 95 of 322
            Pseudo Genes (multiple problems)  :: 48 of 322
            ##Genome-Annotation-Data-END##
FEATURES             Location/Qualifiers
     source          1..19
                     /organism="Bacillus anthracis"
                     /mol_type="genomic DNA"
                     /strain="London_499"
                     /isolation_source="Animal hide"
                     /db_xref="taxon:1392"
```

/country="United Kingdom: London"
/collection_date="2006-12-01"
/collected_by="Public Health England"
gene <1..>19
/locus_tag="DNQ11_13645"
CDS <1..>19
/locus_tag="DNQ11_13645"
/inference="COORDINATES: similar to AA
sequence:RefSeq:NP_845012.1"
/note="Derived by automated computational analysis using
gene prediction method: Protein Homology."
/codon_start=3
/transl_table=11
/product="penicillin-binding protein"
/protein_id="AWU53438.1"
/translation="MKTKDQIDNIVKETYRHIDFSGVILVKEKKGIVYEEVFGYANRN
ECINNTLQTRFGIASGCKIFTAIGICQLVENGVITFQTKLKECLSVNFPNFNEDITIH
HLLTHSSGIPDYFDESIMDNFEDLWKQTPMYLLKSLKDFLPLFQNRDMMYAPGSKFHY
NNAGFIILGLIIEEQTGLTFTEYIEKNIFNPIGMNHSGYFSLDRLPRQTALGYIKDEI
SQTWRTNAYSIPIKGGSDGGAFITASDMLKFWEALFNYEIISQEYTKILLTPHIQVNN
NQSYGYGIWIETRENKIF**KYHVMGY**DPGVSFRSAVYPDLGITLVIPSNKGAGAEKLMI
EIEGSF"

ORIGIN
1 **aatatcatgt aatgggata**
//

Bacillus anthracis strain London_499 chromosome, complete genome
Sequence ID: CP029805.1 **Length: 5228599 Number of Matches: 1**
Range 1: 2472315 to 2472333 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
38.2 bits(19)	2.2	19/19(100%)	0/19(0%)	Plus/Minus

```
CDS: Putative 1       5           G  M  V  H  Y
Query                 1        TATCCCATTACATGATATT  19
                               ||||||||||||||||||
Sbjct                 2472333  TATCCCATTACATGATATT  2472315
CDS:penicillin-bindi  301      Y  G  M  V  H  Y  K
```

Bacillus anthracis strain FDAARGOS_341 chromosome, complete genome

GenBank: CP022044.2

FASTA Graphics

Go to:

LOCUS CP022044 19 bp DNA linear BCT **30-SEP-2019**

DEFINITION Bacillus anthracis strain FDAARGOS_341 chromosome, complete genome.

ACCESSION CP022044 REGION: 43815..43833

VERSION CP022044.2

DBLINK BioProject: PRJNA231221
BioSample: SAMN06173354

KEYWORDS .

SOURCE Bacillus anthracis

ORGANISM Bacillus anthracis
Bacteria; Firmicutes; Bacilli; Bacillales; Bacillaceae; Bacillus;
Bacillus cereus group.

REFERENCE 1 (bases 1 to 19)

AUTHORS Duncan,R., Fisher,C., Tallon,L.J., Sadzewicz,L., Sengamalay,N., Ott,S., Godinez,A., Nagaraj,S., Vavikolanu,K., Vyas,G., Nadendla,S. and Sichtig,H.

TITLE FDA dAtabase for Regulatory Grade micrObial Sequences (FDA-ARGOS): Supporting development and validation of Infectious Disease Dx tests

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 19)

AUTHORS Duncan,R., Fisher,C., Tallon,L.J., Sadzewicz,L., Sengamalay,N., Ott,S., Godinez,A., Nagaraj,S., Vavikolanu,K., Nadendla,S., George,J. and Sichtig,H.

TITLE Direct Submission

JOURNAL **Submitted (20-JUN-2017) Center for Devices and Radiological Health, US Food and Drug Administration, 10903 New Hampshire Avenue, Silver Spring, MD 20993-0002, USA**

REFERENCE 3 (bases 1 to 19)

AUTHORS Duncan,R., Fisher,C., Tallon,L.J., Sadzewicz,L., Sengamalay,N., Ott,S., Godinez,A., Nagaraj,S., Vavikolanu,K., Vyas,G., Nadendla,S. and Sichtig,H.

TITLE Direct Submission

JOURNAL **Submitted (09-DEC-2017) Center for Devices and Radiological Health, US Food and Drug Administration, 10903 New Hampshire Avenue, Silver Spring, MD 20993-0002, USA**

REMARK Sequence update by submitter

COMMENT On Mar 6, 2018 this sequence version replaced CP022044.1.
Annotation was added by the NCBI Prokaryotic Genome Annotation Pipeline (released 2013). Information about the Pipeline can be found here: https://www.ncbi.nlm.nih.gov/genome/annotation_prok/

```
            ##Genome-Assembly-Data-START##
            Assembly Method       :: SMRT v. 2.3.0, HGAP v. 3.0
            Genome Coverage       :: 14.7389309711375x
            Sequencing Technology :: PacBio
            ##Genome-Assembly-Data-END##

            ##Genome-Annotation-Data-START##
            Annotation Provider               :: NCBI
            Annotation Date                   :: 03/05/2018 22:55:31
            Annotation Pipeline               :: NCBI Prokaryotic Genome
                                                 Annotation Pipeline
            Annotation Method                 :: Best-placed reference protein
                                                 set; GeneMarkS+
            Annotation Software revision      :: 4.4
            Features Annotated                :: Gene; CDS; rRNA; tRNA; ncRNA;
                                                 repeat_region
            Genes (total)                     :: 6,003
            CDS (total)                       :: 5,870
            Genes (coding)                    :: 5,569
            CDS (coding)                      :: 5,569
            Genes (RNA)                       :: 133
            rRNAs                             :: 11, 11, 11 (5S, 16S, 23S)
            complete rRNAs                    :: 11, 11, 11 (5S, 16S, 23S)
            tRNAs                             :: 95
            ncRNAs                            :: 5
            Pseudo Genes (total)              :: 301
            Pseudo Genes (ambiguous residues) :: 0 of 301
            Pseudo Genes (frameshifted)       :: 201 of 301
            Pseudo Genes (incomplete)         :: 58 of 301
            Pseudo Genes (internal stop)      :: 89 of 301
            Pseudo Genes (multiple problems)  :: 42 of 301
            ##Genome-Annotation-Data-END##

FEATURES             Location/Qualifiers
     source          1..19
                     /organism="Bacillus anthracis"
                     /mol_type="genomic DNA"
                     /strain="FDAARGOS_341"
                     /isolation_source="Colorado Serum Co., Anthrax Spore
                     Vaccine"
                     /culture_collection="FDA:FDAARGOS_341"
                     /db_xref="taxon:1392"
                     /collection_date="15-Apr-2015"
                     /collected_by="FDA CBER"
     gene            <1..>19
```

/locus_tag="CEQ19_00235"

CDS <1..>19

/locus_tag="CEQ19_00235"

/inference="COORDINATES: similar to AA sequence:RefSeq:NP_845012.1"

/note="Derived by automated computational analysis using gene prediction method: Protein Homology."

/codon_start=3

/transl_table=11

/product="**penicillin-binding protein**"

/protein_id="ASE27585.1"

/translation="MKTKDQIDNIVKETYRHIDFSGVILVKEKKGIVYEEVFGYANRN ECINNTLQTRFGIASGCKIFTAIGICQLVENGVITFQTKLKECLSVNFPNFNEDITIH HLLTHSSGIPDYFDESIMDNFEDLWKQTPMYLLKSLKDFLPLFQNRDMMYAPGSKFHY NNAGFIILGLIIEEQTGLTFTEYIEKNIFNPIGMNHSGYFSLDRLPRQTALGYIKDEI SQTWRTNAYSIPIKGGSDGGAFITASDMLKFWEALFNYEIISQEYTKILLTPHIQVNN NQSYGYGIWIETRENKIF**KYHVMGY**DPGVSFRSAVYPDLGITLVIPSNKGAGAEKLMI EIEGSF"

ORIGIN

1 **aatatcatgt aatgggata**

//

Bacillus anthracis strain FDAARGOS_341 chromosome, complete genome
Sequence ID: CP022044.2 **Length: 5228806 Number of Matches: 1**
Range 1: 43815 to 43833 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
38.2 bits(19)	2.2	19/19(100%)	0/19(0%)	Plus/Minus

```
CDS: Putative 1       5         G  M  V  H  Y
Query                 1      TATCCCATTACATGATATT  19
                             ||||||||||||||||||
Sbjct                 43833  TATCCCATTACATGATATT  43815
CDS:penicillin-bindi  301    Y  G  M  V  H  Y  K
```

Bacillus anthracis strain 14RA5914 chromosome, complete genome

GenBank: CP023001.1

FASTA Graphics

Go to:

LOCUS CP023001 19 bp DNA linear BCT 30-AUG-2017

DEFINITION Bacillus anthracis strain 14RA5914 chromosome, complete genome.

ACCESSION CP023001 REGION: 237598..237616

VERSION CP023001.1

DBLINK BioProject: PRJNA397960
BioSample: SAMN07498358
KEYWORDS .
SOURCE Bacillus anthracis
ORGANISM Bacillus anthracis
Bacteria; Firmicutes; Bacilli; Bacillales; Bacillaceae; Bacillus;
Bacillus cereus group.
REFERENCE 1 (bases 1 to 19)
AUTHORS Elschner,M.C., Busch,A., Schliephake,A., Gaede,W., Zuchantke,E. and
Tomaso,H.
TITLE High quality genome sequence of Bacillus anthracis strain 14RA5914
isolated during an outbreak in Germany 2014
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 19)
AUTHORS Elschner,M.C., Busch,A., Schliephake,A., Gaede,W., Zuchantke,E. and
Tomaso,H.
TITLE Direct Submission
JOURNAL Submitted (24-AUG-2017) Institute for Bacterial Infections and
Zoonoses, Friedrich Loeffler Institut, Federal Research Institute
for Animal Health, Naumburger Strasse 96a, Jena, Thuringia 07743,
Germany
COMMENT Bacteria and source DNA available from FLI.

##Genome-Assembly-Data-START##
Assembly Date :: 02-JUN-2017
Assembly Method :: HGAP3 v. 2.3.0
Assembly Name :: 14RA5914
Genome Representation :: Full
Expected Final Version :: Yes
Genome Coverage :: 100.0x
Sequencing Technology :: PacBio
##Genome-Assembly-Data-END##
FEATURES Location/Qualifiers
source 1..19
/organism="Bacillus anthracis"
/mol type="genomic DNA"
/strain="14RA5914"
/isolation_source="spleen"
/host="Bos primigenius taurus"
/db_xref="taxon:1392"
/country="Germany: Dobichau"
/collection date="16-Apr-2014"
/collected_by="Department of Veterinary Medicine, State
Institute for Consumer Protection of Saxony-Anhalt,
Stendal, Germany"

```
                     /identified_by="Friedrich-Loeffler-Institut, Institute of
                     Bacterial Infections and Zoonoses, Jena, Germany"
ORIGIN
        1 aatatcatgt aatgggata
//
```

Bacillus anthracis strain 14RA5914, complete genome
Sequence ID: CP023001.1 Length: 5245242 Number of Matches: 1
Range 1: 237598 to 237616 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
38.2 bits(19)	2.6	19/19(100%)	0/19(0%)	Plus/Minus

```
Query  1       TATCCCATTACATGATATT  19
               ||||||||||||||||||
Sbjct  237616  TATCCCATTACATGATATT  237598
```

OTHER PRODUCTS OF SIGNIFICANT ALIGNMENTS (CDS FEATURE)

Bacillus thuringiensis strain FDAARGOS_791 chromosome, complete genome
Sequence ID: CP054568.1 **Length: 5281841 Number of Matches: 1**
Range 1: 4982870 to 4982888 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
38.2 bits(19)	2.2	19/19(100%)	0/19(0%)	Plus/Minus

```
CDS: Putative 1       5           G  M  V  H  Y
Query                 1        TATCCCATTACATGATATT  19
                               ||||||||||||||||||
Sbjct                 4982888  TATCCCATTACATGATATT  4982870
CDS:serine hydrolase  301      Y  G  M  V  H  Y  K
```

Bacillus thuringiensis strain HER1410 chromosome, complete genome
Sequence ID: CP050183.1 **Length: 5585577 Number of Matches: 2**
Range 1: 2588894 to 2588912 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
38.2 bits(19)	2.2	19/19(100%)	0/19(0%)	Plus/Minus

```
CDS: Putative 1       5           G  M  V  H  Y
Query                 1        TATCCCATTACATGATATT  19
                               ||||||||||||||||||
Sbjct                 2588912  TATCCCATTACATGATATT  2588894
CDS:serine hydrolase  302      Y  G  M  V  H  Y  K
```

Bacillus thuringiensis strain FDAARGOS_792 chromosome, complete genome
Sequence ID: CP053938.1 **Length: 5251676 Number of Matches: 2**
Range 1: 1731215 to 1731233 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
38.2 bits(19)	2.2	19/19(100%)	0/19(0%)	Plus/Minus

```
CDS: Putative 1       5           G  M  V  H  Y
Query                 1        TATCCCATTACATGATATT  19
                               ||||||||||||||||||
Sbjct                 1731233  TATCCCATTACATGATATT  1731215
CDS:serine hydrolase  301      Y  G  M  V  H  Y  K
```

Range 2: 587161 to 587176 GenBank Graphics

Score Expect Identities Gaps Strand

32.2 bits(16) 136 16/16(100%) 0/16(0%) Plus/Plus

```
CDS: Putative 1       1          P  L  H  D  I
Query                 4        CCCATTACATGATATT  19
                               ||||||||||||||||
Sbjct                 587161   CCCATTACATGATATT  587176
CDS:phosphoadenylyl-  192        P  L  H  D  I
```

Bacillus thuringiensis strain FDAARGOS_794 chromosome, complete genome
Sequence ID: CP053934.1 **Length: 5214223 Number of Matches: 2**
Range 1: 2468315 to 2468333 GenBank Graphics

Score Expect Identities Gaps Strand

38.2 bits(19) 2.2 19/19(100%) 0/19(0%) Plus/Plus

```
CDS: Putative 1       5            G  M  V  H  Y
Query                 1         TATCCCATTACATGATATT  19
                                ||||||||||||||||||
Sbjct                 2468315   TATCCCATTACATGATATT  2468333
CDS:serine hydrolase  301       Y  G  M  V  H  Y  K
```

Range 2: 3602089 to 3602104 GenBank Graphics

Score Expect Identities Gaps Strand

32.2 bits(16) 136 16/16(100%) 0/16(0%) Plus/Minus

```
CDS: Putative 1       1          P  L  H  D  I
Query                 4        CCCATTACATGATATT  19
                               ||||||||||||||||
Sbjct                 3602104  CCCATTACATGATATT  3602089
CDS:phosphoadenylyl-  192        P  L  H  D  I
```

Bacillus cereus strain FDAARGOS_797 chromosome, complete genome
Sequence ID: CP053931.1 **Length: 5413450 Number of Matches: 2**
Range 1: 4336123 to 4336141 GenBank Graphics

Score Expect Identities Gaps Strand

38.2 bits(19) 2.2 19/19(100%) 0/19(0%) Plus/Plus

```
CDS: Putative 1       5            G  M  V  H  Y
Query                 1         TATCCCATTACATGATATT  19
                                ||||||||||||||||||
Sbjct                 4336123   TATCCCATTACATGATATT  4336141
CDS:serine hydrolase  301       Y  G  M  V  H  Y  K
```

Range 2: 216855 to 216870 GenBank Graphics

Score Expect Identities Gaps Strand

32.2 bits(16) 136 16/16(100%) 0/16(0%) Plus/Minus

```
CDS: Putative 1       1          P  L  H  D  I
Query                 4        CCCATTACATGATATT  19
                               ||||||||||||||||
Sbjct                 216870   CCCATTACATGATATT  216855
CDS:phosphoadenylyl-  192        P  L  H  D  I
```

Bacillus cereus strain FDAARGOS_780 chromosome, complete genome
Sequence ID: CP053997.1 **Length: 5271040 Number of Matches: 2**
Range 1: 615231 to 615249 GenBank Graphics

Score Expect Identities Gaps Strand

38.2 bits(19) 2.2 19/19(100%) 0/19(0%) Plus/Plus

```
CDS: Putative 1       5          G  M  V  H  Y
Query                 1        TATCCCATTACATGATATT  19
                               |||||||||||||||||||
Sbjct                 615231   TATCCCATTACATGATATT  615249
CDS:serine hydrolase  301      Y  G  M  V  H  Y  K
```

Range 2: 1757099 to 1757114 GenBank Graphics

Score Expect Identities Gaps Strand

32.2 bits(16) 136 16/16(100%) 0/16(0%) Plus/Minus

```
CDS: Putative 1       1          P  L  H  D  I
Query                 4        CCCATTACATGATATT  19
                               ||||||||||||||||
Sbjct                 1757114  CCCATTACATGATATT  1757099
CDS:phosphoadenylyl-  192        P  L  H  D  I
```

Bacillus cereus strain FDAARGOS_781 chromosome, complete genome
Sequence ID: CP053991.1 **Length: 5271029 Number of Matches: 2**
Range 1: 2062680 to 2062698 GenBank Graphics

Score Expect Identities Gaps Strand

38.2 bits(19) 2.2 19/19(100%) 0/19(0%) Plus/Minus

```
CDS: Putative 1       5          G  M  V  H  Y
Query                 1        TATCCCATTACATGATATT  19
                               |||||||||||||||||||
Sbjct                 2062698  TATCCCATTACATGATATT  2062680
CDS:serine hydrolase  301      Y  G  M  V  H  Y  K
```

Range 2: 920826 to 920841 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
32.2 bits(16)	136	16/16(100%)	0/16(0%)	Plus/Plus

```
CDS: Putative 1       1          P  L  H  D  I
Query                 4        CCCATTACATGATATT  19
                               ||||||||||||||||
Sbjct                 920826   CCCATTACATGATATT  920841
CDS:phosphoadenylyl-  192        P  L  H  D  I
```

Bacillus thuringiensis strain FDAARGOS_795 chromosome, complete genome
Sequence ID: CP053980.1 **Length: 5228070 Number of Matches: 2**
Range 1: 4574402 to 4574420 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
38.2 bits(19)	2.2	19/19(100%)	0/19(0%)	Plus/Minus

```
CDS: Putative 1       5            G  M  V  H  Y
Query                 1         TATCCCATTACATGATATT  19
                                ||||||||||||||||||
Sbjct                 4574420   TATCCCATTACATGATATT  4574402
CDS:serine hydrolase  301       Y  G  M  V  H  Y  K
```

Range 2: 3420689 to 3420704 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
32.2 bits(16)	136	16/16(100%)	0/16(0%)	Plus/Plus

```
CDS: Putative 1       1          P  L  H  D  I
Query                 4        CCCATTACATGATATT  19
                               ||||||||||||||||
Sbjct                 3420689  CCCATTACATGATATT  3420704
CDS:phosphoadenylyl-  192        P  L  H  D  I
```

Bacillus cereus strain FDAARGOS_798 chromosome, complete genome
Sequence ID: CP053954.1 **Length: 5381410 Number of Matches: 1**
Range 1: 5056611 to 5056629 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
38.2 bits(19)	2.2	19/19(100%)	0/19(0%)	Plus/Minus

```
CDS: Putative 1       5            G  M  V  H  Y
Query                 1         TATCCCATTACATGATATT  19
                                ||||||||||||||||||
Sbjct                 5056629   TATCCCATTACATGATATT  5056611
CDS:serine hydrolase  302       Y  G  M  V  H  Y  K
```

Bacillus cereus strain FDAARGOS_802 chromosome, complete genome
Sequence ID: CP053965.1**Length: 5342923Number of Matches: 1**
Range 1: 3056974 to 3056992 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
38.2 bits(19)	2.2	19/19(100%)	0/19(0%)	Plus/Plus

```
CDS: Putative 1       5           G  M  V  H  Y
Query                 1        TATCCCATTACATGATATT  19
                               ||||||||||||||||||
Sbjct                 3056974  TATCCCATTACATGATATT  3056992
CDS:serine hydrolase  301      Y  G  M  V  H  Y  K
```

Bacillus tropicus strain FDAARGOS_782 chromosome, complete genome
Sequence ID: CP053955.1 **Length: 5263142 Number of Matches: 1**
Range 1: 3671059 to 3671077 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
38.2 bits(19)	2.2	19/19(100%)	0/19(0%)	Plus/Plus

```
CDS: Putative 1       5           G  M  V  H  Y
Query                 1        TATCCCATTACATGATATT  19
                               ||||||||||||||||||
Sbjct                 3671059  TATCCCATTACATGATATT  3671077
CDS:serine hydrolase  301      Y  G  M  V  H  Y  K
```

Bacillus cereus strain FDAARGOS_799 chromosome, complete genome
Sequence ID: CP053951.1 **Length: 5381556 Number of Matches: 1**
Range 1: 1001141 to 1001159 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
38.2 bits(19)	2.2	19/19(100%)	0/19(0%)	Plus/Plus

```
CDS: Putative 1       5           G  M  V  H  Y
Query                 1        TATCCCATTACATGATATT  19
                               ||||||||||||||||||
Sbjct                 1001141  TATCCCATTACATGATATT  1001159
CDS:serine hydrolase  302      Y  G  M  V  H  Y  K
```

Bacillus thuringiensis strain FDAARGOS_793 chromosome, complete genome
Sequence ID: CP053981.1 **Length: 5256259 Number of Matches: 1**
Range 1: 3437404 to 3437422 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
38.2 bits(19)	2.2	19/19(100%)	0/19(0%)	Plus/Minus

```
CDS: Putative 1        5           G  M  V  H  Y
Query                  1        TATCCCATTACATGATATT  19
                                 ||||||||||||||||||
Sbjct                  3437422  TATCCCATTACATGATATT  3437404
CDS:serine hydrolase   301      Y  G  M  V  H  Y  K
```

Bacillus cereus strain WPySW2 chromosome, complete genome
Sequence ID: CP053289.1**Length: 5328644 Number of Matches: 2**
Range 1: 2826804 to 2826822 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
38.2 bits(19)	2.2	19/19(100%)	0/19(0%)	Plus/Plus

```
CDS: Putative 1        5           G  M  V  H  Y
Query                  1        TATCCCATTACATGATATT  19
                                 ||||||||||||||||||
Sbjct                  2826804  TATCCCATTACATGATATT  2826822
CDS:serine hydrolase   302      Y  G  M  V  H  Y  K
```

Range 2: 540310 to 540326 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
34.2 bits(17)	34	17/17(100%)	0/17(0%)	Plus/Plus

```
Query  6       CATTACATGATATTATT  22
               |||||||||||||||||
Sbjct  540310  CATTACATGATATTATT  540326
```

Bacillus sp. RZ2MS9 chromosome, complete genome
Sequence ID: CP049978.1 **Length: 5357194 Number of Matches: 1**
Range 1: 2581327 to 2581345 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
38.2 bits(19)	2.2	19/19(100%)	0/19(0%)	Plus/Minus

```
CDS: Putative 1    5           D  W  *  M  I  N
Query              1        TATCCCATTACATGATATT  19
                            ||||||||||||||||||
Sbjct              2581345  TATCCCATTACATGATATT  2581327
CDS: hypothetica   90         D  W  *  M  I  N
```

Bacillus tropicus strain AOA-CPS1 chromosome
Sequence ID: CP049019.1 **Length: 5246960 Number of Matches: 2**
Range 1: 1935094 to 1935112 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
38.2 bits(19)	2.2	19/19(100%)	0/19(0%)	Plus/Minus

```
CDS: Putative 1       5           G  M  V  H  Y
Query                 1        TATCCCATTACATGATATT  19
                               ||||||||||||||||||
Sbjct                 1935112  TATCCCATTACATGATATT  1935094
CDS:serine hydrolase  301      Y  G  M  V  H  Y  K
```

Range 2: 791722 to 791737 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
32.2 bits(16)	136	16/16(100%)	0/16(0%)	Plus/Plus

```
CDS: Putative 1       1         P  L  H  D  I
Query                 4       CCCATTACATGATATT  19
                              ||||||||||||||||
Sbjct                 791722  CCCATTACATGATATT  791737
CDS:phosphoadenylyl-  192       P  L  H  D  I
```

Bacillus wiedmannii PL1 DNA, complete genome
Sequence ID: AP022643.1 **Length: 5309441 Number of Matches: 1**
Range 1: 68515 to 68533 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
38.2 bits(19)	2.2	19/19(100%)	0/19(0%)	Plus/Minus

```
CDS: Putative 1       5         G  M  V  H  Y
Query                 1      TATCCCATTACATGATATT  19
                             ||||||||||||||||||
Sbjct                 68533  TATCCCATTACATGATATT  68515
CDS:penicillin-bindi  302    Y  G  M  V  H  Y  K
```

Bacillus cereus strain SGAir0260 chromosome
Sequence ID: CP028009.1 **Length: 5945656 Number of Matches: 1**
Range 1: 5140082 to 5140100 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
38.2 bits(19)	2.2	19/19(100%)	0/19(0%)	Plus/Plus

```
CDS: Putative 1   5          D  W  *  M  I  N
Query             1        TATCCCATTACATGATATT  19
                           ||||||||||||||||||
Sbjct             5140082  TATCCCATTACATGATATT  5140100
CDS: hypothetica  90         D  W  *  M  I  N
```

Bacillus cereus strain SGAir0263 chromosome, complete genome
Sequence ID: CP027920.1 **Length: 5947158 Number of Matches: 1**
Range 1: 2880936 to 2880954 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
38.2 bits(19)	2.2	19/19(100%)	0/19(0%)	Plus/Minus

```
CDS: Putative 1    5          D  W  *  M  I  N
Query              1          TATCCCATTACATGATATT  19
                              ||||||||||||||||||
Sbjct              2880954    TATCCCATTACATGATATT  2880936
CDS: hypothetica   90         D  W  *  M  I  N
```

Bacillus sp. A260 chromosome, complete genome
Sequence ID: CP046398.1 **Length: 5286434 Number of Matches: 2**
Range 1: 2475253 to 2475271 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
38.2 bits(19)	2.2	19/19(100%)	0/19(0%)	Plus/Minus

```
CDS: Putative 1        5          G  M  V  H  Y
Query                  1          TATCCCATTACATGATATT  19
                                  ||||||||||||||||||
Sbjct                  2475271    TATCCCATTACATGATATT  2475253
CDS:serine hydrolase   302        Y  G  M  V  H  Y  K
```

Bacillus cereus strain SB1 chromosome, complete genome
Sequence ID: CP045606.2 **Length: 5461537 Number of Matches: 2**
Range 1: 2670087 to 2670105 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
38.2 bits(19)	2.2	19/19(100%)	0/19(0%)	Plus/Minus

```
CDS: Putative 1        5          G  M  V  H  Y
Query                  1          TATCCCATTACATGATATT  19
                                  ||||||||||||||||||
Sbjct                  2670105    TATCCCATTACATGATATT  2670087
CDS:serine hydrolase   302        Y  G  M  V  H  Y  K
```

Bacillus thuringiensis strain GA-A07 chromosome, complete genome
Sequence ID: CP042270.1 **Length: 5272357 Number of Matches: 2**
Range 1: 3697477 to 3697495 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
38.2 bits(19)	2.2	19/19(100%)	0/19(0%)	Plus/Minus

```
CDS: Putative 1        5          G  M  V  H  Y
Query                  1          TATCCCATTACATGATATT  19
                                  ||||||||||||||||||
Sbjct                  3697495    TATCCCATTACATGATATT  3697477
CDS:serine hydrolase   302        Y  G  M  V  H  Y  K
```

Bacillus thuringiensis strain BT62 chromosome, complete genome
Sequence ID: CP044978.1 **Length: 5294964 Number of Matches: 2**
Range 1: 3359829 to 3359847 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
38.2 bits(19)	2.2	19/19(100%)	0/19(0%)	Plus/Minus

```
CDS: Putative 1       5           G  M  V  H  Y
Query                 1        TATCCCATTACATGATATT  19
                               ||||||||||||||||||
Sbjct                 3359847  TATCCCATTACATGATATT  3359829
CDS:serine hydrolase  302      Y  G  M  V  H  Y  K
```

PREDICTED: Sitophilus oryzae uncharacterized LOC115891418 (LOC115891418), mRNA
Sequence ID: XM_030911861.1 **Length: 1572 Number of Matches: 1**
Related Information
Gene-associated gene details
Genome Data Viewer-aligned genomic context
Range 1: 1404 to 1422 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
38.2 bits(19)	2.2	19/19(100%)	0/19(0%)	Plus/Minus

```
CDS: Putative 1       6      G  N  C  S  I  I
Query                 4      CCCATTACATGATATTATT  22
                             ||||||||||||||||||
Sbjct                 1422   CCCATTACATGATATTATT  1404
CDS:uncharacterized   474    G  N  C  S  I  I
```

Bacillus sp. JAS24-2 chromosome, complete genome
Sequence ID: CP031068.1 **Length: 5225480 Number of Matches: 1**
Range 1: 2497014 to 2497032 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
38.2 bits(19)	2.2	19/19(100%)	0/19(0%)	Plus/Minus

```
CDS: Putative 1       5           G  M  V  H  Y
Query                 1        TATCCCATTACATGATATT  19
                               ||||||||||||||||||
Sbjct                 2497032  TATCCCATTACATGATATT  2497014
CDS:class C beta-lac  302      Y  G  M  V  H  Y  K
```

Bacillus sp. SH8-8 chromosome, complete genome
Sequence ID: CP031065.1 **Length: 5241886 Number of Matches: 2**
Range 1: 2530402 to 2530420 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
38.2 bits(19)	2.2	19/19(100%)	0/19(0%)	Plus/Minus

```
CDS: Putative 1       5           G  M  V  H  Y
Query                 1        TATCCCATTACATGATATT  19
                               ||||||||||||||||||
Sbjct                 2530420  TATCCCATTACATGATATT  2530402
CDS:class C beta-lac  302      Y  G  M  V  H  Y  K
```

Bacillus sp. AR4-2 chromosome, complete genome
Sequence ID: CP031062.1 **Length: 5241257 Number of Matches: 2**
Range 1: 2529168 to 2529186 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
38.2 bits(19)	2.2	19/19(100%)	0/19(0%)	Plus/Minus

```
CDS: Putative 1       5           G  M  V  H  Y
Query                 1        TATCCCATTACATGATATT  19
                               ||||||||||||||||||
Sbjct                 2529186  TATCCCATTACATGATATT  2529168
CDS:class C beta-lac  302      Y  G  M  V  H  Y  K
```

Bacillus cereus strain 09 chromosome, complete genome
Sequence ID: CP042874.1 **Length: 5250357 Number of Matches: 2**
Range 1: 2565769 to 2565787 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
38.2 bits(19)	2.2	19/19(100%)	0/19(0%)	Plus/Minus

```
CDS: Putative 1       5           G  M  V  H  Y
Query                 1        TATCCCATTACATGATATT  19
                               ||||||||||||||||||
Sbjct                 2565787  TATCCCATTACATGATATT  2565769
CDS:serine hydrolase  302      Y  G  M  V  H  Y  K
```

Bacillus cereus strain Co1-1 chromosome
Sequence ID: CP031778.1 **Length: 6375991 Number of Matches: 2**
Range 1: 2451662 to 2451680 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
38.2 bits(19)	2.2	19/19(100%)	0/19(0%)	Plus/Minus

```
CDS: Putative 1       5           G  M  V  H  Y
Query                 1        TATCCCATTACATGATATT  19
                               ||||||||||||||||||
Sbjct                 2451680  TATCCCATTACATGATATT  2451662
CDS:class C beta-lac  302      Y  G  M  V  H  Y  K
```

Range 2: 231355 to 231370 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
32.2 bits(16)	136	16/16(100%)	0/16(0%)	Plus/Minus

```
CDS: Putative 1       1          L  H  D  I  I
Query                 7       ATTACATGATATTATT  22
                              ||||||||||||||||
Sbjct                 231370  ATTACATGATATTATT  231355
CDS:undecaprenyl-dip  100        L  H  D  I  I
```

Undecaprenyl-diphosphate phosphatase (PRK00281)

PRK00281 is a member of the superfamily cl00858 (***BacA Superfamily***). **Bacitracin resistance protein** (**BacA**) is a ***putative undecaprenol kinase***. BacA confers resistance to bacitracin, probably by phosphorylation of undecaprenol. More recent studies show that BacA has *undecaprenyl pyrophosphate phosphatase activity*. ***Undecaprenyl diphosphate*** is a polyprenol diphosphate compound having eleven prenyl units with undefined stereochemistry about the double bonds. It has a role as a Saccharomyces cerevisiae metabolite. It is a *polyprenyl diphosphate* and an *undecaprenyl phosphate* (https://pubchem.ncbi.nlm.nih.gov/compound/Undecaprenyl-pyrophosphate; https://www.ebi.ac.uk/chebi/searchId.do?chebiId=CHEBI:53042). **Undecaprenyl phosphate** is a key lipid intermediate involved in the synthesis of various bacterial cell wall polymers. Bacitracin, a mixture of related cyclic polypeptide antibiotics, is used to treat surface tissue infections. Its primary mode of action is the inhibition of bacterial cell wall synthesis through sequestration of the essential carrier lipid undecaprenyl pyrophosphate, C55-PP, resulting in the loss of cell integrity and lysis. The characteristic phosphatase sequence-motif in this family is likely to be the PGxSRSGG, compared with the PSGH of the PAP family of phosphatases (https://www.ncbi.nlm.nih.gov/Structure/cdd/cl00858).

Bacillus tropicus strain LM1212-W3 chromosome, complete genome
Sequence ID: CP041071.1 **Length: 5631167 Number of Matches: 1**
Range 1: 1106485 to 1106503 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
38.2 bits(19)	2.2	19/19(100%)	0/19(0%)	Plus/Plus

```
CDS: Putative 1       5           G  M  V  H  Y
Query                 1        TATCCCATTACATGATATT  19
                               |||||||||||||||||||
Sbjct                 1106485  TATCCCATTACATGATATT  1106503
CDS:serine hydrolase  301      Y  G  M  V  H  Y  K
```

Bacillus cereus strain FORC087 chromosome, complete genome
Sequence ID: CP029454.1 **Length: 5271204 Number of Matches: 1**
Range 1: 2546874 to 2546892 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
38.2 bits(19)	2.2	19/19(100%)	0/19(0%)	Plus/Minus

```
CDS: Putative 1       5           G  M  V  H  Y
Query                 1        TATCCCATTACATGATATT  19
                               ||||||||||||||||||
Sbjct                 2546892  TATCCCATTACATGATATT  2546874
CDS:Penicillin-bindi  302      Y  G  M  V  H  Y  K
```

Bacillus cereus ATCC 14579 chromosome, complete genome
Sequence ID: CP034551.1 **Length: 5416249 Number of Matches: 2**
Range 1: 2649594 to 2649612 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
38.2 bits(19)	2.2	19/19(100%)	0/19(0%)	Plus/Minus

```
CDS: Putative 1       5           G  M  V  H  Y
Query                 1        TATCCCATTACATGATATT  19
                               ||||||||||||||||||
Sbjct                 2649612  TATCCCATTACATGATATT  2649594
CDS:class C beta-lac  302      Y  G  M  V  H  Y  K
```

Bacillus cereus strain FORC_086 chromosome, complete genome
Sequence ID: CP029468.1 **Length: 5340498 Number of Matches: 2**
Range 1: 2587916 to 2587934 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
38.2 bits(19)	2.2	19/19(100%)	0/19(0%)	Plus/Minus

```
CDS: Putative 1       5           G  M  V  H  Y
Query                 1        TATCCCATTACATGATATT  19
                               ||||||||||||||||||
Sbjct                 2587934  TATCCCATTACATGATATT  2587916
CDS:penicillin-bindi  301      Y  G  M  V  H  Y  K
```

Bacillus sp. DU-106 chromosome, complete genome
Sequence ID: CP026607.1 **Length: 5415320 Number of Matches: 2**
Range 1: 2524508 to 2524526 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
38.2 bits(19)	2.2	19/19(100%)	0/19(0%)	Plus/Minus

```
CDS: Putative 1       5           G  M  V  H  Y
Query                 1        TATCCCATTACATGATATT  19
                               ||||||||||||||||||
Sbjct                 2524526  TATCCCATTACATGATATT  2524508
CDS:penicillin-bindi  302      Y  G  M  V  H  Y  K
```

Bacillus cereus strain 1000305 chromosome, complete genome
Sequence ID: CP028516.1 **Length: 5330209 Number of Matches: 2**
Range 1: 2600008 to 2600026 GenBank Graphics

Score Expect Identities Gaps Strand

38.2 bits(19) 2.2 19/19(100%) 0/19(0%) Plus/Minus

Features:Penicillin-binding protein

```
CDS: Putative 1        5           G  M  V  H  Y
Query                  1        TATCCCATTACATGATATT  19
                                 ||||||||||||||||||
Sbjct                  2600026  TATCCCATTACATGATATT  2600008
CDS:Penicillin-bindi   302      Y  G  M  V  H  Y  K
```

Range 2: 4838080 to 4838096 GenBank Graphics

Score Expect Identities Gaps Strand

34.2 bits(17) 34 17/17(100%) 0/17(0%) Plus/Minus

Features: hypothetical protein

```
CDS: Putative 1                          *
Query                  6        CATTACATGATATTATT  22
                                 |||||||||||||||||
Sbjct                  4838096  CATTACATGATATTATT  4838080
CDS:hypothetical pro   42                        K
```

Bacillus wiedmannii bv. thuringiensis strain FCC41 chromosome, complete genome
Sequence ID: CP024684.1 **Length: 5513449 Number of Matches: 1**
Range 1: 2707589 to 2707607 GenBank Graphics

Score Expect Identities Gaps Strand

38.2 bits(19) 2.2 19/19(100%) 0/19(0%) Plus/Minus

Features:penicillin-binding protein

```
CDS: Putative 1        5           G  M  V  H  Y
Query                  1        TATCCCATTACATGATATT  19
                                 ||||||||||||||||||
Sbjct                  2707607  TATCCCATTACATGATATT  2707589
CDS:penicillin-bindi   301      Y  G  M  V  H  Y  K
```

Bacillus sp. FDAARGOS_527 chromosome, complete genome
Sequence ID: CP033795.1 **Length: 5277296 Number of Matches: 2**
Range 1: 1774673 to 1774691GenBank Graphics

Score Expect Identities Gaps Strand

38.2 bits(19) 2.2 19/19(100%) 0/19(0%) Plus/Minus

```
CDS: Putative 1        5           G  M  V  H  Y
Query                  1        TATCCCATTACATGATATT  19
                                 ||||||||||||||||||
Sbjct                  1774691  TATCCCATTACATGATATT  1774673
CDS:class C beta-lac   301      Y  G  M  V  H  Y  K
```

Range 2: 693315 to 693330 GenBank Graphics

Score Expect Identities Gaps Strand

32.2 bits(16) 136 16/16(100%) 0/16(0%) Plus/Plus

```
CDS: Putative 1       1          P  L  H  D  I
Query                 4        CCCATTACATGATATT  19
                               ||||||||||||||||
Sbjct                 693315   CCCATTACATGATATT  693330
CDS:phosphoadenylyl-  192        P  L  H  D  I
```

Bacillus mobilis strain ML-A2C4 chromosome, complete genome
Sequence ID: CP031443.1 **Length: 5468285 Number of Matches: 1**
Range 1: 2733986 to 2734004 GenBank Graphics

Score Expect Identities Gaps Strand

38.2 bits(19) 2.2 19/19(100%) 0/19(0%) Plus/Minus

```
CDS: Putative 1       5             G  M  V  H  Y
Query                 1          TATCCCATTACATGATATT  19
                                 ||||||||||||||||||
Sbjct                 2734004    TATCCCATTACATGATATT  2733986
CDS:class C beta-lac  301        Y  G  M  V  H  Y  K
```

Bacillus thuringiensis LM1212 chromosome, complete genome
Sequence ID: CP024771.1 **Length: 5705934 Number of Matches: 1**
Range 1: 4372606 to 4372624 GenBank Graphics

Score Expect Identities Gaps Strand

38.2 bits(19) 2.2 19/19(100%) 0/19(0%) Plus/Plus

```
CDS: Putative 1       5             G  M  V  H  Y
Query                 1          TATCCCATTACATGATATT  19
                                 ||||||||||||||||||
Sbjct                 4372606    TATCCCATTACATGATATT  4372624
CDS:penicillin-bindi  301        Y  G  M  V  H  Y  K
```

Bacillus thuringiensis strain C15 chromosome, complete genome
Sequence ID: CP021436.1 **Length: 5637049 Number of Matches: 2**
Range 1: 2717226 to 2717244 GenBank Graphics

Score Expect Identities Gaps Strand

38.2 bits(19) 2.2 19/19(100%) 0/19(0%) Plus/Minus

```
CDS: Putative 1       5             G  M  V  H  Y
Query                 1          TATCCCATTACATGATATT  19
                                 ||||||||||||||||||
Sbjct                 2717244    TATCCCATTACATGATATT  2717226
CDS:penicillin-bindi  301        Y  G  M  V  H  Y  K
```

Bacillus albus strain PFYN01 chromosome, complete genome
Sequence ID: CP034548.1 **Length: 4939577 Number of Matches: 1**
Range 1: 2730671 to 2730689 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
38.2 bits(19)	2.2	19/19(100%)	0/19(0%)	Plus/Plus

```
CDS: Putative 1       5           G  M  V  H  Y
Query                 1        TATCCCATTACATGATATT  19
                               ||||||||||||||||||
Sbjct                 2730671  TATCCCATTACATGATATT  2730689
CDS:class C beta-lac  301      Y  G  M  V  H  Y  K
```

Bacillus cereus strain FORC60 chromosome, complete genome
Sequence ID: CP020383.1 **Length: 5361178 Number of Matches: 2**
Range 1: 2518418 to 2518436 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
38.2 bits(19)	2.2	19/19(100%)	0/19(0%)	Plus/Minus

```
CDS: Putative 1       5           G  M  V  H  Y
Query                 1        TATCCCATTACATGATATT  19
                               ||||||||||||||||||
Sbjct                 2518436  TATCCCATTACATGATATT  2518418
CDS:Penicillin-bindi  302      Y  G  M  V  H  Y  K
```

Range 2: 4881885 to 4881901 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
34.2 bits(17)	34	17/17(100%)	0/17(0%)	Plus/Minus

```
CDS: Putative 1                          *
Query                 6        CATTACATGATATTATT  22
                               |||||||||||||||||
Sbjct                 4881901  CATTACATGATATTATT  4881885
CDS:hypothetical pro  42                        K
```

Bacillus cereus strain TG1-6 chromosome, complete genome
Sequence ID: CP026678.1 **Length: 5417196 Number of Matches: 2**
Range 1: 4967235 to 4967253 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
38.2 bits(19)	2.2	19/19(100%)	0/19(0%)	Plus/Plus

```
CDS: Putative 1       5           G  M  V  H  Y
Query                 1        TATCCCATTACATGATATT  19
                               ||||||||||||||||||
Sbjct                 4967235  TATCCCATTACATGATATT  4967253
CDS:penicillin-bindi  302      Y  G  M  V  H  Y  K
```

Bacillus sp. HBCD-sjtu chromosome, complete genome
Sequence ID: CP025122.1 **Length: 5230501 Number of Matches: 2**
Range 1: 5166681 to 5166699 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
38.2 bits(19)	2.2	19/19(100%)	0/19(0%)	Plus/Minus

```
CDS: Putative 1       5           G  M  V  H  Y
Query                 1        TATCCCATTACATGATATT  19
                               ||||||||||||||||||
Sbjct                 5166699  TATCCCATTACATGATATT  5166681
CDS:penicillin-bindi  301      Y  G  M  V  H  Y  K
```

Range 2: 4062674 to 4062689 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
32.2 bits(16)	136	16/16(100%)	0/16(0%)	Plus/Plus

```
CDS: Putative 1       1          P  L  H  D  I
Query                 4        CCCATTACATGATATT  19
                               ||||||||||||||||
Sbjct                 4062674  CCCATTACATGATATT  4062689
CDS:phosphoadenylyl-  192        P  L  H  D  I
```

Bacillus cereus strain MLY1 chromosome MLY1.0, complete sequence
Sequence ID: CP024655.1 **Length: 5561815 Number of Matches: 2**
Range 1: 2770639 to 2770657 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
38.2 bits(19)	2.2	19/19(100%)	0/19(0%)	Plus/Plus

```
CDS: Putative 1       5           G  M  V  H  Y
Query                 1        TATCCCATTACATGATATT  19
                               ||||||||||||||||||
Sbjct                 2770639  TATCCCATTACATGATATT  2770657
CDS:penicillin-bindi  302      Y  G  M  V  H  Y  K
```

Bacillus cereus strain BHU1 chromosome
Sequence ID: CP023727.1 **Length: 5204644 Number of Matches: 2**
Range 1: 2405953 to 2405971 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
38.2 bits(19)	2.2	19/19(100%)	0/19(0%)	Plus/Minus

```
CDS: Putative 1       5           G  M  V  H  Y
Query                 1        TATCCCATTACATGATATT  19
                               ||||||||||||||||||
Sbjct                 2405971  TATCCCATTACATGATATT  2405953
CDS:penicillin-bindi  301      Y  G  M  V  H  Y  K
```

Bacillus cereus strain EFR-1 chromosome
Sequence ID: CP064072.1 **Length: 5315863 Number of Matches: 2**
Range 1: 2498140 to 2498158 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
38.2 bits(19)	2.2	19/19(100%)	0/19(0%)	Plus/Minus

```
CDS: Putative 1       5           G  M  V  H  Y
Query                 1        TATCCCATTACATGATATT  19
                               ||||||||||||||||||
Sbjct                 2498158  TATCCCATTACATGATATT  2498140
CDS:serine hydrolase  301      Y  G  M  V  H  Y  K
```

Range 2: 1426675 to 1426690 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
32.2 bits(16)	136	16/16(100%)	0/16(0%)	Plus/Plus

```
CDS: Putative 1       1          P  L  H  D  I
Query                 4        CCCATTACATGATATT  19
                               ||||||||||||||||
Sbjct                 1426675  CCCATTACATGATATT  1426690
CDS:phosphoadenylyl-  192        P  L  H  D  I
```

Bacillus cereus strain BHU2 chromosome
Sequence ID: CP023726.1 Length: 5186653 Number of Matches: 2
Range 1: 4159052 to 4159070 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
38.2 bits(19)	2.6	19/19(100%)	0/19(0%)	Plus/Minus

```
CDS: Putative 1       5           G  M  V  H  Y
Query                 1        TATCCCATTACATGATATT  19
                               ||||||||||||||||||
Sbjct                 4159070  TATCCCATTACATGATATT  4159052
CDS:penicillin-bindi  301      Y  G  M  V  H  Y  K
```

Range 2: 3458973 to 3458987 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
30.2 bits(15)	630	15/15(100%)	0/15(0%)	Plus/Minus

```
CDS: Putative 1       4           M  V  H  Y
Query                 5        CCATTACATGATATT  19
                               |||||||||||||||
Sbjct                 3458987  CCATTACATGATATT  3458973
CDS:TIGR02687 family  382        M  V  H  Y  E
```

TIGR02687 family protein

Members of this family are uncharacterized proteins sporadically distributed in bacteria and archaea, about 880 amino acids in length. This protein is repeatedly found upstream of another uncharacterized protein of about 470 amino acids in length, modeled by TIGR02688. **TIGR02687** is a member of the superfamily cl28728 (https://www.ncbi.nlm.nih.gov/Structure/cdd/TIGR02687).

Bacillus cereus strain CC-1 chromosome, complete genome
Sequence ID: CP023179.1 Length: 5279524 Number of Matches: 2
Range 1: 2700310 to 2700328 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
38.2 bits(19)	2.6	19/19(100%)	0/19(0%)	Plus/Plus

```
CDS: Putative 1       5           G  M  V  H  Y
Query                 1        TATCCCATTACATGATATT  19
                               ||||||||||||||||||
Sbjct                 2700310  TATCCCATTACATGATATT  2700328
CDS:penicillin-bindi  301      Y  G  M  V  H  Y  K
```

Range 2: 1469206 to 1469221 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
32.2 bits(16)	160	16/16(100%)	0/16(0%)	Plus/Plus

```
CDS: Putative 1       1          P  L  H  D  I
Query                 4        CCCATTACATGATATT  19
                               ||||||||||||||||
Sbjct                 1469206  CCCATTACATGATATT  1469221
CDS:phosphoadenylyl-  192        P  L  H  D  I
```

Bacillus cereus strain HBL-AI chromosome, complete genome
Sequence ID: CP023245.1 Length: 5482892 Number of Matches: 2
Range 1: 2709873 to 2709891 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
38.2 bits(19)	2.6	19/19(100%)	0/19(0%)	Plus/Minus

```
CDS: Putative 1       5           G  M  V  H  Y
Query                 1        TATCCCATTACATGATATT  19
                               ||||||||||||||||||
Sbjct                 2709891  TATCCCATTACATGATATT  2709873
CDS:penicillin-bindi  301      Y  G  M  V  H  Y  K
```

Range 2: 4999606 to 4999622 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
34.2 bits(17)	40	17/17(100%)	0/17(0%)	Plus/Minus

```
Query  6         CATTACATGATATTATT  22
                 |||||||||||||||||
Sbjct  4999622   CATTACATGATATTATT  4999606
```

Bacillus anthracis strain 14RA5914, complete genome
Sequence ID: CP023001.1 Length: 5245242 Number of Matches: 1
Range 1: 237598 to 237616 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
38.2 bits(19)	2.6	19/19(100%)	0/19(0%)	Plus/Minus

```
Query  1        TATCCCATTACATGATATT  19
                |||||||||||||||||||
Sbjct  237616   TATCCCATTACATGATATT  237598
```

Bacillus thuringiensis strain ATCC 10792, complete genome
Sequence ID: CP020754.1 Length: 5528419 Number of Matches: 2
Range 1: 990178 to 990196 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
38.2 bits(19)	2.6	19/19(100%)	0/19(0%)	Plus/Plus

```
CDS: Putative 1   5          D  W  *  M  I  N
Query             1        TATCCCATTACATGATATT  19
                           |||||||||||||||||||
Sbjct             990178   TATCCCATTACATGATATT  990196
CDS: penicillin-  277        D  W  *  M  I  N
```

Penicillin-binding protein

See when already discussed on page 286 on ***Serine hydrolase***.

Range 2: 4170375 to 4170391 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
34.2 bits(17)	40	17/17(100%)	0/17(0%)	Plus/Plus

```
Query  6         CATTACATGATATTATT  22
                 |||||||||||||||||
Sbjct  4170375   CATTACATGATATTATT  4170391
```

Bacillus cereus C1L chromosome, complete genome
Sequence ID: CP022445.1 Length: 5312355 Number of Matches: 2
Range 1: 2600914 to 2600932 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
38.2 bits(19)	2.6	19/19(100%)	0/19(0%)	Plus/Minus

```
Query  1        TATCCCATTACATGATATT  19
               ||||||||||||||||||
Sbjct  2600932  TATCCCATTACATGATATT  2600914
```

Range 2: 4846921 to 4846937GenBankGraphics

Score	Expect	Identities	Gaps	Strand
34.2 bits(17)	40	17/17(100%)	0/17(0%)	Plus/Minus

```
Query  6        CATTACATGATATTATT  22
               |||||||||||||||||
Sbjct  4846937  CATTACATGATATTATT  4846921
```

Bacillus thuringiensis strain c25, complete genome
Sequence ID: CP022345.1 Length: 5334660 Number of Matches: 2
Range 1: 2273085 to 2273103 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
38.2 bits(19)	2.6	19/19(100%)	0/19(0%)	Plus/Plus

```
Query  1        TATCCCATTACATGATATT  19
               ||||||||||||||||||
Sbjct  2273085  TATCCCATTACATGATATT  2273103
```

Range 2: 4598157 to 4598172 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
32.2 bits(16)	160	16/16(100%)	0/16(0%)	Plus/Plus

```
Query  7        ATTACATGATATTATT  22
               ||||||||||||||||
Sbjct  4598157  ATTACATGATATTATT  4598172
```

10.2 THE PRINCIPAL "PRODUCTS" IDENTIFIED THROUGH BLAST RESEARCHES CARRIED OUT ON SEQUENCE N ° 3/1 (Hitlist size: 100) AND THE FUNDAMENTAL ROLE PLAYED BY THE SERINE HYDROLASES

The "**products**" of the **Bacilli Paranthracis and Anthracis** sequences identified through the BAST research are the **Serine Hydrolases** [in particular, the *region* "**AmpC**" (*CubicO group peptidase, beta-lactamase class C family;* ***COG1680***, member of the **Transpeptidase Superfamily cl21491**, *Penicillin binding protein transpeptidase domain*)] and the **Phosphoadenylyl-sulfate reductase (thioredoxin)**.

First of all it is necessary to deepen the role played by the **Serine Hydrolases**, more precisely by the **Transmembrane Serine Protease 2** (**TMPRSS2**) and **4** (**TMPRSS4**), in **COVID-19**. The expression of **TMPRSS2** and **TMPRSS4** (two **mucosa-specific serine proteases**) facilitates the fusogenic activity of SARS-Cov-2 Spike and promotes the entry of the virus into host cells (https://www.uniprot.org/citations/32404436).
Serine proteases are known to be involved in the infection process by priming the **coronavirus spike protein**.

In an excellent article by several Italian authors (https://sibbm.zanichelli.it/italiano/2020/08/23/infezione-biologia-strutturale/#proteasi-principale) it is explained that: "Coronavirus entry into susceptible cells is a complex process that requires *concerted action of receptor binding and* ***protein S*** *proteolytic processing to promote* ***fusion between virus and cell***. One of the **cellular proteases** involved in the entry of Cov-2 into host cells is **TMPRSS2**, which is also required for *SARS-Cov infection*. In fact, **inhibition of TMPRS22 blocks the entry of SARS-Cov and Cov-2 into cells**".

It is known that "….**Angiotensin converting enzyme 2 (ACE2) receptor** can bind to **spike protein of SARS-CoV-2**. The *spike protein of SARS-CoV-2* can be cleaved and activated by **transmembrane protease**, **serine 2** (**TMPRSS2**) *of the host cells for SARS-CoV-2 infection*. Therefore, **ACE2** and **TMPRSS2** are potential antiviral targets for treatment of prevention of *SARS-CoV-2* infection (https://www.sciencedirect.com/science/article/pii/S075333222031009X?fbclid=IwAR0Y7RMgCMj_d2RdrGasy5g05QranXJi9dxKZtCnCG-zYJWVrVg0tEZgf-0). In the study just cited (see the link in parentheses), the authors found that "…..**GB-2** can inhibit **ACE2** *mRNA expression and* ***ACE2*** and ***TMPRSS2 protein expression in HepG2*** and **293 T cells** *without cytotoxicity*. **GB-2 treatment** could decrease **ACE2** and **TMPRSS2** *expression level of lung tissue and kidney tissue without adverse effects*….In addition, in the composition of **GB-2**, the authors discovered **theaflavin** could ***inhibit protein expression of ACE2 and TMPRSS2*** and ***mRNA expression of ACE2***. In conclusion,….**GB-2** and **theaflavin** could act as ***potential compounds for ACE2 and TMPRSS2 inhibitors***.

In one study employing human *lung epithelial cells infected with SARS-CoV-2* (https://www.nature.com/articles/s41598-020-78402-2) the authors identified genes co-expressed with *angiotensin I converting enzyme 2 (ACE2)*, the **human cell entry receptor of SARS-CoV-2**, and they have unveiled several genes correlated or inversely correlated with high significance. **TMPRSS4** has been identified among these genes.

TMPRSS4 is a **novel type II Transmembrane Serine Protease** highly expressed on the cell surface in pancreatic, thyroid, colon, and other cancer tissues, it is a plasma membrane-anchored **serine protease** and directly induces processing of ***pro-uPA/PLAU*** [*urokinase-type plasminogen activator* (**uPA**); (PubMed: 24434139)] into the active form through proteolytic activity (https://www.uniprot.org/uniprot/Q9NRS4). **TMPRSS4** mediates cancer cell invasion (upregulates urokinase-type plasminogen activator **uPA**), epithelial-mesenchymal transition, and metastasis. Increased TMPRSS4 expression correlates with also colorectal cancer progression. Basically, **TMPRSS4** could ***increase pro-uPA-mediated invasion in a serine proteolytic activity-dependent manner***. These observations suggest that **TMPRSS4** is an upstream regulator of pro-uPA activation (PubMed: 24434139).

TMPRSS4 seems to be capable of activating ***ENaC*** [By similarity (https://www.uniprot.org/uniprot/Q8VCA5)]. **Sodium balance** is maintained by the precise regulation of the ***activity of the epithelial sodium channel (ENaC) in the kidney***. In one study (https://pubmed.ncbi.nlm.nih.gov/12149280/) it was reported an *extracellular activation of ENaC-mediated sodium transport **(I(Na))** by a GPI-anchored **serine protease*** (mouse channel-activating protein, **mCAP1**) that was isolated from a cortical collecting duct cell line derived from mouse kidney. In this study, the authors have identified ***two additional membrane-bound*** **serine proteases** (**mCAP2** and **mCAP3**) that are expressed in the same cell line. Each of these proteases is able to **increase I(Na) 6-10-fold** in the *Xenopus oocyte* expression system. The authors show that **mCAP1** increases *I(Na) 10-fold* but the number of channels (N) remained unchanged, indicating that **mCAP1** *regulates ENaC activity by increasing its average open probability of the whole cell*. Similar results were observed for **mCAP2** and *mCAP3*. The synergism between **CAPs** and **Sgk1** (*serum- and glucocorticoid-regulated kinase* involved in the *aldosterone*-dependent signaling cascade) on **I(Na)** was always more than additive, indicating a true potentiation (https://pubmed.ncbi.nlm.nih.gov/12149280/).

Finally, **TMPRSS4**, in *gut epithelial cells, facilitates human coronavirus SARS-CoV-2 infection through*, at least, the **cleavage of coronavirus spike glycoproteins which activates the glycoprotein for host cell entry** (https://www.uniprot.org/citations/32404436). The results of this study (https://pubmed.ncbi.nlm.nih.gov/32404436/) highlight the intestine as a potential site of SARS-CoV-2 replication, which may contribute to local and systemic illness and overall disease progression.

An other "**product**" of the **Bacilli Paranthracis and Anthracis** sequences identified by the BLAST research is the **Phosphoadenylyl-sulfate reductase (thioredoxin)**. As reported in the following article (the link is shown in the next second sentence), "The recently identified **coronavirus (SARS-CoV-2) genome** *encodes for various proteins*, such as envelope, membrane, nucleocapsid, and spike protein. Among these, the *protein necessary for the virus to enter inside the host cell is* ***spike protein***". In this research (https://link.springer.com/article/10.1007/s11030-020-10171-6), "…short *synthetic peptide molecules* which bind to a *specific target protein* with a high affinity to exert its function, known as **peptide aptamers**, were designed and targeted at the **spike receptor binding domain of SARS-CoV-2**. The **peptide aptamer** has been designed by using **bacterial Thioredoxin A** as the ***scaffold protein*** and an ***18-residue-long peptide***. The tertiary structure of the *peptide aptamer* is modeled and docked to *spike receptor-binding domain of SARS CoV2*. The *aptamer* binds to *spike receptor-binding domain of SARS-CoV-2* in a **similar pattern as that of ACE2**. The **aptamer** is also predicted to be non-antigenic, non-allergenic, non-hemolytic, non-inflammatory, water-soluble with high affinity toward **ACE2** than *serum albumin*. Thus, **peptide aptamer** can be a novel approach for the therapeutic treatment for **SARS-CoV-2**.
In addition to the *products* mentioned above, the two "**products**" of the **Bacillus Cereus** sequences identified by BAST research are the **Undecaprenyl-diphosphate phosphatase (PRK00281)**, member of the ***BacA Superfamily*** [**Bacitracin resistance protein** (**BacA**), a ***putative undecaprenol kinase***] and the **TIGR02687 family protein**.

In a recent study (https://pubmed.ncbi.nlm.nih.gov/33352458/) FDA approved **drugs** and **nucleoside** analogues (in particular **NA1**, a *lead compound*) were identified as potential **SARS-CoV-2 A1pp domain inhibitor**. *Coronaviruses* possess *highly conserved viral macro domain A1pp having* ***adenosine diphosphate (ADP)-ribose*** *binding and* ***phosphatase activity sites***. **A1pp** inhibits adenosine diphosphate (ADP)-ribosylation in the host and promotes viral infection and pathogenesis. Taken together, the results of this study offers **NA1** as a lead **SARS-CoV-2 A1pp domain inhibitor for future testing and development as therapeutics against human coronavirus**.
With specific reference to the above **Undecaprenyl-diphosphate phosphatase**, some authors, studying compounds capable of interfering with **bacterial cell-wall biosynthesis**, found that the most active compounds to inhibit two enzymes involved in biosynthesis were **undecaprenyl diphosphate synthase** (**UPPS**) and **undecaprenyl diphosphate phosphatase** (**UPPP**), but not *farnesyl diphosphate synthase*, and there were good correlations between *bacterial cell growth inhibition*, **UPPS** inhibition, and **UPPP** inhibition (https://chemistry-europe.onlinelibrary.wiley.com/doi/abs/10.1002/cmdc.201600342).

Finally, the **Pubmed** research did not reveal any relationship between **SARS-CoV-2** and the **TIGR02687 family protein**.

10.3 SEQUENCE N° 3/1 (T=2; C=0; G=3; A=1) (Hitlist size 5000)

Query: **TATCCCATTACATGATATTATT**

Search Parameters	
Program	blastn
Word size	7
Expect value	1000
Hitlist size	**5000**
Match/Mismatch scores	1,-3
Gapcosts	5,2
Filter string	F
Genetic Code	1

Database	

5000 sequences selected:

Accession Seq	Description	Comomm Name	Max Score	Total Score	Query C.	E value	Per. Id.	Acc. Len
.....	.....	.....	.....	.....	.....	.....	.....	.....
CP054800.1	**Bacillus anthracis strain FDAARGOS_702 chromosome**	Bacillus anthracis	38.2	38.2	86%	2.3	100.00%	5272559
CP054797.1	**Bacillus anthracis strain FDAARGOS_703 chromosome**	Bacillus anthracis	38.2	38.2	86%	2.3	100.00%	5261520
.....	.....	.....	.....	.....	.....	.....	.....	.....
CP050973.1	**Bacillus anthracis strain FDAARGOS_699 chromosome**	Bacillus anthracis	38.2	38.2	86%	2.3	100.00%	5227345
CP050972.1	**Bacillus anthracis strain FDAARGOS_697 chromosome**	Bacillus anthracis	38.2	38.2	86%	2.3	100.00%	5240765
CP050971.1	**Bacillus anthracis strain FDAARGOS_696 chromosome**	Bacillus anthracis	38.2	38.2	86%	2.3	100.00%	5255974
CP050970.1	**Bacillus anthracis strain FDAARGOS_694 chromosome**	Bacillus anthracis	38.2	38.2	86%	2.3	100.00%	5251490
.....	.....	.....	.....	.....	.....	.....	.....	.....
CP048687.1	***Bacillus paranthracis strain MN1F chromosome, complete genome***	Bacillus paranthracis	38.2	70.4	86%	2.3	100.00%	5687871
.....	.....	.....	.....	.....	.....	.....	.....	.....
CP041981.1	***Bacillus paranthracis strain NCCP 15910 chromosome, complete genome***	Bacillus paranthracis	38.2	70.4	86%	2.3	100.00%	5208566
CP041750.1	***Bacillus paranthracis strain NCCP 14796 chromosome, complete genome***	Bacillus paranthracis	38.2	70.4	86%	2.3	100.00%	5017580
CP047131.1	**Bacillus anthracis str. BF1 chromosome, complete genome**	Bacillus anthracis str. BF1	38.2	38.2	86%	2.3	100.00%	5213532
CP047111.1	**Bacillus anthracis strain FDAARGOS_704 chromosome, complete genome**	Bacillus anthracis	38.2	38.2	86%	2.3	100.00%	5227450
CP047107.1	**Bacillus anthracis strain FDAARGOS_705 chromosome, complete genome**	Bacillus anthracis	38.2	38.2	86%	2.3	100.00%	5227452
CP047104.1	**Bacillus anthracis strain FDAARGOS_706 chromosome, complete genome**	Bacillus anthracis	38.2	38.2	86%	2.3	100.00%	5227454
CP047099.1	**Bacillus anthracis strain FDAARGOS_700 chromosome, complete genome**	Bacillus anthracis	38.2	38.2	86%	2.3	100.00%	5227419
CP047098.1	**Bacillus anthracis strain FDAARGOS_701 chromosome, complete genome**	Bacillus anthracis	38.2	38.2	86%	2.3	100.00%	5227419
CP047097.1	**Bacillus anthracis strain FDAARGOS_698 chromosome, complete genome**	Bacillus anthracis	38.2	38.2	86%	2.3	100.00%	5225746
CP047085.1	***Bacillus paranthracis strain BC307 chromosome, complete genome***	Bacillus paranthracis	38.2	70.4	86%	2.3	100.00%	5248370
.....	.....	.....	.....	.....	.....	.....	.....	.....
CP045777.1	***Bacillus paranthracis strain CFSAN068816 chromosome, complete***	Bacillus paranthracis	38.2	70.4	86%	2.3	100.00%	5351395

	genome							
.....	.....	.....	.....	.....	.....	.....	.....	.....
LR697109.1	**Chanos chanos genome assembly, chromosome: 4**	Chanos chanos	38.2	70.4	95%	2.3	100.00%	53909382
.....	.....	.....	.....	.....	.....	.....	.....	.....
AP019731.1	**Bacillus anthracis PCr DNA, complete genome**	Bacillus anthracis	38.2	38.2	86%	2.3	100.00%	5209717
.....	.....	.....	.....	.....	.....	.....	.....	.....
CP029323.1	**Bacillus anthracis strain 17OD930 chromosome, complete genome**	Bacillus anthracis	38.2	38.2	86%	2.3	100.00%	5213297
.....	.....	.....	.....	.....	.....	.....	.....	.....
CP026608.1	**Bacillus anthracis strain HDZK-BYSB7 chromosome, complete genome**	Bacillus anthracis	38.2	70.4	86%	2.3	100.00%	5294779
.....	.....	.....	.....	.....	.....	.....	.....	.....
AP018443.1	**Bacillus anthracis CZC5 DNA, complete genome**	Bacillus anthracis	38.2	38.2	86%	2.3	100.00%	5219278
CP031642.1	**Bacillus anthracis strain MCCC 1A02161 chromosome, complete genome**	Bacillus anthracis	38.2	70.4	86%	2.3	100.00%	5231857
CP031643.1	**Bacillus anthracis strain MCCC 1A01412 chromosome, complete genome**	Bacillus anthracis	38.2	70.4	86%	2.3	100.00%	5291783
.....	.....	.....	.....	.....	.....	.....	.....	.....
CP029805.1	**Bacillus anthracis strain London_499 chromosome, complete genome**	Bacillus anthracis	38.2	38.2	86%	2.3	100.00%	5228599
.....	.....	.....	.....	.....	.....	.....	.....	.....
CP022044.2	**Bacillus anthracis strain FDAARGOS_341 chromosome, complete genome**	Bacillus anthracis	38.2	38.2	86%	2.3	100.00%	5228806
.....	.....	.....	.....	.....	.....	.....	.....	.....
CP023001.1	**Bacillus anthracis strain 14RA5914, complete genome**	Bacillus anthracis	38.2	38.2	86%	2.3	100.00%	5245242
.....	.....	.....	.....	.....	.....	.....	.....	.....
CP019726.1	**Bacillus anthracis strain Sterne 34F2 genome**	Bacillus anthracis	38.2	38.2	86%	2.3	100.00%	5209112
.....	.....	.....	.....	.....	.....	.....	.....	.....
CP019588.1	**Bacillus anthracis strain SPV842_15, complete genome**	Bacillus anthracis	38.2	38.2	86%	2.3	100.00%	5228667
.....	.....	.....	.....	.....	.....	.....	.....	.....
CP018903.1	**Bacillus anthracis strain Tyrol 4675, complete genome**	Bacillus anthracis	38.2	38.2	86%	2.3	100.00%	5227565
.....	.....	.....	.....	.....	.....	.....	.....	.....
CP001974.2	**Bacillus anthracis str. A16R, complete genome**	Bacillus anthracis str. A16R	38.2	38.2	86%	2.3	100.00%	5228828
CP001970.2	**Bacillus anthracis str. A16, complete genome**	Bacillus anthracis str. A16	38.2	38.2	86%	2.3	100.00%	5228842
.....	.....	.....	.....	.....	.....	.....	.....	.....
CP012728.1	**Bacillus anthracis strain PR10-4, complete genome**	Bacillus anthracis	38.2	38.2	86%	2.3	100.00%	5227079
CP012727.1	**Bacillus anthracis strain PR09-4, complete genome**	Bacillus anthracis	38.2	38.2	86%	2.3	100.00%	5227090
CP012726.1	**Bacillus anthracis strain PR09-1, complete genome**	Bacillus anthracis	38.2	38.2	86%	2.3	100.00%	5227060
CP012725.1	**Bacillus anthracis strain PR08, complete genome**	Bacillus anthracis	38.2	38.2	86%	2.3	100.00%	5227091
CP012724.1	**Bacillus anthracis strain PR07, complete genome**	Bacillus anthracis	38.2	38.2	86%	2.3	100.00%	5227094
CP012723.1	**Bacillus anthracis strain PR06, complete genome**	Bacillus anthracis	38.2	38.2	86%	2.3	100.00%	5228656
CP012722.1	**Bacillus anthracis strain PR05, complete genome**	Bacillus anthracis	38.2	38.2	86%	2.3	100.00%	5228658
CP012721.1	**Bacillus anthracis strain PR02, complete genome**	Bacillus anthracis	38.2	38.2	86%	2.3	100.00%	5228657
CP012720.1	**Bacillus anthracis strain PR01, complete genome**	Bacillus anthracis	38.2	38.2	86%	2.3	100.00%	5228656
CP012730.1	**Bacillus anthracis strain Parent1, complete genome**	Bacillus anthracis	38.2	38.2	86%	2.3	100.00%	5228660
CP012729.1	**Bacillus anthracis strain Parent2, complete genome**	Bacillus anthracis	38.2	38.2	86%	2.3	100.00%	5227092
.....	.....	.....	.....	.....	.....	.....	.....	.....
CP015779.1	**Bacillus anthracis strain Tangail-1, complete genome**	Bacillus anthracis	38.2	38.2	86%	2.3	100.00%	5227292
.....	.....	.....	.....	.....	.....	.....	.....	.....
CP014179.1	**Bacillus anthracis strain Stendal, complete genome**	Bacillus anthracis	38.2	38.2	86%	2.3	100.00%	5226939

CP012519.1	**Bacillus anthracis strain Larissa, complete genome**	Bacillus anthracis	38.2	38.2	86%	2.3	100.00%	5226878
AP014833.1	**Bacillus anthracis DNA, nearly complete genome, strain: Shikan-NIID**	Bacillus anthracis	38.2	38.2	86%	2.3	100.00%	5228065
CP010852.1	**Bacillus anthracis strain A1144, complete genome**	Bacillus anthracis	38.2	38.2	86%	2.3	100.00%	5200257
CP010322.1	**Bacillus anthracis strain Canadian_bison, complete genome**	Bacillus anthracis	38.2	38.2	86%	2.3	100.00%	5229224
CP009981.1	**Bacillus anthracis strain Ames, complete genome**	Bacillus anthracis	38.2	38.2	86%	2.3	100.00%	5227467
CP009902.1	**Bacillus anthracis strain 2002013094, complete genome**	Bacillus anthracis	38.2	70.4	86%	2.3	100.00%	5321900
CP009598.1	**Bacillus anthracis str. V770-NP-1R, complete genome**	Bacillus anthracis str. V770-NP-1R	38.2	38.2	86%	2.3	100.00%	5228687
CP009700.1	**Bacillus anthracis strain BA1035, complete genome**	Bacillus anthracis	38.2	38.2	86%	2.3	100.00%	5210596
CP009544.1	**Bacillus anthracis strain BA1015, complete genome**	Bacillus anthracis	38.2	38.2	86%	2.3	100.00%	5214671
CP009541.1	**Bacillus anthracis str. Sterne, complete genome**	Bacillus anthracis str. Sterne	38.2	38.2	86%	2.3	100.00%	5227496
CP009697.1	**Bacillus anthracis strain RA3, complete genome**	Bacillus anthracis	38.2	38.2	86%	2.3	100.00%	5213209
CP009476.1	**Bacillus anthracis strain Pasteur, complete genome**	Bacillus anthracis	38.2	38.2	86%	2.3	100.00%	5200082
CP009464.1	**Bacillus anthracis strain SK-102, complete genome**	Bacillus anthracis	38.2	38.2	86%	2.3	100.00%	5230088
CP009341.1	**Bacillus anthracis strain Ohio ACB, complete genome**	Bacillus anthracis	38.2	38.2	86%	2.3	100.00%	5222171
CP009328.1	**Bacillus anthracis strain Vollum 1B, complete genome**	Bacillus anthracis	38.2	38.2	86%	2.3	100.00%	5229969
CP009331.1	**Bacillus anthracis strain K3, complete genome**	Bacillus anthracis	38.2	38.2	86%	2.3	100.00%	5228730
CP009325.1	**Bacillus anthracis strain PAK-1, complete genome**	Bacillus anthracis	38.2	38.2	86%	2.3	100.00%	5221627
CP009315.1	**Bacillus anthracis str. Turkey32, complete genome**	Bacillus anthracis str. Turkey32	38.2	38.2	86%	2.3	100.00%	5228731
CP010813.1	**Bacillus anthracis strain Pollino sequence**	Bacillus anthracis	38.2	38.2	86%	2.3	100.00%	5227410
CP008854.1	**Bacillus anthracis strain Han, complete genome**	Bacillus anthracis	38.2	38.2	86%	2.3	100.00%	5225433
CP008853.1	**Bacillus anthracis strain Cvac02, complete genome**	Bacillus anthracis	38.2	38.2	86%	2.3	100.00%	5227170
CP008752.1	**Bacillus anthracis strain delta Sterne genome**	Bacillus anthracis	38.2	38.2	86%	2.3	100.00%	5229650
CP007666.1	**Bacillus anthracis str. Vollum, complete genome**	Bacillus anthracis str. Vollum	38.2	38.2	86%	2.3	100.00%	5229469
CP007704.1	**Bacillus anthracis strain BFV genome**	Bacillus anthracis	38.2	38.2	86%	2.3	100.00%	5230053
CP007618.1	**Bacillus anthracis strain 2000031021, complete genome**	Bacillus anthracis	38.2	70.4	86%	2.3	100.00%	5236981
CP008846.1	**Bacillus anthracis strain HYU01, complete genome**	Bacillus anthracis	38.2	38.2	86%	2.3	100.00%	5213498
CP006742.1	**Bacillus anthracis str. SVA11, complete genome**	Bacillus anthracis str. SVA11	38.2	38.2	86%	2.3	100.00%	5210966
CP066168.1	**Bacillus anthracis strain STI1 chromosome, complete genome**	Bacillus anthracis	38.2	38.2	86%	2.3	100.00%	5229731

OD002082.1	1_Tps_b3v08	Timema poppensis	38.2	38.2	86%	2.3	100.00%	94204
CP065149.1	*Bacillus paranthracis strain Gxun-30 chromosome, complete genome*	Bacillus paranthracis	38.2	70.4	86%	2.3	100.00%	5149464
OA568019.1	1_Tdi_b3v08	Timema douglasi	38.2	38.2	86%	2.3	100.00%	76308
OA564779.1	1_Tdi_b3v08	Timema douglasi	38.2	38.2	86%	2.3	100.00%	266255
CP064079.1	*Bacillus paranthracis strain EFR-4 chromosome*	Bacillus paranthracis	38.2	70.4	86%	2.3	100.00%	5315901
CP002091.1	**Bacillus anthracis str. H9401, complete genome**	Bacillus anthracis str. H9401	38.2	38.2	86%	2.3	100.00%	5218947
CP060194.1	*Bacillus anthracis strain Kanchipuram chromosome*	Bacillus anthracis	38.2	38.2	86%	2.3	100.00%	5227419
CP001598.1	**Bacillus anthracis str. A0248, complete genome**	Bacillus anthracis str. A0248	38.2	38.2	86%	2.3	100.00%	5227419
CP001215.1	**Bacillus anthracis str. CDC 684, complete genome**	Bacillus anthracis str. CDC 684	38.2	38.2	86%	2.3	100.00%	5230115
EF040444.1	**Synthetic construct Bacillus anthracis clone FLH247716.01L BA2650 gene, complete sequence**	synthetic construct	38.2	38.2	86%	2.3	100.00%	1023
AE017334.2	**Bacillus anthracis str. 'Ames Ancestor', complete genome**	Bacillus anthracis str. 'Ames Ancestor'	38.2	38.2	86%	2.3	100.00%	5227419
AE017225.1	**Bacillus anthracis str. Sterne, complete genome**	Bacillus anthracis str. Sterne	38.2	38.2	86%	2.3	100.00%	5228663
AL589951.6	**Human DNA sequence from clone RP11-55F17 on chromosome 13, complete sequence**	Homo sapiens	38.2	38.2	86%	2.3	100.00%	55627
AE016879.1	**Bacillus anthracis str. Ames, complete genome**	Bacillus anthracis str. Ames	38.2	38.2	86%	2.3	100.00%	5227293
LR131992.1	**Gouania willdenowi genome assembly, chromosome: 16**	Gouania willdenowi	36.2	66.4	81%	9.2	100.00%	40234402
CP025664.1	**Ipomoea triloba cultivar NCNSP0323 chromosome 5**	Ipomoea triloba	36.2	36.2	100%	9.2	95.45%	32570749
OE014649.1	4_Tte_b3v08	Timema tahoe	36.2	36.2	100%	9.2	95.45%	4969
OC316763.1	3_Tce_b3v08	Timema cristinae	36.2	36.2	100%	9.2	95.45%	416485
OB795602.1	3_Tms_b3v08	Timema monikensis	36.2	36.2	100%	9.2	95.45%	93002
OD565139.1	4_Tbi_b3v08	Timema bartmani	36.2	36.2	100%	9.2	95.45%	213605
OC002533.1	2_Tsi_b3v08	Timema shepardi	36.2	36.2	81%	9.2	100.00%	111484
LR722980.1	**Thalassophryne amazonica genome assembly, chromosome: 15**	Thalassophryne amazonica	34.2	98.6	100%	40	100%	97232431
LR697108.1	**Chanos chanos genome assembly, chromosome: 3**	Chanos chanos	34.2	94.6	81%	37	100.00%	58651890
LR537125.1	**Sparus aurata genome assembly, chromosome: 5**	Sparus aurata	34.2	34.2	77%	37	100.00%	38249111
LR597561.1	**Myripristis murdjan genome assembly, chromosome: 12**	Myripristis murdjan	34.2	94.6	81%	37	100.00%	32414123
LR597554.1	**Myripristis murdjan genome assembly, chromosome: 5**	Myripristis murdjan	34.2	128	77%	37	100.00%	39210961
LR584437.1	**Salmo trutta genome assembly, chromosome:**	Salmo trutta	34.2	64.4	81%	37	100.00%	52728040

	21							
LR584429.1	Salmo trutta genome assembly, chromosome: 20	Salmo trutta	34.2	64.4	90%	37	100.00%	55161844
LR584428.1	Salmo trutta genome assembly, chromosome: 13	Salmo trutta	34.2	157	100%	37	100.00%	91488822
.....	.....	.....	.....	.....	.....	.....	.....	.....
LR131999.1	Gouania willdenowi genome assembly, chromosome: 13	Gouania willdenowi	34.2	66.4	77%	37	100.00%	44521366
LR131994.1	Gouania willdenowi genome assembly, chromosome: 4	Gouania willdenowi	34.2	124	77%	37	100.00%	43801156
LR131986.1	Gouania willdenowi genome assembly, chromosome: 17	Gouania willdenowi	34.2	94.6	77%	37	100.00%	37169619
LR131984.1	Gouania willdenowi genome assembly, chromosome: 11	Gouania willdenowi	34.2	130	81%	37	100.00%	32583565
.....	.....	.....	.....	.....	.....	.....	.....	.....
CP025648.1	Ipomoea trifida cultivar NCNSP0306 chromosome 5	Ipomoea trifida	34.2	34.2	77%	37	100.00%	26165527
.....	.....	.....	.....	.....	.....	.....	.....	.....
HG975450.1	Solanum pennellii chromosome ch11, complete genome	Solanum pennellii	34.2	96.6	90%	37	100.00%	66223686
HG975447.1	Solanum pennellii chromosome ch08, complete genome	Solanum pennellii	34.2	124	86%	37	100.00%	70546659
HG975445.1	Solanum pennellii chromosome ch06, complete genome	Solanum pennellii	34.2	64.4	95%	37	100.00%	60730942
.....	.....	.....	.....	.....	.....	.....	.....	.....
OD003851.1	1_Tps_b3v08	Timema poppensis	34.2	34.2	77%	37	100.00%	60017
OD001380.1	1_Tps_b3v08	Timema poppensis	34.2	34.2	77%	37	100.00%	120435
OD001097.1	1_Tps_b3v08	Timema poppensis	34.2	34.2	77%	37	100.00%	134835
OE002038.1	4_Tte_b3v08	Timema tahoe	34.2	34.2	77%	37	100.00%	139209
OE001353.1	4_Tte_b3v08	Timema tahoe	34.2	34.2	95%	37	95.24%	178969
OD565296.1	4_Tbi_b3v08	Timema bartmani	34.2	34.2	77%	37	100.00%	194494
OD211154.1	1_Tps_b3v08	Timema poppensis	34.2	34.2	77%	37	100.00%	519
.....	.....	.....	.....	.....	.....	.....	.....	.....
OD516663.1	1_Tps_b3v08	Timema poppensis	34.2	34.2	77%	37	100.00%	266
OC316498.1	3_Tce_b3v08	Timema cristinae	34.2	34.2	77%	37	100.00%	1009467
OB793053.1	3_Tms_b3v08	Timema monikensis	34.2	34.2	77%	37	100.00%	455776
OE000270.1	4_Tte_b3v08	Timema tahoe	34.2	34.2	77%	37	100.00%	367521
OD565288.1	4_Tbi_b3v08	Timema bartmani	34.2	34.2	95%	37	95.24%	195391
OD564554.1	4_Tbi_b3v08	Timema bartmani	34.2	34.2	77%	37	100.00%	360533
.....	.....	.....	.....	.....	.....	.....	.....	.....
OC209956.1	2_Tsi_b3v08	Timema shepardi	34.2	34.2	77%	37	100.00%	316
OC001206.1	2_Tsi_b3v08	Timema shepardi	34.2	34.2	77%	37	100.00%	182350
OC000708.1	2_Tsi_b3v08	Timema shepardi	34.2	34.2	77%	37	100.00%	238676
OA565959.1	1_Tdi_b3v08	Timema douglasi	34.2	34.2	77%	37	100.00%	141192
OA565241.1	1_Tdi_b3v08	Timema douglasi	34.2	34.2	77%	37	100.00%	193593
.....	.....	.....	.....	.....	.....	.....	.....	.....
OA611562.1	1_Tdi_b3v08	Timema douglasi	34.2	34.2	77%	37	100.00%	1171
.....	.....	.....	.....	.....	.....	.....	.....	.....
LR877239.1	Acomys russatus genome assembly, chromosome: 28	Acomys russatus	34.2	34.2	95%	37	95.24%	43538471
LR877229.1	Acomys russatus genome assembly, chromosome: 18	Acomys russatus	34.2	64.4	86%	37	100.00%	64120505
.....	.....	.....	.....	.....	.....	.....	.....	.....
AL157777.5	Human DNA sequence from clone RP11-76H14 on chromosome 6, complete sequence	Homo sapiens	34.2	34.2	77%	40	100.00%	169669
.....	.....	.....	.....	.....	.....	.....	.....	.....
XM_033117539.1	PREDICTED: Rhinolophus ferrumequinum methyltransferase like 25 (METTL25), transcript variant X2, mRNA	Rhinolophus ferrumequinum	32.2	32.2	72%	162	100.00%	1884
XM_033117538.1	PREDICTED: Rhinolophus ferrumequinum methyltransferase like 25 (METTL25), transcript variant X1, mRNA	Rhinolophus ferrumequinum	32.2	32.2	72%	162	100.00%	1994

LR778290.1	Coregonus sp. 'balchen' genome assembly, chromosome: 38	NA	32.2	92.7	90%	162	95.00%	36774138
LR778289.1	Coregonus sp. 'balchen' genome assembly, chromosome: 37	NA	32.2	64.4	72%	162	100.00%	43663377
LR778288.1	Coregonus sp. 'balchen' genome assembly, chromosome: 36	NA	32.2	32.2	90%	162	95.00%	42009912
LR778287.1	Coregonus sp. 'balchen' genome assembly, chromosome: 35	NA	32.2	64.4	90%	162	95.00%	42609905
LR778286.1	Coregonus sp. 'balchen' genome assembly, chromosome: 34	NA	32.2	64.4	90%	162	95.00%	40727438
LR778285.1	Coregonus sp. 'balchen' genome assembly, chromosome: 33	NA	32.2	126	90%	162	100.00%	44662967
LR778284.1	Coregonus sp. 'balchen' genome assembly, chromosome: 32	NA	32.2	191	90%	162	95.00%	44616205
LR778283.1	Coregonus sp. 'balchen' genome assembly, chromosome: 31	NA	32.2	94.6	90%	162	95.00%	48446552
LR778282.1	Coregonus sp. 'balchen' genome assembly, chromosome: 30	NA	32.2	225	90%	162	100.00%	48675208
LR778281.1	Coregonus sp. 'balchen' genome assembly, chromosome: 29	NA	32.2	94.6	90%	162	95.00%	48977775
LR778280.1	Coregonus sp. 'balchen' genome assembly, chromosome: 28	NA	32.2	64.4	90%	162	95.00%	46671285
LR778279.1	Coregonus sp. 'balchen' genome assembly, chromosome: 27	NA	32.2	32.2	90%	162	95.00%	48683376
LR778277.1	Coregonus sp. 'balchen' genome assembly, chromosome: 25	NA	32.2	64.4	90%	162	95.00%	51033154
LR778276.1	Coregonus sp. 'balchen' genome assembly, chromosome: 24	NA	32.2	94.6	90%	162	95.00%	50329371
LR778275.1	Coregonus sp. 'balchen' genome assembly, chromosome: 23	NA	32.2	96.6	90%	162	95.00%	52020451
LR778274.1	Coregonus sp. 'balchen' genome assembly, chromosome: 22	NA	32.2	155	90%	162	100.00%	56862223
LR778271.1	Coregonus sp. 'balchen' genome assembly, chromosome: 19	NA	32.2	94.6	90%	162	95.00%	59907985
LR778270.1	Coregonus sp. 'balchen' genome assembly, chromosome: 18	NA	32.2	64.4	90%	162	95.00%	51949489
LR778269.1	Coregonus sp. 'balchen' genome assembly, chromosome: 17	NA	32.2	62.4	90%	162	95.00%	54216998
LR778268.1	Coregonus sp. 'balchen' genome assembly, chromosome: 16	NA	32.2	193	90%	162	95.00%	54025139
LR778267.1	Coregonus sp. 'balchen' genome assembly, chromosome: 15	NA	32.2	155	95%	162	95.00%	55641933
LR778266.1	Coregonus sp. 'balchen' genome assembly, chromosome: 14	NA	32.2	64.4	90%	162	95.00%	47256133
LR778264.1	Coregonus sp. 'balchen' genome assembly, chromosome: 12	NA	32.2	310	90%	162	100.00%	63881516
LR778263.1	Coregonus sp. 'balchen' genome assembly, chromosome: 11	NA	32.2	225	90%	162	95.00%	63177489
LR778262.1	Coregonus sp. 'balchen' genome assembly, chromosome: 10	NA	32.2	159	90%	162	95.00%	60468309
LR778261.1	Coregonus sp. 'balchen' genome assembly, chromosome: 9	NA	32.2	191	90%	162	95.00%	65193448
LR778260.1	Coregonus sp. 'balchen' genome assembly, chromosome: 8	NA	32.2	159	90%	162	95.00%	68138733
LR778259.1	Coregonus sp. 'balchen' genome assembly, chromosome: 7	NA	32.2	189	90%	162	95.00%	65391737
LR778257.1	Coregonus sp. 'balchen' genome assembly, chromosome: 5	NA	32.2	157	100%	162	95.00%	45591172
LR778256.1	Coregonus sp. 'balchen' genome assembly, chromosome: 4	NA	32.2	318	100%	162	95.00%	92224161
LR778255.1	Coregonus sp. 'balchen' genome assembly, chromosome: 3	NA	32.2	64.4	90%	162	95.00%	42764345
LR778254.1	Coregonus sp. 'balchen' genome assembly, chromosome: 2	NA	32.2	62.4	72%	162	100.00%	43329510
LR778253.1	Coregonus sp. 'balchen' genome assembly, chromosome: 1	NA	32.2	124	90%	162	95.00%	93459789
LR697112.1	Chanos chanos genome assembly, chromosome: 7	milkfish	32.2	92.7	90%	162	100.00%	49187548
LR537144.1	Sparus aurata genome assembly, chromosome: 24	gilthead seabream	32.2	32.2	72%	162	100.00%	22435136
LR537143.1	Sparus aurata genome assembly, chromosome: 23	gilthead seabream	32.2	62.4	72%	162	100.00%	30526999

LR537141.1	Sparus aurata genome assembly, chromosome: 21	gilthead seabream	32.2	94.6	86%	162	100.00%	35160130
LR537139.1	Sparus aurata genome assembly, chromosome: 19	gilthead seabream	32.2	32.2	72%	162	100.00%	31047742
LR537138.1	Sparus aurata genome assembly, chromosome: 18	gilthead seabream	32.2	92.7	72%	162	100.00%	37104330
LR537127.1	Sparus aurata genome assembly, chromosome: 7	gilthead seabream	32.2	32.2	72%	162	100.00%	35997798
LR537122.1	Sparus aurata genome assembly, chromosome: 2	gilthead seabream	32.2	64.4	100%	162	100.00%	36133474
LR537144.1	Sparus aurata genome assembly, chromosome: 24	gilthead seabream	32.2	32.2	72%	162	100.00%	22435136
LR537143.1	Sparus aurata genome assembly, chromosome: 23	gilthead seabream	32.2	62.4	72%	162	100.00%	30526999
LR537141.1	Sparus aurata genome assembly, chromosome: 21	gilthead seabream	32.2	94.6	86%	162	100.00%	35160130
LR537139.1	Sparus aurata genome assembly, chromosome: 19	gilthead seabream	32.2	32.2	72%	162	100.00%	31047742
LR537138.1	Sparus aurata genome assembly, chromosome: 18	gilthead seabream	32.2	92.7	72%	162	100.00%	37104330
LR537127.1	Sparus aurata genome assembly, chromosome: 7	gilthead seabream	32.2	32.2	72%	162	100.00%	35997798
LR537122.1	Sparus aurata genome assembly, chromosome: 2	gilthead seabream	32.2	64.4	100%	162	100.00%	36133474
.....	.....	.....	.....	.....	.....	.....	.....	.....
XM_030051815.1	PREDICTED: Myripristis murdjan acidic mammalian chitinase-like (LOC115359376), mRNA	pinecone soldierfish	32.2	32.2	72%	162	100.00%	1538
.....	.....	.....	.....	.....	.....	.....	.....	.....
LR597567.1	Myripristis murdjan genome assembly, chromosome: 18	pinecone soldierfish	32.2	32.2	72%	162	100.00%	29228082
LR597564.1	Myripristis murdjan genome assembly, chromosome: 15	pinecone soldierfish	32.2	126	72%	162	100.00%	33424759
LR597555.1	Myripristis murdjan genome assembly, chromosome: 6	pinecone soldierfish	32.2	94.6	90%	162	100.00%	35357120
.....	.....	.....	.....	.....	.....	.....	.....	.....
LR584441.1	Salmo trutta genome assembly, chromosome: 12	river trout	32.2	32.2	72%	162	100.00%	97529106
LR584438.1	Salmo trutta genome assembly, chromosome: 11	river trout	32.2	62.4	77%	162	100.00%	22956941
LR584414.1	Salmo trutta genome assembly, chromosome: 33	river trout	32.2	96.6	77%	162	100.00%	44893998
LR584413.1	Salmo trutta genome assembly, chromosome: 39	river trout	32.2	32.2	72%	162	100.00%	25826328
LR584411.1	Salmo trutta genome assembly, chromosome: 14	river trout	32.2	32.2	72%	162	100.00%	86253525
LR584410.1	Salmo trutta genome assembly, chromosome: 1	river trout	32.2	276	95%	162	100.00%	81542925
LR584408.1	Salmo trutta genome assembly, chromosome: 32	river trout	32.2	92.7	90%	162	100.00%	44953446
.....	.....	.....	.....	.....	.....	.....	.....	.....
LR132001.1	Gouania willdenowi genome assembly, chromosome: 21	blunt-snouted clingfish	32.2	92.7	72%	162	100.00%	38978045
LR131995.1	Gouania willdenowi genome assembly, chromosome: 10	blunt-snouted clingfish	32.2	213	72%	162	100.00%	42970088
LR131987.1	Gouania willdenowi genome assembly, chromosome: 15	blunt-snouted clingfish	32.2	189	77%	162	100.00%	37027010
.....	.....	.....	.....	.....	.....	.....	.....	.....
CP025670.1	Ipomoea triloba cultivar NCNSP0323 chromosome 11	trilobed morning glory	32.2	32.2	72%	162	100.00%	25640617
CP025669.1	Ipomoea triloba cultivar NCNSP0323 chromosome 10	trilobed morning glory	32.2	62.4	72%	162	100.00%	29809665
CP025662.1	Ipomoea triloba cultivar NCNSP0323 chromosome 3	trilobed morning glory	32.2	32.2	72%	162	100.00%	33448119
CP025655.1	Ipomoea trifida cultivar NCNSP0306 chromosome 12	NA	32.2	62.4	72%	162	100.00%	24193007
CP025654.1	Ipomoea trifida cultivar NCNSP0306 chromosome 11	NA	32.2	32.2	72%	162	100.00%	19221892
CP025653.1	Ipomoea trifida cultivar NCNSP0306 chromosome 10	NA	32.2	92.7	72%	162	100.00%	24884449
CP025652.1	Ipomoea trifida cultivar NCNSP0306 chromosome 9	NA	32.2	32.2	72%	162	100.00%	23347646

CP025649.1	Ipomoea trifida cultivar NCNSP0306 chromosome 6	NA	32.2	32.2	72%	162	100.00%	2608917
CP025646.1	Ipomoea trifida cultivar NCNSP0306 chromosome 3	NA	32.2	32.2	72%	162	100.00%	2861972
.....	.....	.....	.....	.....	.....	.....	.....	.....
XM_015212391.2	PREDICTED: Solanum pennellii flavonol 4'-sulfotransferase-like (LOC107012525), mRNA	NA	32.2	32.2	72%	162	100.00%	1245
.....	.....	.....	.....	.....	.....	.....	.....	.....
CP012620.1	Oryza sativa Indica Group cultivar RP Bio-226 chromosome 12 sequence	long-grained rice	32.2	32.2	72%	162	100.00%	2138165
CP012615.1	Oryza sativa Indica Group cultivar RP Bio-226 chromosome 7 sequence	long-grained rice	32.2	62.4	72%	162	100.00%	2666061
CP012614.1	Oryza sativa Indica Group cultivar RP Bio-226 chromosome 6 sequence	long-grained rice	32.2	32.2	72%	162	100.00%	3090386
CP012611.1	Oryza sativa Indica Group cultivar RP Bio-226 chromosome 3 sequence	long-grained rice	32.2	62.4	77%	162	100.00%	3950426
CP012609.1	Oryza sativa Indica Group cultivar RP Bio-226 chromosome 1 sequence	long-grained rice	32.2	94.6	72%	162	100.00%	4452663
.....	.....	.....	.....	.....	.....	.....	.....	.....
HG975449.1	Solanum pennellii chromosome ch10, complete genome	NA	32.2	153	90%	162	100.00%	8252994
HG975448.1	Solanum pennellii chromosome ch09, complete genome	NA	32.2	306	100%	162	100.00%	8405750
HG975446.1	Solanum pennellii chromosome ch07, complete genome	NA	32.2	153	77%	162	100.00%	7929216
HG975444.1	Solanum pennellii chromosome ch05, complete genome	NA	32.2	157	95%	162	100.00%	7799110
HG975443.1	Solanum pennellii chromosome ch04, complete genome	NA	32.2	183	95%	162	100.00%	7719730
HG975442.1	Solanum pennellii chromosome ch03, complete genome	NA	32.2	122	100%	162	100.00%	7541401
.....	.....	.....	.....	.....	.....	.....	.....	.....

.....	.....	.....	.....	.....	.....	.....	.....	.....
CP056063.1	Oryza sativa Indica Group cultivar Zhenshan 97 chromosome 12	Oryza sativa Indica Group	32.2	32.2	72%	144	100.00%	2723243
CP056062.1	Oryza sativa Indica Group cultivar Zhenshan 97 chromosome 11	Oryza sativa Indica Group	32.2	92.7	95%	144	100.00%	3262430
CP056058.1	Oryza sativa Indica Group cultivar Zhenshan 97 chromosome 7	Oryza sativa Indica Group	32.2	94.6	72%	144	100.00%	3084055
CP056057.1	Oryza sativa Indica Group cultivar Zhenshan 97 chromosome 6	Oryza sativa Indica Group	32.2	32.2	72%	144	100.00%	3211991
CP056054.1	Oryza sativa Indica Group cultivar Zhenshan 97 chromosome 3	Oryza sativa Indica Group	32.2	62.4	77%	144	100.00%	3963616
CP054687.1	Oryza sativa Indica Group cultivar Minghui 63 chromosome 12	Oryza sativa Indica Group	32.2	32.2	72%	144	100.00%	2694288
CP054686.1	Oryza sativa Indica Group cultivar Minghui 63 chromosome 11	Oryza sativa Indica Group	32.2	92.7	95%	144	100.00%	3410058
CP054683.1	Oryza sativa Indica Group cultivar Minghui 63 chromosome 8	Oryza sativa Indica Group	32.2	32.2	72%	144	100.00%	3049230
CP054682.1	Oryza sativa Indica Group cultivar Minghui 63 chromosome 7	Oryza sativa Indica Group	32.2	94.6	72%	144	100.00%	3087707
CP054681.1	Oryza sativa Indica Group cultivar Minghui 63 chromosome 6	Oryza sativa Indica Group	32.2	32.2	72%	144	100.00%	3192118
CP054678.1	Oryza sativa Indica Group cultivar Minghui 63 chromosome 3	Oryza sativa Indica Group	32.2	62.4	77%	144	100.00%	3989325
.....	.....	.....	.....	.....	.....	.....	.....	.....
OD565841.1	4_Tbi_b3v08	Timema bartmani	32.2	32.2	72%	144	100.00%	151582
OE008157.1	4_Tte_b3v08	Timema tahoe	32.2	32.2	72%	144	100.00%	25833
OD113652.1	1_Tps_b3v08	Timema poppensis	32.2	32.2	72%	144	100.00%	835
OD000614.1	1_Tps_b3v08	Timema poppensis	32.2	32.2	72%	144	100.00%	174942
OE003899.1	4_Tte_b3v08	Timema tahoe	32.2	32.2	72%	144	100.00%	79256
OE001030.1	4_Tte_b3v08	Timema tahoe	32.2	32.2	72%	144	100.00%	208271
OC341357.1	3_Tce_b3v08	Timema cristinae	32.2	32.2	72%	144	100.00%	2591
OD567612.1	4_Tbi_b3v08	Timema bartmani	32.2	32.2	72%	144	100.00%	81314
OD566451.1	4_Tbi_b3v08	Timema bartmani	32.2	32.2	72%	144	100.00%	119606
OC322116.1	3_Tce_b3v08	Timema cristinae	32.2	32.2	72%	144	100.00%	36104
OB801225.1	3_Tms_b3v08	Timema	32.2	32.2	72%	144	100.00%	7147

		monikensis						
OC318741.1	3_Tce_b3v08	Timema cristinae	32.2	32.2	72%	144	100.00%	116317
OC318108.1	3_Tce_b3v08	Timema cristinae	32.2	32.2	72%	144	100.00%	150161
OD017994.1	1_Tps_b3v08	Timema poppensis	32.2	64.4	72%	144	100.00%	6204
OC317138.1	3_Tce_b3v08	Timema cristinae	32.2	32.2	72%	144	100.00%	272383
OC316733.1	3_Tce_b3v08	Timema cristinae	32.2	32.2	72%	144	100.00%	440874
OC316666.1	3_Tce_b3v08	Timema cristinae	32.2	32.2	72%	144	100.00%	506915
OC316580.1	3_Tce_b3v08	Timema cristinae	32.2	32.2	72%	144	100.00%	622769
OC316534.1	3_Tce_b3v08	Timema cristinae	32.2	32.2	72%	144	100.00%	723779
OB796106.1	3_Tms_b3v08	Timema monikensis	32.2	32.2	72%	144	100.00%	73987
OD009905.1	1_Tps_b3v08	Timema poppensis	32.2	32.2	72%	144	100.00%	17846
OB794560.1	3_Tms_b3v08	Timema monikensis	32.2	32.2	72%	144	100.00%	152572
OB793891.1	3_Tms_b3v08	Timema monikensis	32.2	32.2	72%	144	100.00%	222894
OB793863.1	3_Tms_b3v08	Timema monikensis	32.2	32.2	72%	144	100.00%	227279
OB793791.1	3_Tms_b3v08	Timema monikensis	32.2	32.2	72%	144	100.00%	237592
OB793273.1	3_Tms_b3v08	Timema monikensis	32.2	32.2	72%	144	100.00%	352819
OB793159.1	3_Tms_b3v08	Timema monikensis	32.2	32.2	72%	144	100.00%	397913
OB793041.1	3_Tms_b3v08	Timema monikensis	32.2	32.2	72%	144	100.00%	464793
OE002111.1	4_Tte_b3v08	Timema tahoe	32.2	32.2	72%	144	100.00%	135631
OE000844.1	4_Tte_b3v08	Timema tahoe	32.2	32.2	72%	144	100.00%	228278
OE000625.1	4_Tte_b3v08	Timema tahoe	32.2	32.2	72%	144	100.00%	263078
OE000398.1	4_Tte_b3v08	Timema tahoe	32.2	32.2	72%	144	100.00%	321069
OE000320.1	4_Tte_b3v08	Timema tahoe	32.2	32.2	72%	144	100.00%	340682
OD564725.1	4_Tbi_b3v08	Timema bartmani	32.2	32.2	72%	144	100.00%	298343
OD564660.1	4_Tbi_b3v08	Timema bartmani	32.2	32.2	72%	144	100.00%	319700
OD564385.1	4_Tbi_b3v08	Timema bartmani	32.2	32.2	72%	144	100.00%	504703
OD564279.1	4_Tbi_b3v08	Timema bartmani	32.2	32.2	72%	144	100.00%	852534
OD004884.1	1_Tps_b3v08	Timema poppensis	32.2	32.2	72%	144	100.00%	48018
OD000371.1	1_Tps_b3v08	Timema poppensis	32.2	32.2	72%	144	100.00%	212703
.....	.....	.....	.....	.....	.....	.....	.....	.....
OC007453.1	2_Tsi_b3v08	Timema shepardi	32.2	32.2	72%	144	100.00%	30778
OC004557.1	2_Tsi_b3v08	Timema shepardi	32.2	32.2	72%	144	100.00%	63513
OC004547.1	2_Tsi_b3v08	Timema shepardi	32.2	32.2	72%	144	100.00%	63626
OC003968.1	2_Tsi_b3v08	Timema shepardi	32.2	32.2	72%	144	100.00%	74580
OC002709.1	2_Tsi_b3v08	Timema shepardi	32.2	32.2	72%	144	100.00%	106344
OC000605.1	2_Tsi_b3v08	Timema shepardi	32.2	32.2	72%	144	100.00%	255980
.....	.....	.....	.....	.....	.....	.....	.....	.....
OA568639.1	1_Tps_b3v08	Timema douglasi	32.2	32.2	72%	144	100.00%	65513
OA567745.1	1_Tps_b3v08	Timema douglasi	32.2	32.2	72%	144	100.00%	81924
OA564956.1	1_Tps_b3v08	Timema douglasi	32.2	32.2	72%	144	100.00%	231413
OA564414.1	1_Tps_b3v08	Timema douglasi	32.2	32.2	72%	144	100.00%	406296
.....	.....	.....	.....	.....	.....	.....	.....	.....
XR_005026893.1	PREDICTED: Manis pentadactyla uncharacterized LOC118917760 (LOC118917760), ncRNA	Manis pentadactyla	32.2	32.2	72%	144	100.00%	3131
.....	.....	.....	.....	.....	.....	.....	.....	.....
MW109311.1	Influenza A virus (A/environment/Anhui/2.22_YHZGS004-E/2019(mixed)) segment 3 polymerase PA (PA) and PA-X protein (PA-X) genes, complete cds	Influenza A virus	32.2	32.2	72%	144	100.00%	2175
MW109303.1	Influenza A virus	Influenza A virus	32.2	32.2	72%	144	100.00%	2174

MW109287.1	(A/environment/Anhui/2.22_YHZGS003-E/2019(mixed)) segment 3 polymerase PA (PA) and PA-X protein (PA-X) genes, complete cds Influenza A virus (A/environment/Anhui/2.22_YHZGS001-E/2019(mixed)) segment 3 polymerase PA (PA) and PA-X protein (PA-X) genes, complete cds	Influenza A virus	32.2	32.2	72%	144	100.00%	2174
MW102519.1	Influenza A virus (A/chicken/Anhui/2.22_YHZGS018-C/2019(mixed)) segment 3 polymerase PA (PA) and PA-X protein (PA-X) genes, complete cds	Influenza A virus	32.2	32.2	72%	144	100.00%	2175
.....	.....	.....	.....	.....	.....	.....	.....	.....
LR877241.1	Acomys russatus genome assembly, chromosome: 30	Acomys russatus	32.2	122	100%	160	100.00%	40934070
LR877233.1	Acomys russatus genome assembly, chromosome: 22	Acomys russatus	32.2	32.2	72%	160	100.00%	59984821
LR877232.1	Acomys russatus genome assembly, chromosome: 21	Acomys russatus	32.2	32.2	72%	160	100.00%	61132863
LR877219.1	Acomys russatus genome assembly, chromosome: 8	Acomys russatus	32.2	92.7	77%	160	100.00%	77196262
LR877216.1	Acomys russatus genome assembly, chromosome: 5	Acomys russatus	32.2	62.4	86%	160	100.00%	79927262
LR877215.1	Acomys russatus genome assembly, chromosome: 4	Acomys russatus	32.2	94.6	81%	160	100.00%	85305823
.....	.....	.....	.....	.....	.....	.....	.....	.....
LR862376.1	Pipistrellus pipistrellus genome assembly, chromosome: 20	Pipistrellus pipistrellus	32.2	32.2	72%	160	100.00%	15565596
LR862373.1	Pipistrellus pipistrellus genome assembly, chromosome: 17	Pipistrellus pipistrellus	32.2	274	72%	160	100.00%	45571527
LR862372.1	Pipistrellus pipistrellus genome assembly, chromosome: 16	Pipistrellus pipistrellus	32.2	92.7	72%	160	100.00%	46231460
LR862367.1	Pipistrellus pipistrellus genome assembly, chromosome: 11	Pipistrellus pipistrellus	32.2	183	90%	160	100.00%	69386144
LR862363.1	Pipistrellus pipistrellus genome assembly, chromosome: 7	Pipistrellus pipistrellus	32.2	155	86%	160	100.00%	84548104
LR862362.1	Pipistrellus pipistrellus genome assembly, chromosome: 6	Pipistrellus pipistrellus	32.2	304	72%	160	100.00%	85793147
.....	.....	.....	.....	.....	.....	.....	.....	.....
AC146523.2	Oryza sativa Japonica Group cultivar Nipponbare chromosome 11 clone OSJNBa0045D11 map near 50283S, complete sequence	Oryza sativa Japonica Group	32.2	32.2	72%	160	100.00%	148306
.....	.....	.....	.....	.....	.....	.....	.....	.....

Bacillus anthracis strain FDAARGOS_702 chromosome

GenBank: CP054800.1

FASTA Graphics

Go to:

LOCUS CP054800 19 bp DNA linear BCT **17-JUN-2020**

DEFINITION Bacillus anthracis strain FDAARGOS_702 chromosome.

ACCESSION CP054800 REGION: 208258..208276

VERSION CP054800.1

DBLINK BioProject: PRJNA231221
BioSample: SAMN11056417

KEYWORDS .

SOURCE Bacillus anthracis

ORGANISM Bacillus anthracis
Bacteria; Firmicutes; Bacilli; Bacillales; Bacillaceae; Bacillus;
Bacillus cereus group.

REFERENCE 1 (bases 1 to 19)

AUTHORS Sozhamannan,S., Rosenzweig,N., Tallon,L., Sadzewicz,L.,
Vavikolanu,K., Mehta,A., Aluvathingal,J., Nadendla,S., Yan,Y. and
Sichtig,H.

TITLE FDA dAtabase for Regulatory Grade micrObial Sequences (FDA-ARGOS): Supporting development and validation of Infectious Disease Dx tests

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 19)

AUTHORS Sozhamannan,S., Rosenzweig,N., Tallon,L., Sadzewicz,L.,
Vavikolanu,K., Mehta,A., Aluvathingal,J., Nadendla,S., Yan,Y. and
Sichtig,H.

TITLE Direct Submission

JOURNAL **Submitted (25-NOV-2019) Center for Devices and Radiological Health, US Food and Drug Administration, 10903 New Hampshire Avenue, Silver Spring, MD 20993-0002, USA**

COMMENT Assembly not circularized.
The annotation was added by the NCBI Prokaryotic Genome Annotation
Pipeline (PGAP). Information about PGAP can be found here:
https://www.ncbi.nlm.nih.gov/genome/annotation_prok/
##Genome-Assembly-Data-START##
Assembly Method :: SMRT v. 2.3.0, HGAP v. 3, SPAdes v. 3.6.0
Genome Coverage :: 923.93x
Sequencing Technology :: Pacbio; Illumina
##Genome-Assembly-Data-END##

##Genome-Annotation-Data-START##
Annotation Provider :: NCBI
Annotation Date :: 06/10/2020 00:01:37
Annotation Pipeline :: NCBI Prokaryotic Genome

```
                                                Annotation Pipeline (PGAP)
            Annotation Method                  :: Best-placed reference protein
                                                set; GeneMarkS-2+
            Annotation Software revision       :: 4.11
            Features Annotated                 :: Gene; CDS; rRNA; tRNA; ncRNA;
                                                repeat_region
            Genes (total)                      :: 5,533
            CDSs (total)                       :: 5,400
            Genes (coding)                     :: 5,101
            CDSs (with protein)                :: 5,101
            Genes (RNA)                        :: 133
            rRNAs                              :: 11, 11, 11 (5S, 16S, 23S)
            complete rRNAs                     :: 11, 11, 11 (5S, 16S, 23S)
            tRNAs                              :: 95
            ncRNAs                             :: 5
            Pseudo Genes (total)               :: 299
            CDSs (without protein)             :: 299
            Pseudo Genes (ambiguous residues)  :: 0 of 299
            Pseudo Genes (frameshifted)        :: 215 of 299
            Pseudo Genes (incomplete)          :: 38 of 299
            Pseudo Genes (internal stop)       :: 67 of 299
            Pseudo Genes (multiple problems)   :: 18 of 299
            ##Genome-Annotation-Data-END##
FEATURES             Location/Qualifiers
     source          1..19
                     /organism="Bacillus anthracis"
                     /mol_type="genomic DNA"
                     /strain="FDAARGOS_702"
                     /culture_collection="FDA:FDAARGOS_702"
                     /db_xref="taxon:1392"
                     /country="USA: MD"
                     /collected_by="U.S. Army Edgewood Chemical Biological
                      Center"
                     /note="genetically modified strain"
     gene            complement(<1..>19)
                     /locus_tag="FOC10_01165"
     CDS             complement(<1..>19)
                     /locus_tag="FOC10_01165"
                     /inference="COORDINATES: similar to AA
                     sequence:RefSeq:WP_000849936.1"
                     /note="Derived by automated computational analysis using
                     gene prediction method: Protein Homology."
                     /codon_start=3
                     /transl_table=11
                     /product="serine hydrolase"
```

/protein_id="QKU12340.1"
/translation="MKTKDQIDNIVKETYRHIDFSGVILVKEKKGIVYEEVFGYANRN
ECINNTLQTRFGIASGCKIFTAIGICQLVENGVITFQTKLKECLSVNFPNFNEDITIH
HLLTHSSGIPDYFDESIMDNFEDLWKQTPMYLLKSLKDFLPLFQNRDMMYAPGSKFHY
NNAGFIILGLIIEEQTGLTFTEYIEKNIFNPIGMNHSGYFSLDRLPRQTALGYIKDEI
SQTWRTNAYSIPIKGGSDGGAFITASDMLKFWEALFNYEIISQEYTKILLTPHIQVNN
NQSYGYGIWIETRENKIF**KYHVMGY**DPGVSFRSAVYPDLGITLVIPSNKGAGAEKLMI
EIEGSF"

ORIGIN

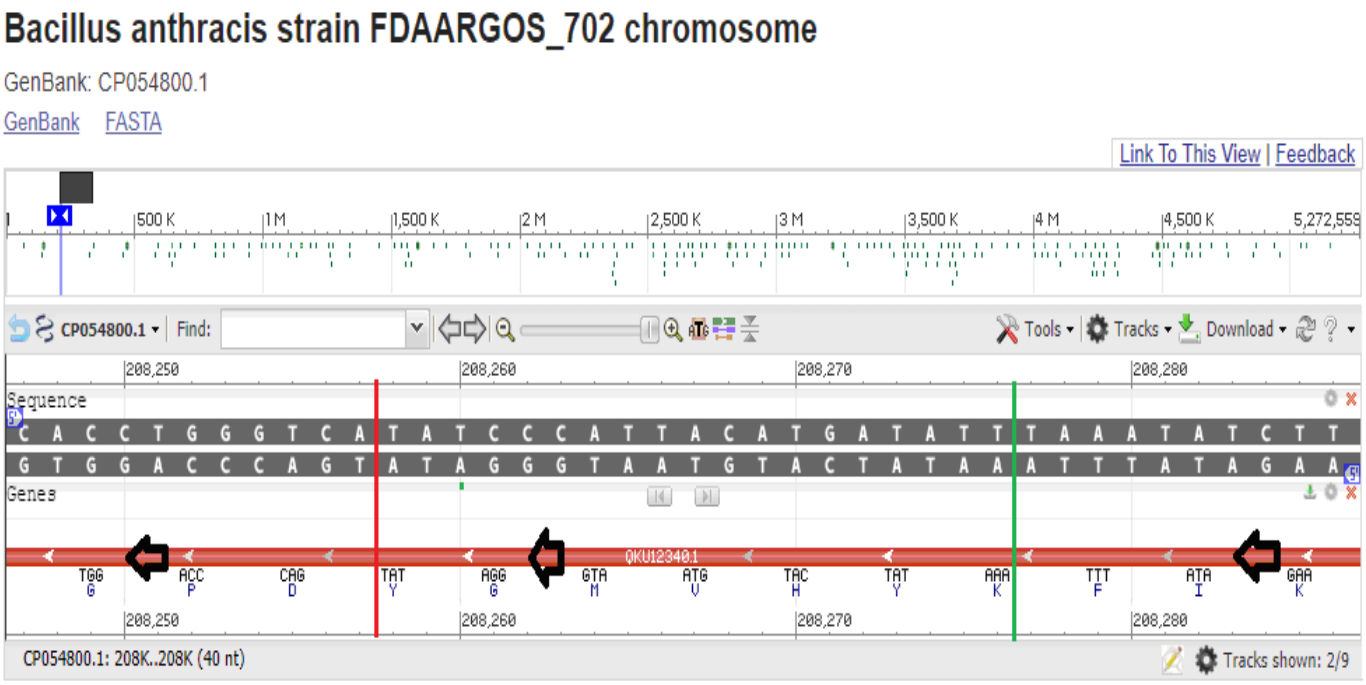
1 **tatcccatta catgatatt**

//

Bacillus anthracis strain FDAARGOS_702 chromosome
Sequence ID: CP054800.1 **Length: 5272559 Number of Matches: 1**
Range 1: 208258 to 208276 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
38.2 bits(19)	2.2	19/19(100%)	0/19(0%)	Plus/Plus

```
CDS: Putative 1        5          G  M  V  H  Y
Query                  1       TATCCCATTACATGATATT  19
                               ||||||||||||||||||
Sbjct                  208258  TATCCCATTACATGATATT  208276
CDS:serine hydrolase   301        Y  G  M  V  H  Y  K
```



Bacillus anthracis strain FDAARGOS_703 chromosome

GenBank: CP054797.1

FASTA Graphics

Go to:

LOCUS CP054797 19 bp DNA linear BCT **17-JUN-2020**

DEFINITION Bacillus anthracis strain FDAARGOS_703 chromosome.

ACCESSION CP054797 REGION: 65981..65999

VERSION CP054797.1

DBLINK BioProject: PRJNA231221
BioSample: SAMN11056418

KEYWORDS .

SOURCE Bacillus anthracis

ORGANISM Bacillus anthracis
Bacteria; Firmicutes; Bacilli; Bacillales; Bacillaceae; Bacillus;
Bacillus cereus group.

REFERENCE 1 (bases 1 to 19)

AUTHORS Sozhamannan,S., Rosenzweig,N., Tallon,L., Sadzewicz,L.,
Vavikolanu,K., Mehta,A., Aluvathingal,J., Nadendla,S., Yan,Y. and
Sichtig,H.

TITLE FDA dAtabase for Regulatory Grade micrObial Sequences (FDA-ARGOS): Supporting development and validation of Infectious Disease Dx tests

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 19)

AUTHORS Sozhamannan,S., Rosenzweig,N., Tallon,L., Sadzewicz,L.,
Vavikolanu,K., Mehta,A., Aluvathingal,J., Nadendla,S., Yan,Y. and
Sichtig,H.

TITLE Direct Submission

JOURNAL **Submitted (25-NOV-2019) Center for Devices and Radiological Health, US Food and Drug Administration, 10903 New Hampshire Avenue, Silver Spring, MD 20993-0002, USA**

COMMENT Assembly not circularized.
The annotation was added by the NCBI Prokaryotic Genome Annotation Pipeline (PGAP). Information about PGAP can be found here:
https://www.ncbi.nlm.nih.gov/genome/annotation_prok/

##Genome-Assembly-Data-START##
Assembly Method :: SMRT v. 2.3.0, HGAP v. 3, SPAdes v. 3.6.0
Genome Coverage :: 942.4x
Sequencing Technology :: Pacbio; Illumina
##Genome-Assembly-Data-END##
##Genome-Annotation-Data-START##
Annotation Provider :: NCBI
Annotation Date :: 06/10/2020 00:12:20
Annotation Pipeline :: NCBI Prokaryotic Genome

```
                                                Annotation Pipeline (PGAP)
            Annotation Method                 :: Best-placed reference protein
                                                set; GeneMarkS-2+
            Annotation Software revision      :: 4.11
            Features Annotated                :: Gene; CDS; rRNA; tRNA; ncRNA;
                                                repeat_region
            Genes (total)                     :: 5,514
            CDSs (total)                      :: 5,381
            Genes (coding)                    :: 5,095
            CDSs (with protein)               :: 5,095
            Genes (RNA)                       :: 133
            rRNAs                             :: 11, 11, 11 (5S, 16S, 23S)
            complete rRNAs                    :: 11, 11, 11 (5S, 16S, 23S)
            tRNAs                             :: 95
            ncRNAs                            :: 5
            Pseudo Genes (total)              :: 286
            CDSs (without protein)            :: 286
            Pseudo Genes (ambiguous residues) :: 0 of 286
            Pseudo Genes (frameshifted)       :: 201 of 286
            Pseudo Genes (incomplete)         :: 36 of 286
            Pseudo Genes (internal stop)      :: 71 of 286
            Pseudo Genes (multiple problems)  :: 19 of 286
            ##Genome-Annotation-Data-END##
FEATURES             Location/Qualifiers
     source          1..19
                     /organism="Bacillus anthracis"
                     /mol_type="genomic DNA"
                     /strain="FDAARGOS_703"
                     /culture_collection="FDA:FDAARGOS_703"
                     /db_xref="taxon:1392"
                     /country="USA: MD"
                     /collected_by="U.S. Army Edgewood Chemical Biological
                       Center"
                     /note="genetically modified strain"
     gene            <1..>19
                     /locus_tag="FOC11_00315"
     CDS             <1..>19
                     /locus_tag="FOC11_00315"
                     /inference="COORDINATES: similar to AA
                     sequence:RefSeq:WP_000849936.1"
                     /note="Derived by automated computational analysis using
                     gene prediction method: Protein Homology."
                     /codon_start=3
                     /transl_table=11
                     /product="serine hydrolase"
```

/protein_id="QKU07095.1"
/translation="MKTKDQIDNIVKETYRHIDFSGVILVKEKKGIVYEEVFGYANRN
ECINNTLQTRFGIASGCKIFTAIGICQLVENGVITFQTKLKECLSVNFPNFNEDITIH
HLLTHSSGIPDYFDESIMDNFEDLWKQTPMYLLKSLKDFLPLFQNRDMMYAPGSKFHY
NNAGFIILGLIIEEQTGLTFTEYIEKNIFNPIGMNHSGYFSLDRLPRQTALGYIKDEI
SQTWRTNAYSIPIKGGSDGGAFITASDMLKFWEALFNYEIISQEYTKILLTPHIQVNN
NQSYGYGIWIETRENKIF**KYHVMGY**DPGVSFRSAVYPDLGITLVIPSNKGAGAEKLMI
EIEGSF"

ORIGIN

1 **aatatcatgt aatgggata**

//

ATTENTION:

The sequence "**aatatcatgt aatgggata**" shown above is the *reverse and mirror sequence* of the sequence "**TATCCCATTACATGATATT**" (partial sequence of **Sequence n° 3/1**, "**TATCCCATTACATGATATTATT**"):

ATAGGGTAATGTACTATAA

||||||||||||||||||||

TATCCCATTACATGATATT

Bacillus anthracis strain FDAARGOS_703 chromosome
Sequence ID: CP054797.1 **Length: 5261520 Number of Matches: 1**
Range 1: 65981 to 65999 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
38.2 bits(19)	2.2	19/19(100%)	0/19(0%)	Plus/Minus

```
CDS: Putative 1       5          G  M  V  H  Y
Query                 1      TATCCCATTACATGATATT  19
                             ||||||||||||||||||
Sbjct                 65999  TATCCCATTACATGATATT  65981
CDS:serine hydrolase  301    Y  G  M  V  H  Y  K
```

Bacillus anthracis strain FDAARGOS_703 chromosome

GenBank: CP054797.1

GenBank FASTA

Bacillus paranthracis strain NCCP 15910 chromosome, complete genome

GenBank: CP041981.1

FASTA Graphics

Go to:

LOCUS CP041981 19 bp DNA linear BCT **15-JAN-2020**

DEFINITION Bacillus paranthracis strain NCCP 15910 chromosome, complete genome.

ACCESSION CP041981 REGION: 2708037..2708055

VERSION CP041981.1

DBLINK BioProject: PRJNA556101
BioSample: SAMN12340779

KEYWORDS .

SOURCE Bacillus paranthracis

ORGANISM Bacillus paranthracis
Bacteria; Firmicutes; Bacilli; Bacillales; Bacillaceae; Bacillus;
Bacillus cereus group.

REFERENCE 1 (bases 1 to 19)
AUTHORS Yu,W.S., Cheong,H.-M., Choi,Y., Hwang,K.J., Jung,K., Lee,S. and Choi,C.
TITLE Direct Submission
JOURNAL Submitted (23-JUL-2019) Pathogen Resource TF, Center for Infectious Diseases Research, Korea National Institute of Health, Korea Centers for Disease Control & Prevention (KCDC), Osong Health Technology Administration Complex, 187, Osongsaengmyeong 2-ro, Osong-eup, Heungdeok-gu, Cheongju-si, Chungcheongbuk-do 28159, South Korea
COMMENT The annotation was added by the NCBI Prokaryotic Genome Annotation Pipeline (PGAP). Information about PGAP can be found here: https://www.ncbi.nlm.nih.gov/genome/annotation_prok/

##Genome-Assembly-Data-START##
Assembly Date :: JUL-2017
Assembly Method :: SPAdes v. 3.1
Genome Representation :: Full
Expected Final Version :: Yes
Genome Coverage :: 207.0x
Sequencing Technology :: IonTorrent; PacBio
##Genome-Assembly-Data-END##

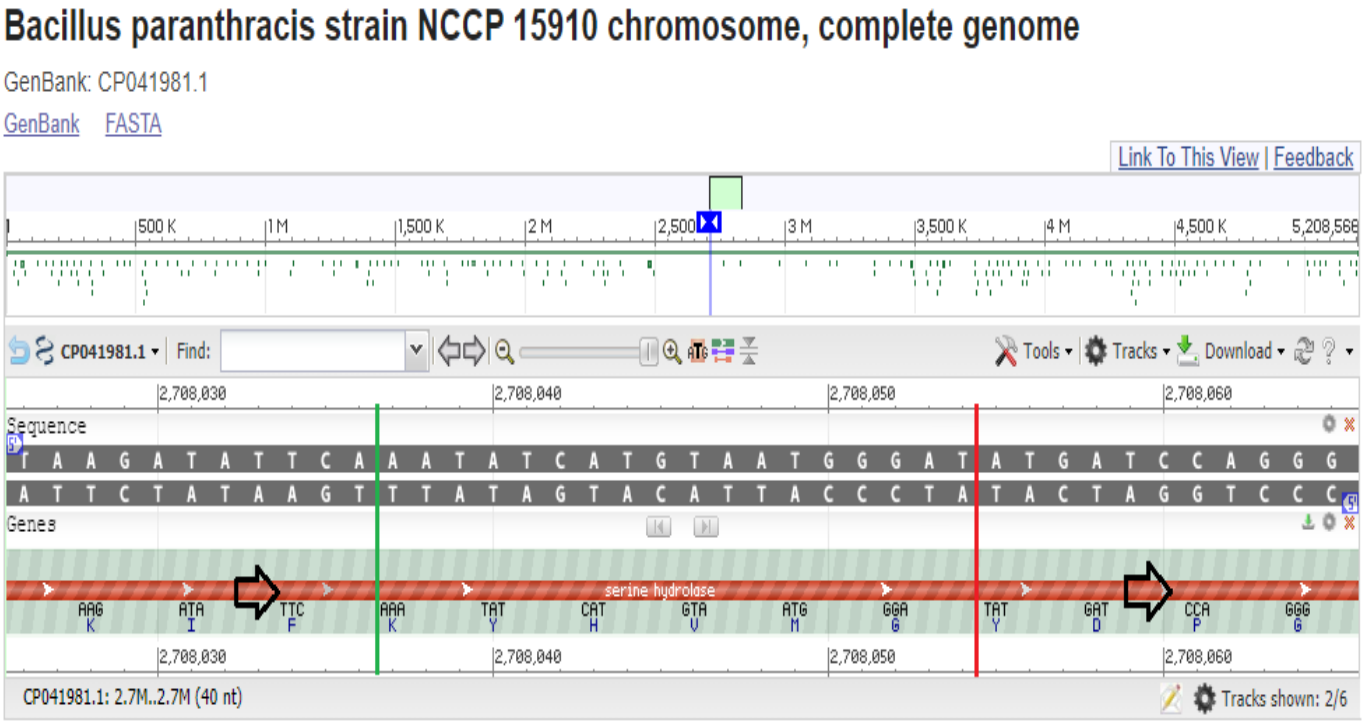
##Genome-Annotation-Data-START##
Annotation Provider :: NCBI
Annotation Date :: 07/24/2019 15:40:42
Annotation Pipeline :: NCBI Prokaryotic Genome Annotation Pipeline (PGAP)
Annotation Method :: Best-placed reference protein set; GeneMarkS-2+
Annotation Software revision :: 4.8
Features Annotated :: Gene; CDS; rRNA; tRNA; ncRNA; repeat_region
Genes (total) :: 5,737
CDSs (total) :: 5,589
Genes (coding) :: 4,975
CDSs (with protein) :: 4,975
Genes (RNA) :: 148
rRNAs :: 13, 13, 13 (5S, 16S, 23S)
complete rRNAs :: 13, 13, 13 (5S, 16S, 23S)
tRNAs :: 104
ncRNAs :: 5
Pseudo Genes (total) :: 614
CDSs (without protein) :: 614
Pseudo Genes (ambiguous residues) :: 0 of 614

```
            Pseudo Genes (frameshifted)       :: 497 of 614
            Pseudo Genes (incomplete)         :: 114 of 614
            Pseudo Genes (internal stop)      :: 65 of 614
            Pseudo Genes (multiple problems)  :: 53 of 614
            ##Genome-Annotation-Data-END##
FEATURES             Location/Qualifiers
     source          1..19
                     /organism="Bacillus paranthracis"
                     /mol_type="genomic DNA"
                     /strain="NCCP 15910"
                     /isolation_source="stool"
                     /host="Homo sapiens"
                     /culture_collection="NCCP:15910"
                     /db_xref="taxon:2026186"
                     /country="South Korea: Seoul"
                     /lat_lon="37.547889 N 126.941889 E"
                     /collection_date="2013-05-20"
                     /collected_by="Korea National Institute of Health"
     gene            <1..>19
                     /locus tag="FPL02 15370"
                     /pseudo
     CDS             <1..>19
                     /locus_tag="FPL02_15370"
                     /inference="COORDINATES: similar to AA
                     sequence:RefSeq:WP_000849936.1"
                     /note="incomplete; partial on complete genome; missing
                     C-terminus; Derived by automated computational analysis
                     using gene prediction method: Protein Homology."
                     /pseudo
                     /codon_start=3
                     /transl_table=11
                     /product="serine hydrolase"
ORIGIN
        1 aatatcatgt aatgggata
//
```

Bacillus paranthracis strain NCCP 15910 chromosome, complete genome
Sequence ID: CP041981.1 Length: 5208566 Number of Matches: 2
Range 1: 2708037 to 2708055 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
38.2 bits(19)	2.3	19/19(100%)	0/19(0%)	Plus/Minus

```
CDS: Putative 1    5          G  M  V  H  Y
Query              1          TATCCCATTACATGATATT  19
                              ||||||||||||||||||||
Sbjct              2708055    TATCCCATTACATGATATT  2708037
CDS: serine hydr   301        Y  G  M  V  H  Y  K
```



Range 2: 1645460 to 1645475 GenBank Graphics

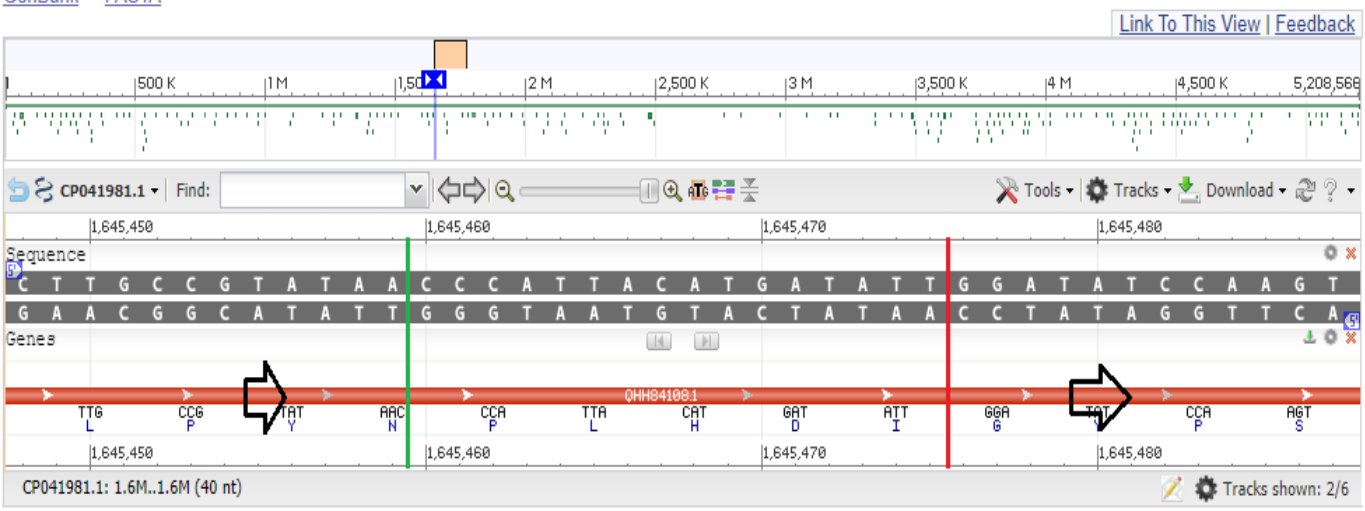
Score	Expect	Identities	Gaps	Strand
32.2 bits(16)	144	16/16(100%)	0/16(0%)	Plus/Plus

```
CDS: Putative 1        1          P  L  H  D  I
Query                  4          CCCATTACATGATATT  19
                                  ||||||||||||||||
Sbjct                  1645460    CCCATTACATGATATT  1645475
CDS:phosphoadenylyl-   192        P  L  H  D  I
```

Bacillus paranthracis strain NCCP 15910 chromosome, complete genome

GenBank: CP041981.1

GenBank FASTA



PREDICTED: Rhinolophus ferrumequinum methyltransferase like 25 (METTL25), transcript variant X2, mRNA

NCBI Reference Sequence: XM_033117539.1

FASTA Graphics

Go to:

LOCUS XM_033117539 16 bp mRNA linear MAM **23-MAR-2020**

DEFINITION PREDICTED: Rhinolophus ferrumequinum methyltransferase like 25 (METTL25), transcript variant X2, mRNA.

ACCESSION XM_033117539 REGION: 1140..1155

VERSION XM_033117539.1

DBLINK BioProject: PRJNA613604

KEYWORDS RefSeq.

SOURCE Rhinolophus ferrumequinum (greater horseshoe bat)

ORGANISM Rhinolophus ferrumequinum

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Laurasiatheria; Chiroptera; Microchiroptera; Rhinolophidae; Rhinolophinae; Rhinolophus.

COMMENT MODEL REFSEQ: This record is predicted by automated computational analysis. This record is derived from a genomic sequence (NC_046293.1) annotated using gene prediction method: Gnomon, supported by mRNA evidence.
Also see:
Documentation of NCBI's Annotation Process

##Genome-Annotation-Data-START##
Annotation Provider :: NCBI
Annotation Status :: Full annotation
Annotation Name :: Rhinolophus ferrumequinum Annotation Release 100
Annotation Version :: 100
Annotation Pipeline :: NCBI eukaryotic genome annotation pipeline
Annotation Software Version :: 8.4
Annotation Method :: Best-placed RefSeq; Gnomon
Features Annotated :: Gene; mRNA; CDS; ncRNA
##Genome-Annotation-Data-END##

FEATURES Location/Qualifiers
source 1..16
/organism="Rhinolophus ferrumequinum"
/mol_type="mRNA"
/isolate="MPI-CBG mRhiFer1"
/db_xref="taxon:59479"
/chromosome="10"
/sex="male"
/tissue type="liver"
/country="United Kingdom: Woodchester Mansion, Gloucestershire"
/lat_lon="51.7108 N 2.2776 W"
/collection_date="2016-05-31"
/collected_by="Gareth Jones"
gene <1..>16
/gene="METTL25"
/note="Derived by automated computational analysis using gene prediction method: Gnomon. Supporting evidence includes similarity to: 2 mRNAs, and 100% coverage of the annotated genomic feature by RNAseq alignments, including 2 samples with support for all annotated introns"
/db_xref="GeneID:117028680"
CDS <1..>16
/gene="METTL25"
/codon_start=2
/product="**methyltransferase-like protein 25 isoform X2**"

/protein_id="XP_032973430.1"
/db_xref="GeneID:117028680"
/translation="MAASCPLPVTQDLSTLRAKLQGLLRFLREALSISNAHTVDFYTE
SVWEQLIDLPPETVLAVLKKSTEEAEARPSEARPLVEAERASGITDFPKIFCETSQKL
LSVEAFALAAKYYSLQNLGVCTPLEQLLVALRGNQTQKTGENVKPDEFMNLKKSHEVQ
AMSELISSIADYYGIKQVIDLGSGKGYLSSFLSLKYGLKVYGIDSSNTNTHGAQERNR
KLKKHWKICHTQSKSDVNGLALKMAKGKKMQGEIPYKANTEEVCNNSPTSQEMMSTSE
FLPDFSVSVISNIRKQMENLHVCSHREENLCFANAFSLIDLLPIDAIEPTSSSRMPKR
KMLEANKERRKTTSKSTESNIYSPLTSFVTADSELHDIIKDLEDCLMVGLHTCGDLAP
NTLRIFTSKSEIKGVCSVGCCYHLLSEEFEDPHKECTQEKWGFPMCHYLKEERWCCGR
NARMSACLALERVAVGRGLPTESLFYRAVLQDIIKDCYGITKCDRHVGKIYSKSSSFL
DYVRKSLRKLGLDESKVRKLKHKWTA"

ORIGIN

1 attacatgat attatt

//

PREDICTED: Rhinolophus ferrumequinum methyltransferase like 25 (METTL25), transcript variant X2, mRNA
Sequence ID: XM_033117539.1 Length: 1884 Number of Matches: 1
Range 1: 1140 to 1155 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
32.2 bits(16)	160	16/16(100%)	0/16(0%)	Plus/Plus

```
CDS: Putative 1        1        L  H  D  I  I
Query                  7     ATTACATGATATTATT  22
                             ||||||||||||||||
Sbjct                  1140  ATTACATGATATTATT  1155
CDS:methyltransferas   369      L  H  D  I  I
```

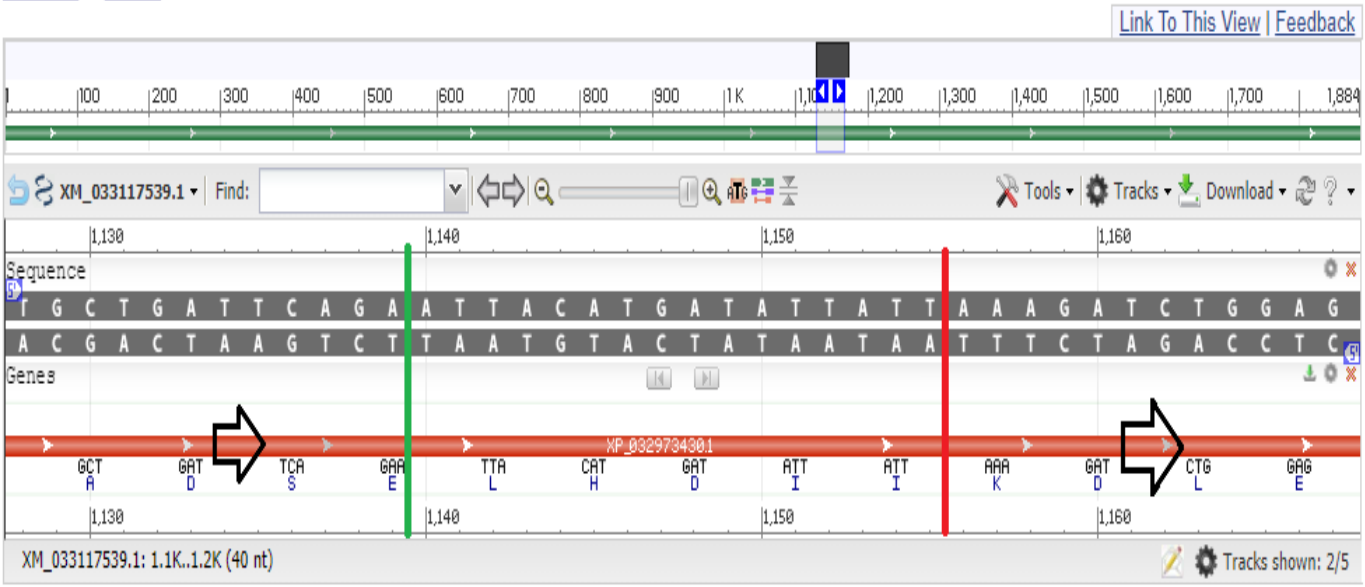
Methyltransferase Like 25

A **Methyltransferase** is an enzyme that is part of the class *transferase* can transfer a *methyl group* from a *donor molecule* to an *acceptor*. ***Methyltransferase Like 25*** (**METTL25**) is a Protein Coding gene. Diseases associated with METTL25 include Optic Atrophy 7 With Or Without Auditory Neuropathy. Gene Ontology (GO) annotations related to this gene include *methyltransferase activity*. An important paralog of this gene is **RRNAD1** (https://www.genecards.org/cgi-bin/carddisp.pl?gene=METTL25#:~:text=METTL25%20(Methyltransferase%20Like%2025)%20is,of%20this%20gene%20is%20RRNAD1). **RRNAD1** (*Ribosomal RNA Adenine Dimethylase Domain Containing 1*) is a Protein Coding gene. Gene Ontology (GO) annotations related to this gene include *rRNA (adenine-N6,N6-)-dimethyltransferase activity* (https://www.genecards.org/cgi-bin/carddisp.pl?gene=RRNAD1). See also the discussions on pages 208 and 433.

PREDICTED: Rhinolophus ferrumequinum methyltransferase like 25 (METTL25), transcript variant X2, mRNA

NCBI Reference Sequence: XM_033117539.1

GenBank FASTA



PREDICTED: Rhinolophus ferrumequinum methyltransferase like 25 (METTL25), transcript variant X1, mRNA

NCBI Reference Sequence: XM_033117538.1

FASTA Graphics

Go to:

LOCUS XM_033117538 16 bp mRNA linear MAM **23-MAR-2020**

DEFINITION PREDICTED: Rhinolophus ferrumequinum methyltransferase like 25 (METTL25), transcript variant X1, mRNA.

ACCESSION XM_033117538 REGION: 1140..1155

VERSION XM_033117538.1

DBLINK BioProject: PRJNA613604

KEYWORDS RefSeq.

SOURCE Rhinolophus ferrumequinum (greater horseshoe bat)

ORGANISM Rhinolophus ferrumequinum

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Laurasiatheria; Chiroptera; Microchiroptera;

Rhinolophidae; Rhinolophinae; Rhinolophus.

COMMENT MODEL REFSEQ: This record is predicted by automated computational analysis. This record is derived from a genomic sequence (NC_046293.1) annotated using gene prediction method: Gnomon, supported by mRNA evidence.

Also see:

Documentation of NCBI's Annotation Process

##Genome-Annotation-Data-START##

Annotation Provider :: NCBI

Annotation Status :: Full annotation

Annotation Name :: Rhinolophus ferrumequinum Annotation Release 100

Annotation Version :: 100

Annotation Pipeline :: NCBI eukaryotic genome annotation pipeline

Annotation Software Version :: 8.4

Annotation Method :: Best-placed RefSeq; Gnomon

Features Annotated :: Gene; mRNA; CDS; ncRNA

##Genome-Annotation-Data-END##

FEATURES Location/Qualifiers

source 1..16

/organism="Rhinolophus ferrumequinum"

/mol_type="mRNA"

/isolate="MPI-CBG mRhiFer1"

/db_xref="taxon:59479"

/chromosome="10"

/sex="male"

/tissue type="liver"

/country="United Kingdom: Woodchester Mansion, Gloucestershire"

/lat_lon="51.7108 N 2.2776 W"

/collection_date="2016-05-31"

/collected_by="Gareth Jones"

gene <1..>16

/gene="METTL25"

/note="Derived by automated computational analysis using gene prediction method: Gnomon. Supporting evidence includes similarity to: 4 mRNAs, 6 Proteins, and 100% coverage of the annotated genomic feature by RNAseq alignments, including 5 samples with support for all annotated introns"

/db_xref="GeneID:117028680"

CDS <1..>16

/gene="METTL25"

/codon_start=2

/product="**methyltransferase-like protein 25 isoform X1**"
/protein_id="XP_032973429.1"
/db_xref="GeneID:117028680"
/translation="MAASCPLPVTQDLSTLRAKLQGLLRFLREALSISNAHTVDFYTE
SVWEQLIDLPPETVLAVLKKSTEEAEARPSEARPLVEAERASGITDFPKIFCETSQKL
LSVEAFALAAKYYSLQNLGVCTPLEQLLVALRGNQTQKTGENVKPDEFMNLKKSHEVQ
AMSELISSIADYYGIKQVIDLGSGKGYLSSFLSLKYGLKVYGIDSSNTNTHGAQERNR
KLKKHWKICHTQSKSDVNGLALKMAKGKKMQGEIPYKANTEEVCNNSPTSQEMMSTSE
FLPDFSVSVISNIRKQMENLHVCSHREENLCFANAFSLIDLLPIDAIEPTSSSRMPKR
KMLEANKERRKTTSKSTESNIYSPLTSFVTADSELHDIIKDLEDCLMVGLHTCGDLAP
NTLRIFTSKSEIKGVCSVGCCYHLLSEEFEDPHKECTQEKWGFPMCHYLKEERWCCGR
NARMSACLALERVAVGRGLPTESLFYRAVLQDIIKDCYGITKCDRHVGKIYSKSSSFL
DYVRKSLRKLGLDESKLPEKLIMDYYEKYKPRMNELEAFNMLKVVLAPCIETLILLDR
LCYLKEQEDIAWSALVKLFDPVKSPRCYAVIALKQQQ"

ORIGIN

1 attacatgat attatt

//

PREDICTED: Rhinolophus ferrumequinum methyltransferase like 25 (METTL25), transcript variant X1, mRNA
Sequence ID: XM_033117538.1Length: 1994Number of Matches: 1
Range 1: 1140 to 1155 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
32.2 bits(16)	160	16/16(100%)	0/16(0%)	Plus/Plus

```
CDS: Putative 1       1       L  H  D  I  I
Query                 7     ATTACATGATATTATT  22
                            ||||||||||||||||
Sbjct                 1140  ATTACATGATATTATT  1155
CDS:methyltransferas  369     L  H  D  I  I
```

PREDICTED: Manis pentadactyla uncharacterized LOC118917760 (LOC118917760), ncRNA

NCBI Reference Sequence: XR_005026893.1

FASTA Graphics

Go to:

LOCUS XR_005026893 3131 bp RNA linear MAM **28-OCT-2020**

DEFINITION PREDICTED: Manis pentadactyla uncharacterized LOC118917760 (LOC118917760), ncRNA.

ACCESSION XR_005026893

VERSION XR_005026893.1

DBLINK BioProject: PRJNA666705

KEYWORDS RefSeq.

SOURCE Manis pentadactyla (Chinese pangolin)

ORGANISM Manis pentadactyla

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Laurasiatheria; Pholidota; Manidae; Manis.

COMMENT MODEL REFSEQ: This record is predicted by automated computational analysis. This record is derived from a genomic sequence (NW_023458232.1) annotated using gene prediction method: Gnomon.

Also see:

Documentation of NCBI's Annotation Process

##Genome-Annotation-Data-START##

Annotation Provider :: NCBI

Annotation Status :: Full annotation

Annotation Name :: Manis pentadactyla Annotation Release 100

Annotation Version :: 100

Annotation Pipeline :: NCBI eukaryotic genome annotation pipeline

Annotation Software Version :: 8.5

Annotation Method :: Best-placed RefSeq; Gnomon

Features Annotated :: Gene; mRNA; CDS; ncRNA

##Genome-Annotation-Data-END##

FEATURES Location/Qualifiers

source 1..3131

/organism="Manis pentadactyla"

/mol_type="transcribed RNA"

/isolate="MP20"

/db_xref="taxon:143292"

/chromosome="Unknown"

/dev_stage="adult"

/collection_date="2017"

gene 1..3131

/gene="LOC118917760"

/note="Derived by automated computational analysis using gene prediction method: Gnomon. Supporting evidence includes similarity to: 100% coverage of the annotated genomic feature by RNAseq alignments"
/db_xref="GeneID:118917760"

ncRNA 1..3131

/ncRNA_class="lncRNA"

/gene="LOC118917760"

/product="uncharacterized LOC118917760"

/db_xref="GeneID:118917760"

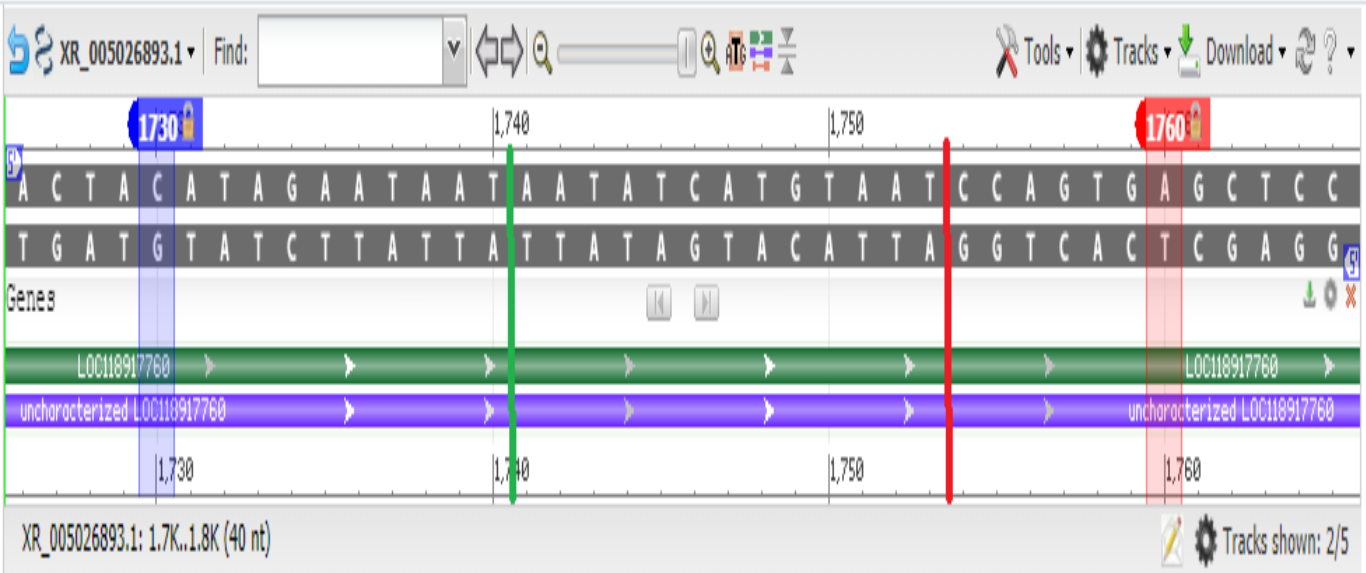
ORIGIN

```
        1 cactctcccc accaccacag ggcaggcccc tcttggtttg ctcccacatc agttctccat
       61 tcctcgctgt cctgggctct gaggccgact gtacagacca aaccacacgg ggcccctcac
      121 cctttggttt cccacacact cagccaaagg gaggctggaa ggggatgtgg gggctgagga
      181 gggagaggcg gggcttgttc cctccccacc tgggtgcaac tcggcagggg accctccaac
      241 cttggccaca gctgccgagc cacaccacgc cgttccctct ccggtcccct caggcctgga
      301 ggggactatc ccaagggtcc caccgccccc tgtttctgta accctgccaa ccccttcagt
      361 gaactttctt taattatcac tcagtttagt ttaatgtaaa cagtcccttc attaaactaa
      421 ttaccacttt ttgttttcca tttctttcct cccaggatgc taacagacac acctgcctcc
      481 catctaccta tcaaatcttg tgtggaactc tcaaactcaa gctgaaccct tgctgtgaga
      541 tttgctaatg gtaccagccc atcctgagct gtcctggaga gaattcaaca gcttcactaa
      601 aagagataga ccttcccttg ggtatgtagt gccacctgct ggacatccag agaaacagcg
      661 ccatctagct ttgaagtccc tgatgaagaa agcccacttg gaaggtagca atcactcttc
      721 aaatgaattg acttatgaag aaaaaaaaca cagcatgaaa tagagaactt taaatgagga
      781 ccagaaaaga aggaaaggag aatgttaaaa agtataagag tacaaggcga gtctccttct
      841 tccagagata aagaaatgga gatattcaga tttcttaaaa gggcccttcc tgtcctgcga
      901 catatgctgt attatgatta tgttgttcat ttcctcattt tcattacttc tgtcctcttg
      961 gttaagtagc ctgattctaa agacctaagg gctagaacta acactgacaa gatggaaaaa
     1021 aagctcttga tgtgcactta gctgctccaa atgaacatgc aaacaggatc tatggttcct
     1081 aaagcccact gatgagataa ttcatccaat atcaaaaatc ctttggcaat cttttgggta
     1141 gaattctgcc cttactatta ttatacaata ggcatcaccc tacagtcgtt tcatgtgtaa
     1201 ttcctaatct tcataaaatc tcttaatatt ttacaaatga ggaaacaaag gctcaaaaga
     1261 gttaaacaac ttcctcaagc ttcctagtac ataagtggta gagctgggct ttgaatccag
     1321 gtctctgtga tacaaaaaaa aaaacccaaa aactatattt ttcctatatg ttaagctgcc
     1381 tttcagaacc ctatctactt cagatcaaga gcttccaata tgcaaggcat agggctagat
     1441 gcattacctc atttagtcac agagtaatgc atttgagttc attatcatgg tcatttgata
     1501 ttgacattgc tagcaagatt ttaaaaccag taattaaaaa ggttttgagc aatggggaaa
     1561 ggattcaatc actggaaact agtaaggctt tgctatgagt aaatcatgcc tgcctatctt
     1621 cacttccttt taaatttggc caacagatga tagagaaatg gtggagttat ttagacaggc
     1681 atttgaaaaa gtctttctta tgatcctctt ttatgaacaa cctgcactac atagaataat
     1741 aatatcatgt aatccagtga gctccaccca aagagctttg actgatgtac tgacttctca
     1801 gatcagtctt ttttgattta ttaatttgct aggatgagga catggatgat atcaactcct
     1861 ctgatcctct gaagctgaga ggaatagctg atatatacga atgaacctag tccaagacat
     1921 tgtcaacaca ccgggatgat gagcagatcc taatgagatg aagtttaaca gaactgaaag
     1981 ggaggtcctc cactatgtcc caggaaaaag tgctcaaggg agacgaacaa atcatttata
```

```
     2041 tgaaaaggct agcaatggca gaaagaatga actacagtgt agtcagaatc tgaaggaggc
     2101 ggttcatggc cccacttctc cacttgttca caggggagct cctgctgggc tttcaagggg
     2161 actcacggga cacatgctaa agggagtttc tgtcctgtgg tatgtctcca tgtacctgat
     2221 caacaattaa aggaactggg aaggtttaac ctagagaaga ctttttggga ggtgggatgt
     2281 ctgtatccaa acttctgatg ggctacgttg atgatgagac atggaaaagt agttggtgta
     2341 agtgggctca gagagtagaa ctaagagtgg aagttagatg gggtagtgtc tggcttcata
     2401 tacggaggaa ctttctaaca gccgtagaaa aacaaaatga actatttcaa agaagtaagt
     2461 cccttactct gaaacttttg ttccaccaga ggcaggatga ctgaccattt gagagggatg
     2521 gtggaggga aattctggca ttcagtggag gcttgatcgt attactgcta aaacccccttc
     2581 taagccagag actctacaat actctgcaaa tgaaactgtg catcttctaa gtataagatg
     2641 cagggacata ctaaaagaag agaagaagaa agtgatgagt ccagtcagat gaaaaaagtg
     2701 ttgtgctggc tgagcgaaga tgacataacc ctctgcaagc ttgctggaca cagttacggt
     2761 ggttttggac tcggcttgca gacagatatg gtggccatca aacccccatt gatggacggt
     2821 gctgtaaagt cacattgctc actacttgga aaagccttca tgttttccaa aagagacaaa
     2881 cttctcaatc ccttcaatta tattattagg tgggctctta atggcagcca gcttggccct
     2941 tgcatattgc tgcacccttt tattccagac tttaagccct tcagtcactg aaacttgagc
     3001 cacttatgtt acaatcaagc cacaaagaga acaggggaaa atggtatgcc ttcttttact
     3061 ttgaagagat gacttcatca ccagtggtta tgtttctttt aagtctcaga atacatgata
     3121 tattacttct a
//
```

Query: aatatcatgt aat

PREDICTED: Manis pentadactyla uncharacterized LOC118917760 (LOC118917760), ncRNA
NCBI Reference Sequence: XR_005026893.1



Influenza A virus (A/environment/Anhui/2.22_YHZGS004-E/2019(mixed)) segment 3 polymerase PA (PA) and PA-X protein (PA-X) genes, complete cds

GenBank: MW109311.1

FASTA Graphics

Go to:

LOCUS MW109311 16 bp cRNA linear VRL **19-OCT-2020**

DEFINITION Influenza A virus (A/environment/Anhui/2.22_YHZGS004-E/2019(mixed)) segment 3 polymerase PA (PA) and PA-X protein (PA-X) genes, complete cds.

ACCESSION MW109311 REGION: 903..918

VERSION MW109311.1

KEYWORDS .

SOURCE Influenza A virus

ORGANISM Influenza A virus

Viruses; Riboviria; Orthornavirae; Negarnaviricota; Polyploviricotina; Insthoviricetes; Articulavirales; Orthomyxoviridae; Alphainfluenzavirus.

REFERENCE 1 (bases 1 to 16)

AUTHORS Bi,Y., Li,J., Li,S., Fu,G., Jin,T., Zhang,C., Yang,Y., Ma,Z., Tian,W., Li,J., Xiao,S., Li,L., Yin,R., Zhang,Y., Wang,L., Qin,Y., Yao,Z., Meng,F., Hu,D., Li,D., Wong,G., Liu,F., Lv,N., Wang,L., Fu,L., Yang,Y., Peng,Y., Ma,J., Sharshov,K., Shestopalov,A., Gulyaeva,M., Gao,G.F., Chen,J., Shi,Y., Liu,W.J., Chu,D., Huang,Y., Liu,Y., Liu,L., Liu,W., Chen,Q. and Shi,W.

TITLE Dominant subtype switch in avian influenza viruses during 2016-2019 in China

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 16)

AUTHORS Bi,Y., Li,J., Li,S., Fu,G., Jin,T., Zhang,C., Yang,Y., Ma,Z., Tian,W., Li,J., Xiao,S., Li,L., Yin,R., Zhang,Y., Wang,L., Qin,Y., Yao,Z., Meng,F., Hu,D., Li,D., Wong,G., Liu,F., Lv,N., Wang,L., Fu,L., Yang,Y., Peng,Y., Ma,J., Sharshov,K., Shestopalov,A., Gulyaeva,M., Gao,G.F., Chen,J., Shi,Y., Liu,W.J., Chu,D., Huang,Y., Liu,Y., Liu,L., Liu,W., Chen,Q. and Shi,W.

TITLE Direct Submission

JOURNAL Submitted (09-OCT-2020) Key Laboratory of Etiology and Epidemiology of Emerging Infectious Diseases in Universities of Shandong, Shandong First Medical University & Shandong Academy of Medical Sciences, Yingsheng East Road No. 2, Taian, Shandong 271000, China

COMMENT ##Assembly-Data-START##
Assembly Method :: SAMtools v. 1.4
Sequencing Technology :: Illumina
##Assembly-Data-END##

FEATURES Location/Qualifiers

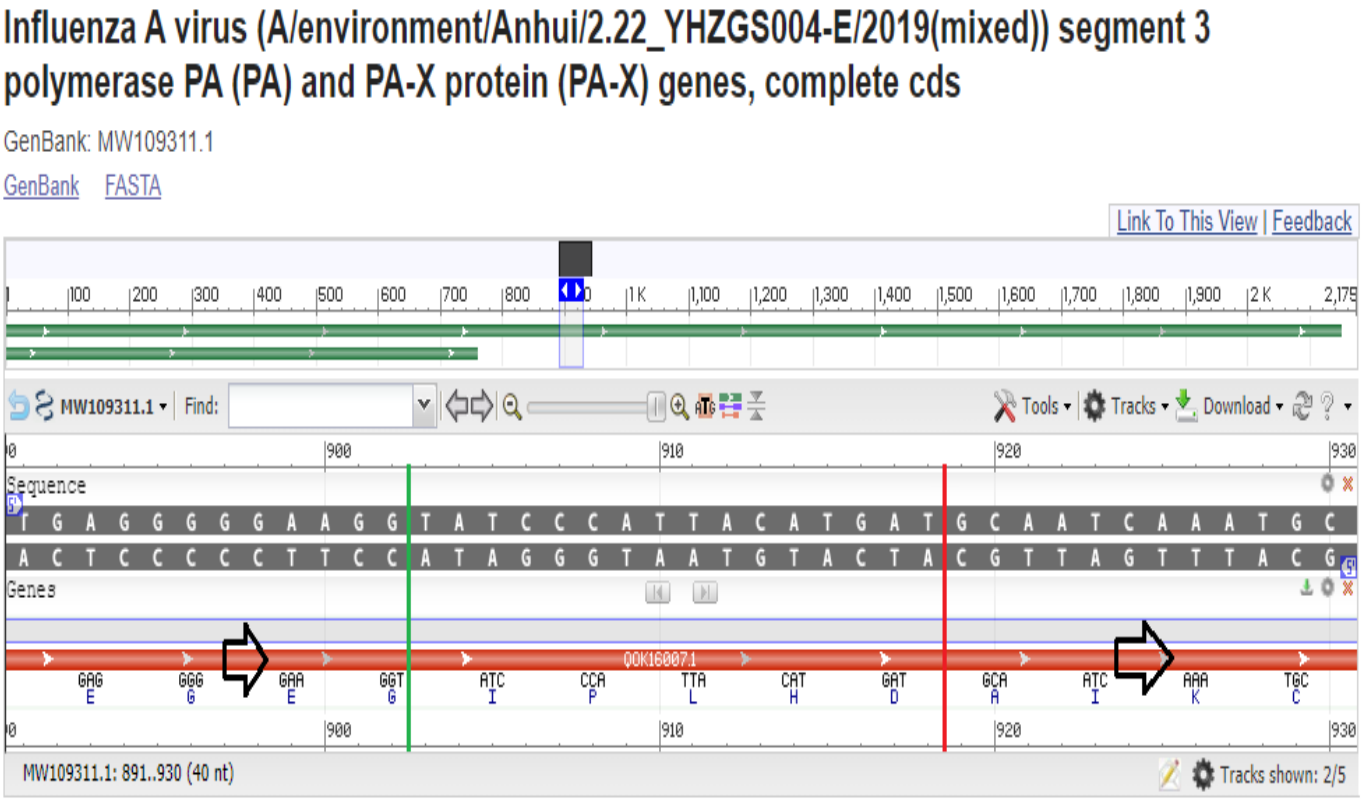
source 1..16

```
                     /organism="Influenza A virus"
                     /mol_type="viral cRNA"
                     /strain="A/environment/Anhui/2.22_YHZGS004-E/2019"
                     /serotype="mixed"
                     /isolate="2.22_YHZGS004-E"
                     /isolation_source="swab"
                     /db_xref="taxon:11320"
                     /segment="3"
                     /country="China: Anhui"
                     /collection_date="22-Feb-2019"
     gene            <1..>16
                     /gene="PA"
     CDS             <1..>16
                     /gene="PA"
                     /codon_start=2
                     /product="polymerase PA"
                     /protein_id="QOK16007.1"
                     /translation="MEDFVRQCFNPMIVELAEKTMKEYGESPKIETNKFAAICTHLEV
                     CFMYSDFHFINERGESIIAESGDPNALLKHRFEIIEGRDRTMAWTVVNSICNTTGVDK
                     PKFLPDLYDYKENRFIEIGVTRREVHIYYLEKANKIKSERTHIHIFSFTGEEMATKAD
                     YTLDEESRARIKTRLFTIRQEMASRGLWDSFRQSERGEXTIEERFEITGTMRRLADQS
                     LPPNFSSLENFRAYVDGFEPNGCIEGKLSQMSKEVNARIEPFXKTTPRPLRLPDGPPC
                     SQRSKFLLMDALKLSIEDPSHEGEGIPLHDAIKCMKTFFGWKEPNIVKPHDKGINPNY
                     LLAWKQVLAELQDIENEDKIPKTKNMKKTSQLKWALGENMAPEKVDFEDCKDVSDLKQ
                     YDSDEPEPRSLSSWVQSEFNKACELTDSSWIELDEIGEDVAPIEHIASMRRNYFTAEV
                     SHCRATEYIMKGVYINTALLNASCAAMDDFQLIPMISKCRTREGRRKTNLYGFIIKGR
                     SHLRNDTDVVNFVSMEFSLTDPRLEPHKWVKYCVLEIGDMLLRTAIGQVSRPMFLYVR
                     TNGTSKIKMKWGMEMRRCLLQSLQQIESMIEAESSVKEKDMTKEFFETKSETWPIGES
                     PKGVEESSIGKVCRTLLAKSVFNSLYASPQLEGFSAESRKLLLIVQALRDNLEPGTFD
                     LGGLYDAIEECLINDPWVLLNASWFNSFLTHALK"
ORIGIN
        1 tatcccatta catgat
//
```

Influenza A virus (A/environment/Anhui/2.22_YHZGS004-E/2019(mixed)) segment 3 polymerase PA (PA) and PA-X protein (PA-X) genes, complete cds
Sequence ID: MW109311.1 Length: 2175 Number of Matches: 1
Range 1: 903 to 918 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
32.2 bits(16)	144	16/16(100%)	0/16(0%)	Plus/Plus

```
CDS: Putative 1        1      I  P  L  H  D
Query                  1      TATCCCATTACATGAT   16
                              ||||||||||||||||
Sbjct                  903    TATCCCATTACATGAT   918
CDS:polymerase PA [I   302    I  P  L  H  D
```



Polymerase PA (PA) - Polymerase acidic protein

Polymerase acidic protein (or **RNA-directed RNA polymerase subunit P2**) of **PA gene** plays an essential role in *viral RNA transcription* and *replication* by forming the *heterotrimeric polymerase complex together with PB1 and PB2 subunits*. The *complex transcribes viral mRNAs* by using a unique mechanism called **cap-snatching**. It consists in the hijacking and cleavage of host capped pre-mRNAs. These short capped RNAs are then used as **primers for viral mRNAs**. The **PB2 subunit** is responsible for the binding of the 5' cap of cellular pre-mRNAs which are subsequently cleaved after 10-13 nucleotides by the **PA** subunit that carries the endonuclease activity (https://www.uniprot.org/uniprot/P03433; https://pubmed.ncbi.nlm.nih.gov/23847103/).

Influenza A virus (A/environment/Anhui/2.22_YHZGS003-E/2019(mixed)) segment 3 polymerase PA (PA) and PA-X protein (PA-X) genes, complete cds
Sequence ID: MW109303.1 Length: 2174 Number of Matches: 1
Range 1: 903 to 918 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
32.2 bits(16)	144	16/16(100%)	0/16(0%)	Plus/Plus

```
CDS: Putative 1       1      I  P  L  H  D
Query                 1    TATCCCATTACATGAT  16
                           ||||||||||||||||
Sbjct                 903  TATCCCATTACATGAT  918
CDS:polymerase PA [I  302    I  P  L  H  D
```

Influenza A virus (A/environment/Anhui/2.22_YHZGS001-E/2019(mixed)) segment 3 polymerase PA (PA) and PA-X protein (PA-X) genes, complete cds
Sequence ID: MW109287.1 Length: 2174 Number of Matches: 1
Range 1: 903 to 918 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
32.2 bits(16)	144	16/16(100%)	0/16(0%)	Plus/Plus

```
CDS: Putative 1       1      I  P  L  H  D
Query                 1    TATCCCATTACATGAT  16
                           ||||||||||||||||
Sbjct                 903  TATCCCATTACATGAT  918
CDS:polymerase PA [I  302    I  P  L  H  D
```

Influenza A virus (A/chicken/Anhui/2.22_YHZGS018-C/2019(mixed)) segment 3 polymerase PA (PA) and PA-X protein (PA-X) genes, complete cds
Sequence ID: MW102519.1 Length: 2175 Number of Matches: 1
Range 1: 903 to 918 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
32.2 bits(16)	144	16/16(100%)	0/16(0%)	Plus/Plus

```
CDS: Putative 1       1      I  P  L  H  D
Query                 1    TATCCCATTACATGAT  16
                           ||||||||||||||||
Sbjct                 903  TATCCCATTACATGAT  918
CDS:polymerase PA [I  302    I  P  L  H  D
```

LR862376.1
Pipistrellus pipistrellus genome assembly, chromosome: 20

LR862373.1
Pipistrellus pipistrellus genome assembly, chromosome: 17

LR862372.1
Pipistrellus pipistrellus genome assembly, chromosome: 16

LR862367.1
Pipistrellus pipistrellus genome assembly, chromosome: 11

LR862363.1
Pipistrellus pipistrellus genome assembly, chromosome: 7

LR862362.1
Pipistrellus pipistrellus genome assembly, chromosome: 6

10.4 THE PRINCIPAL "PRODUCTS" IDENTIFIED THROUGH BLAST RESEARCHES CARRIED OUT ON SEQUENCE N ° 3/1 (Hitlist size: 5000)

In addition to the "**products**" of the **Bacilli Paranthracis, Anthracis, and Cereus** sequences identified through previous the BLAST research, two other "products" were considered very interesting for the purposes of this work: "**Methyltransferase like 25**" and "**Polymerase PA (PA) - Protein Polymerase Acid**".

In a recent study (https://pubmed.ncbi.nlm.nih.gov/33685335/), the authors aim identify a potential ***novel inhibitor for SARS COV-2 2'-O-methyltransferase (nsp16)*** which is one of the most attractive targets in the virus life cycle, responsible for the *viral RNA protection* via a cap formation process. The outputs revealed the ***compound 11*** as the ***best potential nsp16 inhibitor***, as it displayed a *better stability and average binding free energy for the ligand-enzyme complex* compared to *Sinefungin*.
In another study of silico identification and validation of ***natural antiviral compounds*** as ***potential inhibitors of SARS-CoV-2 methyltransferase*** (https://pubmed.ncbi.nlm.nih.gov/33583328/), the authors consider **botanical** an important resource for several **efficacious antiviral agents**, **natural compounds** gaining significant attention for **COVID-19 treatment**. The authors found four *natural antiviral compounds* **Amentoflavone**, **Baicalin**, **Daidzin** and **Luteoloside** as ***strong inhibitors of methyltranferase of SARS-CoV-2***. Their analysis suggests that there is a very high probability that these compounds may have a good potential to *inhibit Methyltransferase (MTase) of SARS-CoV-2* and to be used in the treatment of COVID-19.

In an in-depth study (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7217786/), some researchers discuss of the structures, transmission characteristics, clinical symptoms, diagnosis, pathological changes, treatment and prevention of **two kinds of viruses**, **CoVs** (**SARS** e **MERS**) and **IAVs** (**Influenza A Viruses**). "The respiratory disease that has caused COVID-19 has similar clinical symptoms with other **two CoV diseases**, ***severe acute respiratory syndrome*** and ***Middle East respiratory syndrome*** (**SARS** and **MERS**), of which causative viruses are **SARS-CoV** and **MERS-CoV**, respectively. These three CoVs resulting diseases also share many clinical symptoms with other respiratory diseases caused by **IAVs**". Since both **CoVs** and **IAVs** are general pathogens responsible for seasonal cold, during the changing of seasons, clinicians and public heath may have to distinguish **COVID-19 pneumonia** from **other kinds of viral pneumonia** (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7217786/). In this regard, suitable pharmacotherapy is also essential. "Currently, for ***clinical treatment of influenza infection***, there are **only three** *classes of FDA (Food and Drug Administration)-approved antiviral drugs available*, including ***M2 ion channel inhibitors*** (amantadine and rimantadine), ***neuraminidase inhibitors*** (oseltamivir, zanamivir, peramivir and laninamivir), and the most recently ***RNA polymerase inhibitors*** (*balxavir marboxil*)".

"**Balxavir marboxil**, with a brand name of *Xofloza*, can target to **viral polymerase acidic protein (PA)**, *block its endonuclease function, leading to the inhibition of virus mRNA transcription and effectively* ***prevent influenza A virus infection***. The *antiviral spectrum of Xofloza* includes **seasonal influenza strains** and **influenza strains resistant to** "***oseltamivir***" [a ***neuraminidase inhibitors*** (NAI)], significantly reducing the viral load compared with *oseltamivir*. Also **T-705**, a *new type of RNA polymerase inhibitor*, can inhibit a wide range of viruses by directly entering into the viral RNA chain or by binding directly to the viral RNA polymerase domain to block viral RNA chain replication and transcription" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7217786/).

CHAPTER XI

BLAST RESEARCH AND DISCUSSION OF THE RESULTS CARRIED OUT ON ALIGNMENTS OF SEQUENCE N ° 8/1

GOALS OF CHAPTER XI

This chapter contains the results of BLAST research carried out on **Sequence n° 8/1**. This sequence is one of many generated by the **Original Sequence**, the sequence "**GTGAAATGGTCATGTGTGGCGG**" of the RdRp Gene and from which it differs by **90,90%** of the bases. The results of the significant alignments of the BLAST research (**Hitlist size 100**) involved two species of bats (*Rhinolophus ferrumequinum* and *Pipistrellus pipistrellus*) and covered several **common "organisms"** (distinguishable in different colours in the following pages) to the **"organisms"** identified by BLAST research carried out on sequences to which **a base of the original sequence has been replaced** (see Chapters V and VI).
The common "organisms" identified are: **Chanos chanos**, **Ipomoea triloba**, **Ipomoea Trifida**, **Coregonus sp. 'balchen'**, and **Salmo trutta**.

WARNING: **The enormous amount of work that has been necessary until now to analyze and select all the significant sequences of the different BLAST researches shown in the previous chapters makes it difficult to proceed quickly with such meticulousness. The need to publish as soon as possible the results of this extensive document, forces me, from this page on, to leave scientists concerned with the arduous task (which I absolutely recommend to perform) to deepen the analysis and selection of the sequences considered relevant according to the indications proposed in the previous chapters.**

For the reasons set out above, BLAST in-depth research will be limited in scope.

Extending the BLAST research (**Hitlist size 250**) another sequence of the ***Pipistrellus pipistrellus*** and a considerable number of sequences of the ***Bacilli Paranthracis*** and ***Anthracis*** were identified.

SEQUENCE N° 8/1

(T=1; C=0; G=3; A=2)

11.1 SEQUENCE N° 8/1 (T=1; C=0; G=3; A=2) – (Hitlist size: 100)

Original Sequence	G	T	G	A	A	A	T	G	G	T	C	A	T	G	T	G	T	G	G	C	G	G
Trend n° 8 (T=1; C=0; G=3; A=2)	3	1	3	2	2	2	1	3	3	1	0	2	1	3	1	3	1	3	3	0	3	3
Trend n° 8 (I i D d)	-	D	I	D	d	d	D	I	i	D	D	I	D	I	D	I	D	I	i	D	I	i
Trend n° 8/1	2	0	2	1	1	1	0	3	3	2	1	3	0	2	0	2	0	2	2	1	2	2
Sequence n° 8/1 (90,90%)	A	C	A	T	T	T	C	G	G	A	T	G	C	A	C	A	C	A	A	T	A	A

Query: **ACATTTCGGATGCACACAATAA**

Search Parameters	
Program	Blastn
Word size	7
Expect value	1000
Hitlist size	100
Match/Mismatch scores	1,-3
Gapcosts	5,2
Filter string	F
Genetic Code	1

100 sequences selected:

Accession Seq	Description	Comomm Name	Max Score	Total Score	Query C.	E value	Per. Id.	Acc. Len
LR884471.1	Acanthopagrus latus isolate v.2019 genome assembly, chromosome: 12	yellowfin seabream	40.1	40.1	90%	0.57	100.00%	26679254
XM_030786843.1 1/2 2/2 3/1 3/3 3/4	PREDICTED: Chanos chanos V-type proton ATPase subunit B, brain isoform-like (LOC115822892), mRNA	milkfish	38.2	38.2	86%	2.2	100.00%	2465
LR697115.1 1/2 2/2 3/1 3/3 3/4	Chanos chanos genome assembly, chromosome: 10	milkfish	38.2	38.2	86%	2.2	100.00%	40347274
LR812078.1	Danio rerio genome assembly, chromosome: 16	zebrafish	36.2	36.2	81%	8.9	100.00%	55350661
XM_017289635.2	PREDICTED: Drosophila miranda zinc finger protein 622 (LOC108157528), mRNA	Drosophila miranda	36.2	36.2	81%	8.9	100.00%	1577
XM_001355143.4	PREDICTED: Drosophila pseudoobscura zinc finger protein 622 (LOC4815046), mRNA	Drosophila pseudoobscura	36.2	36.2	81%	8.9	100.00%	1573
XM_031276121.1	PREDICTED: Ipomoea triloba FAD synthetase 1, chloroplastic-like (LOC116033344), partial mRNA	trilobed morning glory	36.2	36.2	81%	8.9	100.00%	1153
XM_002022755.2	PREDICTED: Drosophila persimilis zinc finger protein 622 (LOC6597574), mRNA	Drosophila persimilis	36.2	36.2	81%	8.9	100.00%	1532
CP025660.1	Ipomoea triloba cultivar NCNSP0323 chromosome 1	trilobed morning glory	36.2	36.2	81%	8.9	100.00%	38093535
CP032234.1	Anopheles stephensi strain SDA-500 chromosome 3R	Asian malaria mosquito	36.2	66.4	81%	8.9	100.00%	51436555
CP032595.1	Lateolabrax maculatus linkage group 2 sequence	spotted sea bass	36.2	36.2	81%	8.9	100.00%	29991467
CP027278.1	Lateolabrax maculatus chromosome Lm17	spotted sea bass	36.2	36.2	81%	8.9	100.00%	22884195
CP026254.1	Scophthalmus maximus chromosome 12	turbot	36.2	36.2	100%	8.9	95.45%	24811384
XM_019320759.1	PREDICTED: Ipomoea nil FAD synthetase 2, chloroplastic-like (LOC109171689), transcript variant X3, mRNA	Japanese morning glory	36.2	36.2	81%	8.9	100.00%	1421
XM_019320758.1	PREDICTED: Ipomoea nil FAD synthetase 2, chloroplastic-like (LOC109171689), transcript variant X2, mRNA	Japanese morning glory	36.2	36.2	81%	8.9	100.00%	1401
XM_019320757.1	PREDICTED: Ipomoea nil FAD synthetase 2, chloroplastic-like (LOC109171689), transcript variant X1, mRNA	Japanese morning glory	36.2	36.2	81%	8.9	100.00%	1405
XM_013256488.1	PREDICTED: Stomoxys calcitrans	stable fly	36.2	36.2	81%	8.9	100.00%	1230

	homocysteine S-methyltransferase-like (LOC106090333), mRNA							
XM_034397081.1	PREDICTED: Thrips palmi uncharacterized LOC117652271 (LOC117652271), transcript variant X4, mRNA	Thrips palmi	36.2	36.2	81%	8.9	100.00%	910
XM_034397072.1	PREDICTED: Thrips palmi uncharacterized LOC117652271 (LOC117652271), transcript variant X3, mRNA	Thrips palmi	36.2	36.2	81%	8.9	100.00%	910
XM_034397054.1	PREDICTED: Thrips palmi uncharacterized LOC117652271 (LOC117652271), transcript variant X1, mRNA	Thrips palmi	36.2	36.2	81%	8.9	100.00%	931
LR738627.1	Sciurus vulgaris genome assembly, chromosome: 16	Eurasian red squirrel	36.2	66.4	100%	8.9	100.00%	79696672
LR738607.1	Sciurus carolinensis genome assembly, chromosome: 17	gray squirrel	36.2	66.4	100%	8.9	100.00%	66045118
CP025644.1	**Ipomoea trifida cultivar NCNSP0306 chromosome 1**	Ipomoea trifida	36.2	36.2	81%	8.9	100.00%	32247286
LT594792.1	Theobroma cacao genome assembly, chromosome: V	cacao	36.2	66.4	90%	8.9	100.00%	39371879
LR812514.1	Danio aesculapii genome assembly, chromosome: 18	Danio aesculapii	36.2	36.2	81%	8.9	100.00%	56436977
JN949980.1	Mus musculus targeted non-conditional, lacZ-tagged mutant allele Ggt6:tm1e(KOMP)Wtsi; transgenic	house mouse	36.2	36.2	81%	8.9	100.00%	39835
JN949979.1	Mus musculus targeted KO-first, conditional ready, lacZ-tagged mutant allele Ggt6:tm1a(KOMP)Wtsi; transgenic	house mouse	36.2	36.2	81%	8.9	100.00%	39885
JN949026.1	Mus musculus targeted non-conditional, lacZ-tagged mutant allele Mybbp1a:tm1e(EUCOMM)Hmgu; transgenic	house mouse	36.2	36.2	81%	8.9	100.00%	38333
JN949025.1	Mus musculus targeted KO-first, conditional ready, lacZ-tagged mutant allele Mybbp1a:tm1a(EUCOMM)Hmgu; transgenic	house mouse	36.2	36.2	81%	8.9	100.00%	38404
AF345929.1	Mus musculus Myb-binding protein p160 1a (Mybbp1a) gene, promoter region and partial cds	house mouse	36.2	36.2	81%	8.9	100.00%	549
AL662812.12	Mouse DNA sequence from clone RP23-48A2 on chromosome 11, complete sequence	house mouse	36.2	36.2	81%	8.9	100.00%	258521
LR792833.1	Digitaria exilis annotation	Digitaria exilis	34.2	34.2	77%	35	100.00%	29094342
LR761618.1	Digitaria exilis genome assembly, chromosome: 7B	Digitaria exilis	34.2	34.2	77%	35	100.00%	29094342
LR792548.1	Pseudochaenichthys georgianus genome assembly, chromosome: 3	South Georgia icefish	34.2	34.2	95%	35	95.24%	53425585
XR_004425352.1	**PREDICTED: Rhinolophus ferrumequinum serine/arginine repetitive matrix protein 1-like (LOC117036333), transcript variant X5, misc_RNA**	greater horseshoe bat	34.2	34.2	77%	35	100.00%	2643
XM_033131157.1	**PREDICTED: Rhinolophus ferrumequinum serine/arginine repetitive matrix protein 1-like (LOC117036333), transcript variant X4, mRNA**	greater horseshoe bat	34.2	34.2	77%	35	100.00%	2775
XM_033131156.1	**PREDICTED: Rhinolophus ferrumequinum serine/arginine repetitive matrix protein 1-like (LOC117036333), transcript variant X3, mRNA**	greater horseshoe bat	34.2	34.2	77%	35	100.00%	2844
XM_033131154.1	**PREDICTED: Rhinolophus ferrumequinum serine/arginine repetitive matrix protein 1-like (LOC117036333), transcript variant X1, mRNA**	greater horseshoe bat	34.2	34.2	77%	35	100.00%	2758
LR778290.1 **1/2 2/2 3/3 9/1**	Coregonus sp. 'balchen' genome assembly, chromosome: 38	Coregonus sp. 'balchen'	34.2	132	77%	35	100.00%	36774138
LR778289.1 **1/2 2/2 3/3 9/1**	Coregonus sp. 'balchen' genome assembly, chromosome: 37	Coregonus sp. 'balchen'	34.2	96.6	95%	35	100.00%	43663377
LR584434.1 **1/2 2/2 3/3 3/4 3/6**	Salmo trutta genome assembly, chromosome: 34	river trout	34.2	34.2	77%	35	100.00%	42900288
LR584426.1 **1/2 2/2 3/3 3/4 3/6**	Salmo trutta genome assembly, chromosome: 17	river trout	34.2	68.4	95%	35	95.24%	59764774

LR584421.1 1/2 2/2 3/3 3/4 3/6	Salmo trutta genome assembly, chromosome: 23	river trout	34.2	34.2	95%	35	95.24%	51488243
LR584414.1 1/2 2/2 3/3 3/4 3/6	Salmo trutta genome assembly, chromosome: 33	river trout	34.2	34.2	77%	35	100.00%	44893998
XM_026300864.1	PREDICTED: Mastacembelus armatus endoplasmic reticulum membrane sensor NFE2L1-like (LOC113126748), mRNA	zig-zag eel	34.2	34.2	77%	35	100.00%	3081
XM_002071777.3	PREDICTED: Drosophila willistoni chromosome-associated kinesin KIF4A (LOC6648886), mRNA	Drosophila willistoni	34.2	34.2	77%	35	100.00%	4044
CP023332.1	Drosophila melanogaster strain sitter (fors) chromosome 2R	fruit fly	34.2	34.2	77%	35	100.00%	21146707
XM_019095370.1	PREDICTED: Cyprinus carpio uncharacterized LOC109080287 (LOC109080287), transcript variant X2, mRNA	common carp	34.2	34.2	95%	35	95.24%	1325
XM_019095369.1	PREDICTED: Cyprinus carpio uncharacterized LOC109080287 (LOC109080287), transcript variant X1, mRNA	common carp	34.2	34.2	95%	35	95.24%	1325
XM_017162352.1	PREDICTED: Drosophila kikkawai uncharacterized LOC108071571 (LOC108071571), mRNA	Drosophila kikkawai	34.2	34.2	77%	35	100.00%	1341
LN029746.1	Spirometra erinaceieuropaei genome assembly S_erinaceieuropaei, scaffold SPER_scaffold0029177	Spirometra erinaceieuropaei	34.2	34.2	77%	35	100.00%	7894
AE013599.5	Drosophila melanogaster chromosome 2R	fruit fly	34.2	34.2	77%	35	100.00%	25286936
XM_007136828.1	Phaseolus vulgaris hypothetical protein (PHAVU_009G082600g) mRNA, complete cds	Phaseolus vulgaris	34.2	34.2	77%	35	100.00%	5319
OA888277.1	Notodromas monacha	Notodromas monacha	34.2	34.2	77%	35	100.00%	13315
XM_037692747.1	PREDICTED: Nematolebias whitei nuclear factor, erythroid 2-like 3 (nfe2l3), mRNA	Rio pearlfish	34.2	34.2	77%	35	100.00%	2051
LR812054.1	Danio rerio genome assembly, chromosome: 17	zebrafish	34.2	34.2	77%	35	100.00%	54645068
LR812560.1	Danio rerio strain Cooch Behar (CB) genome assembly, chromosome: 17	zebrafish	34.2	34.2	77%	35	100.00%	50025757
CP002563.1	Carnobacterium sp. 17-4, complete genome	Carnobacterium sp. 17-4	34.2	34.2	77%	35	100.00%	2635294
AC146941.8	Medicago truncatula clone mth2-133o10, complete sequence	barrel medic	34.2	34.2	77%	35	100.00%	134901
LR812081.1	Danio rerio genome assembly, chromosome: 19	zebrafish	34.2	34.2	77%	35	100.00%	49536214
LR812071.1	Danio rerio genome assembly, chromosome: 9	zebrafish	34.2	64.4	77%	35	100.00%	56856035
LR778269.1 1/2 2/2 3/3 9/1	Coregonus sp. 'balchen' genome assembly, chromosome: 17	Coregonus sp. 'balchen'	34.2	124	95%	35	100.00%	54216998
LR778267.1 1/2 2/2 3/3 9/1	Coregonus sp. 'balchen' genome assembly, chromosome: 15	Coregonus sp. 'balchen'	34.2	94.6	95%	35	95.24%	55641933
LR778260.1 1/2 2/2 3/3 9/1	Coregonus sp. 'balchen' genome assembly, chromosome: 8	Coregonus sp. 'balchen'	34.2	124	86%	35	100.00%	68138733
XM_029729683.1 1/2 2/2 3/3 3/4 3/6	PREDICTED: Salmo trutta coiled-coil domain-containing protein 85C-A-like (LOC115172326), transcript variant X3, mRNA	river trout	34.2	34.2	77%	35	100.00%	3497
XM_029729682.1 1/2 2/2 3/3 3/4 3/6	PREDICTED: Salmo trutta coiled-coil domain-containing protein 85C-A-like (LOC115172326), transcript variant X2, mRNA	river trout	34.2	34.2	77%	35	100.00%	3601
XM_029729681.1 1/2 2/2 3/3 3/4 3/6	PREDICTED: Salmo trutta coiled-coil domain-containing protein 85C-A-like (LOC115172326), transcript variant X1, mRNA	river trout	34.2	34.2	77%	35	100.00%	3604
LR584435.1 1/2 2/2 3/3 3/4 3/6	Salmo trutta genome assembly, chromosome: 18	river trout	34.2	94.6	95%	35	95.24%	59142494
LR584433.1 1/2 2/2 3/3 3/4 3/6	Salmo trutta genome assembly, chromosome: 5	river trout	34.2	34.2	95%	35	95.24%	67763595
LR584423.1 1/2 2/2 3/3 3/4 3/6	Salmo trutta genome assembly, chromosome: 31	river trout	34.2	34.2	77%	35	100.00%	45589842

LR535843.1	Mastacembelus armatus genome assembly, chromosome: 11	zig-zag eel	34.2	34.2	77%	35	100.00%	22941789
LR131966.1	Parambassis ranga genome assembly, chromosome: 16	Indian glassy fish	34.2	34.2	77%	35	100.00%	22465611
CP025673.1	**Ipomoea triloba cultivar NCNSP0323 chromosome 14**	trilobed morning glory	34.2	34.2	95%	35	95.24%	23651743
LT594791.1	Theobroma cacao genome assembly, chromosome: IV	cacao	34.2	34.2	77%	35	100.00%	31929104
LT594790.1	Theobroma cacao genome assembly, chromosome: III	cacao	34.2	34.2	77%	35	100.00%	36364294
CP014500.1	Sugiyamaella lignohabitans strain CBS 10342 chromosome C, complete sequence	Sugiyamaella lignohabitans	34.2	34.2	77%	35	100.00%	2773601
AP015036.1	Vigna angularis var. angularis DNA, chromosome 3, almost complete sequence, cultivar: Shumari	Vigna angularis var. angularis	34.2	34.2	95%	35	95.24%	43462759
XM_014194909.1	PREDICTED: Salmo salar coiled-coil domain-containing protein 85C-A-like (LOC106602325), transcript variant X3, mRNA	Atlantic salmon	34.2	34.2	77%	35	100.00%	5155
XM_014194899.1	PREDICTED: Salmo salar coiled-coil domain-containing protein 85C-A-like (LOC106602325), transcript variant X2, mRNA	Atlantic salmon	34.2	34.2	77%	35	100.00%	5260
XM_014194890.1	PREDICTED: Salmo salar coiled-coil domain-containing protein 85C-A-like (LOC106602325), transcript variant X1, mRNA	Atlantic salmon	34.2	34.2	77%	35	100.00%	5261
CP012024.1	Bacillus smithii strain DSM 4216, complete genome	Bacillus smithii	34.2	34.2	77%	35	100.00%	3368778
LK066269.1	Apteryx australis mantelli genome assembly AptMant0, scaffold scaffold1697	Apteryx australis mantelli	34.2	102	77%	35	100.00%	483239
LN013793.1	Spirometra erinaceieuropaei genome assembly S_erinaceieuropaei, scaffold SPER_contig0000183	Spirometra erinaceieuropaei	34.2	34.2	77%	35	100.00%	11008
LN595178.1	Cyprinus carpio genome assembly common carp genome, scaffold 000001327	common carp	34.2	34.2	95%	35	95.24%	212377
LR899427.1	Corylus avellana genome assembly, chromosome: ca5	Corylus avellana	34.2	34.2	77%	35	100.00%	36653616
LR862152.1	Ananas comosus var. bracteatus genome assembly, chromosome: 24	Ananas comosus var. bracteatus	34.2	34.2	77%	35	100.00%	21245074
LR812056.1	Danio rerio genome assembly, chromosome: 19	zebrafish	34.2	34.2	77%	35	100.00%	49847318
LR812046.1	Danio rerio genome assembly, chromosome: 9	zebrafish	34.2	34.2	77%	35	100.00%	57070137
HE601631.2	Schistosoma mansoni strain Puerto Rico genome assembly, chromosome: ZW	Schistosoma mansoni	34.2	96.6	86%	35	100.00%	88385488
LR812612.1	Danio rerio genome assembly, chromosome: 19	zebrafish	34.2	34.2	77%	35	100.00%	47671397
LR812602.1	Danio rerio genome assembly, chromosome: 9	zebrafish	34.2	34.2	77%	35	100.00%	56263738
LR812587.1	Danio rerio strain Nadia (NA) genome assembly, chromosome: 19	zebrafish	34.2	34.2	77%	35	100.00%	48376370
LR812585.1	Danio rerio strain Nadia (NA) genome assembly, chromosome: 17	zebrafish	34.2	34.2	77%	35	100.00%	54799276
LR812577.1	Danio rerio strain Nadia (NA) genome assembly, chromosome: 9	zebrafish	34.2	64.4	77%	35	100.00%	57525603
LR812552.1	Danio rerio strain Cooch Behar (CB) genome assembly, chromosome: 9	zebrafish	34.2	64.4	77%	35	100.00%	54069560
AC136507.19	Medicago truncatula clone mth2-24o19, complete sequence	barrel medic	34.2	34.2	77%	35	100.00%	131025
CR391919.5	Zebrafish DNA sequence from clone DKEY-174K18 in linkage group 9, complete sequence	zebrafish	34.2	34.2	77%	35	100.00%	179764
BX510640.4	Zebrafish DNA sequence from clone RP71-57J15 in linkage group 19, complete sequence	zebrafish	34.2	34.2	77%	35	100.00%	139917
AC006247.12	Drosophila melanogaster, chromosome 2R, region 49E6-49F8, BAC clone BACR48I10, complete sequence	fruit fly	34.2	34.2	77%	35	100.00%	163162
CP046680.1 **1/2 2/2 3/2 3/3 3/5 6/1 9/1 10/1**	Solanum tuberosum cultivar P8 chromosome 2	potato	30.2	60.5	68%	547	100.00%	48611948

ALIGNMENT VIEW FLAT QUERY-ANCHORED WITH LETTERS FOR IDENTITIES

Query range 1: 1 to 22

Query	1	ACATTTCGGATGCACACAATAA	22
LR884471.1	5872729	ACATTTCGGATGCACACAAT	5872748
XM_030786843.1	1749	ACATTTCGGATGCACACAA	1731
LR697115.1	19324111	ACATTTCGGATGCACACAA	19324093
LR812078.1	6416688	CATTTCGGATGCACACAA	6416671
XM_017289635.2	165	ACATTTCGGATGCACACA	148
XM_001355143.4	171	ACATTTCGGATGCACACA	154
XM_031276121.1	386	ATTTCGGATGCACACAAT	369
XM_002022755.2	130	ACATTTCGGATGCACACA	113
CP025660.1	13048241	ATTTCGGATGCACACAAT	13048224
CP032234.1	29284203	ACATTTCGGATGCACACA	29284186
CP032234.1	30390915	TTTCGGATGCACACA	30390901
CP032595.1	11460181	CATTTCGGATGCACACAA	11460164
CP027278.1	9140189	CATTTCGGATGCACACAA	9140172
CP026254.1	7772702	ACATTGCGGATGCACACAATAA	7772681
XM_019320759.1	469	ATTTCGGATGCACACAAT	452
XM_019320758.1	449	ATTTCGGATGCACACAAT	432
XM_019320757.1	453	ATTTCGGATGCACACAAT	436
XM_013256488.1	851	ATTTCGGATGCACACAAT	834
XM_034397081.1	826	ACATTTCGGATGCACACA	843
XM_034397072.1	826	ACATTTCGGATGCACACA	843
XM_034397054.1	847	ACATTTCGGATGCACACA	864
LR738627.1	9776344	ACATTTCGGATGCACACA	9776361
LR738627.1	43382214	GGATGCACACAATAA	43382200
LR738607.1	12873654	ACATTTCGGATGCACACA	12873671
LR738607.1	65672302	GGATGCACACAATAA	65672316
CP025644.1	8402783	ATTTCGGATGCACACAAT	8402800
LT594792.1	27997305	ATTTCGGATGCACACAAT	27997322
LT594792.1	7456216	GGATGCACACAATAA	7456202
LR812514.1	16160607	ACATTTCGGATGCACACA	16160624
JN949980.1	27127	TTCGGATGCACACAATAA	27144
JN949979.1	27177	TTCGGATGCACACAATAA	27194
JN949026.1	11815	TTCGGATGCACACAATAA	11832
JN949025.1	11815	TTCGGATGCACACAATAA	11832
AF345929.1	185	TTCGGATGCACACAATAA	202
AL662812.12	58729	TTCGGATGCACACAATAA	58746
LR792833.1	3481679	TTCGGATGCACACAATA	3481663
LR761618.1	3481679	TTCGGATGCACACAATA	3481663
LR792548.1	27700846	CATTTCGGAAGCACACAATAA	27700826
XR_004425352.1	869	CATTTCGGATGCACACA	853
XM_033131157.1	886	CATTTCGGATGCACACA	870
XM_033131156.1	955	CATTTCGGATGCACACA	939
XM_033131154.1	869	CATTTCGGATGCACACA	853
LR778290.1	30972461	CATTTCGGATGCACACA	30972445
LR778290.1	31640047	CATTTCGGATGCACACA	31640031
LR778290.1	31724491	CATTTCGGATGCACACA	31724507
LR778290.1	25939133	CATTTCGGATGCACA	25939147
LR778289.1	37891225	CATTTGGGATGCACACAATA	37891244
LR778289.1	38138938	CATTTCGGATGCACACA	38138922
LR778289.1	1178049	GGATGCACACAATAA	1178063
LR584434.1	4737613	CATTTCGGATGCACACA	4737597
LR584426.1	40716693	CATTTCGGATGCACACTATAA	40716673
LR584426.1	44568833	CATTTCGGATGCACACA	44568849
LR584421.1	6926956	CATTTGGGATGCACACAATAA	6926936

LR584414.1	8526326	CATTTCGGATGCACACA	8526310
XM_026300864.1	583	ACATTTCGGATGCACAC	567
XM_002071777.3	408	TTTCGGATGCACACAAT	392
CP023332.1	8950321	TTTCGGATGCACACAAT	8950305
XM_019095370.1	861	ACATTTCGGATGCTCACAATA	841
XM_019095369.1	861	ACATTTCGGATGCTCACAATA	841
XM_017162352.1	914	ACATTTCGGATGCACAC	898
LN029746.1	3786	ACATTTCGGATGCACAC	3770
AE013599.5	13062817	TTTCGGATGCACACAAT	13062801
XM_007136828.1	1139	TTTCGGATGCACACAAT	1123
OA888277.1	4987	ATTTCGGATGCACACAA	4971
XM_037692747.1	308	ACATTTCGGATGCACAC	292
LR812054.1	43062162	ACATTTCGGATGCACAC	43062146
LR812560.1	42722772	ACATTTCGGATGCACAC	42722756
CP002563.1	1553457	TTCGGATGCACACAATA	1553441
AC146941.8	83883	CATTTCGGATGCACACA	83867
LR812081.1	42184890	ACATTTCGGATGCACAC	42184906
LR812071.1	20700303	TTTCGGATGCACACAAT	20700319
LR812071.1	7146821	TTCGGATGCACACAA	7146835
LR778269.1	2268914	CATTTCGGATGCAGACAAT	2268896
LR778269.1	12469841	CATTTGGGATGCACACAAT	12469823
LR778269.1	40386131	CATTTCGGATGCACACA	40386147
LR778269.1	21275431	GGATGCACACAATAA	21275417
LR778267.1	2833714	CATTTGGGATGCACACAATAA	2833734
LR778267.1	42529115	GGATGCACACAATAA	42529101
LR778267.1	12460471	GGATGCACACAATAA	12460485
LR778260.1	16977017	CATTTGGGATGCACACAAT	16976999
LR778260.1	43845202	CATTTGGGATGCACACAAT	43845220
LR778260.1	18485245	CATTTCGGATGCACACA	18485261
LR778260.1	13433975	ATTTCGGATGCACAC	13433961
XM_029729683.1	3338	CATTTCGGATGCACACA	3354
XM_029729682.1	3442	CATTTCGGATGCACACA	3458
XM_029729681.1	3445	CATTTCGGATGCACACA	3461
LR584435.1	4663289	CATTTAGGATGCACACAATAA	4663309
LR584435.1	56328727	CATTTGGGATGCACACAAT	56328709
LR584435.1	50726731	CATTTCGGATGCACA	50726717
LR584433.1	2698528	CATTTAGGATGCACACAATAA	2698548
LR584423.1	23636461	CATTTCGGATGCACACA	23636477
LR535843.1	9318085	ACATTTCGGATGCACAC	9318101
LR131966.1	1611448	TCGGATGCACACAATAA	1611464
CP025673.1	18984285	CATTTCGGATGCACAGAATAA	18984305
LT594791.1	27812647	TTTCGGATGCACACAAT	27812663
LT594790.1	5296331	TTCGGATGCACACAATA	5296347
CP014500.1	975022	ACATTTCGGATGCACAC	975038
AP015036.1	43118478	ACATTTCGAATGCACACAATA	43118498
XM_014194909.1	3809	CATTTCGGATGCACACA	3825
XM_014194899.1	3914	CATTTCGGATGCACACA	3930
XM_014194890.1	3915	CATTTCGGATGCACACA	3931
CP012024.1	588716	TTCGGATGCACACAATA	588732
LK066269.1	301368	TCGGATGCACACAATAA	301384
LK066269.1	313490	TCGGATGCACACAATAA	313506
LK066269.1	325800	TCGGATGCACACAATAA	325816
LN013793.1	6993	CATTTCGGATGCACACA	7009
LN595178.1	201332	ACATTTCGGATGCTCACAATA	201352
LR899427.1	14480733	ACATTTCGGATGCACAC	14480749
LR862152.1	6253400	TTTCGGATGCACACAAT	6253416
LR812056.1	42149186	ACATTTCGGATGCACAC	42149202
LR812046.1	20731743	TTTCGGATGCACACAAT	20731759
HE601631.2	58795411	CATTTCGTATGCACACAAT	58795393
HE601631.2	55394958	TTTCGGATGCACACAAT	55394974

HE601631.2	17951684	CATTTCGGATGCACAC	17951669
LR812612.1	40748804	ACATTTCGGATGCACAC	40748820
LR812602.1	20725446	TTTCGGATGCACACAAT	20725462
LR812587.1	41935364	ACATTTCGGATGCACAC	41935380
LR812585.1	43351023	ACATTTCGGATGCACAC	43351039
LR812577.1	21923491	TTTCGGATGCACACAAT	21923507
LR812577.1	7565339	TTCGGATGCACACAA	7565353
LR812552.1	20433661	TTTCGGATGCACACAAT	20433677
LR812552.1	5522106	TTCGGATGCACACAA	5522120
AC136507.19	74014	CATTTCGGATGCACACA	74030
CR391919.5	156569	TTTCGGATGCACACAAT	156585
BX510640.4	83940	ACATTTCGGATGCACAC	83956
AC006247.12	139248	TTTCGGATGCACACAAT	139264
CP046680.1	41617088	GGATGCACACAATAA	41617102
CP046680.1	41620099	GGATGCACACAATAA	41620113

PRODUCTS OF SIGNIFICANT ALIGNMENTS (CDS FEATURE)

- **Zinc finger protein 622** GenBank; GenBank; GenBank
- **FAD synthetase 1, chloroplastic-like** GenBank; GenBank; GenBank; GenBank
- **Homocysteine S-methyltransferase-like** GenBank
- **Uncharacterized protein LOC117652271** GenBank; GenBank; GenBank
- **Endoplasmic reticulum membrane sensor NFE2L1-like** GenBank
- **Chromosome-associated kinesin KIF4A** GenBank
- **Uncharacterized protein LOC109080287** GenBank; GenBank
- **Uncharacterized protein LOC108071571** GenBank
- **Hypothetical protein** GenBank
- **Nuclear factor erythroid 2-related factor 3** GenBank
- **Protein tyrosine-phosphatase** GenBank
- **Manganese-dependent protein-tyrosine phosphatase** GenBank

11.2 SEQUENCE N° 8/1 (T=1; C=0; G=3; A=2) (Hitlist size 1000)

Query: **ACATTTCGGATGCACACAATAA**

Search Parameters	
Program	blastn
Word size	7
Expect value	1000
Hitlist size	250
Match/Mismatch scores	1,-3
Gapcosts	5,2
Filter string	F
Genetic Code	1

245 sequences selected:

Accession Seq	Description	Comomm Name	Max Score	Total Score	Query C.	E value	Per. Id.	Acc. Len
.....	.....	.....	.....	.....	.....	.....	.....	.....
CP054816.1	**Bacillus anthracis strain FDAARGOS_695 chromosome**	Bacillus anthracis	32.2	32.2	72%	138	100.00%	5135792
CP054797.1	**Bacillus anthracis strain FDAARGOS_703 chromosome**	Bacillus anthracis	32.2	32.2	72%	138	100.00%	5261520
CP050971.1	**Bacillus anthracis strain FDAARGOS_696 chromosome**	Bacillus anthracis	32.2	32.2	72%	138	100.00%	5255974
CP050970.1	**Bacillus anthracis strain FDAARGOS_694 chromosome**	Bacillus anthracis	32.2	32.2	72%	138	100.00%	5251490
.....	.....	.....	.....	.....	.....	.....	.....	.....
CP041981.1	**Bacillus paranthracis strain NCCP 15910 chromosome, complete genome**	Bacillus paranthracis	32.2	32.2	72%	138	100.00%	5208566
CP047131.1	**Bacillus anthracis str. BF1 chromosome, complete genome**	Bacillus anthracis str. BF1	32.2	32.2	72%	138	100.00%	5213532
.....	.....	.....	.....	.....	.....	.....	.....	.....
CP046564.1 1/2 3/1	Bacillus sp. N1-1 chromosome, complete genome	Bacillus sp. N1-1	32.2	32.2	72%	138	100.00%	4497340
CP047104.1	**Bacillus anthracis strain FDAARGOS_706 chromosome, complete genome**	Bacillus anthracis	32.2	32.2	72%	138	100.00%	5227454
CP047099.1	**Bacillus anthracis strain FDAARGOS_700 chromosome, complete genome**	Bacillus anthracis	32.2	32.2	72%	138	100.00%	5227419
CP047097.1	**Bacillus anthracis strain FDAARGOS_698 chromosome, complete genome**	Bacillus anthracis	32.2	32.2	72%	138	100.00%	5225746
.....	.....	.....	.....	.....	.....	.....	.....	.....
CP065149.1	**Bacillus paranthracis strain Gxun-30 chromosome, complete genome**	Bacillus paranthracis	32.2	32.2	72%	138	100.00%	5149464
.....	.....	.....	.....	.....	.....	.....	.....	.....
CP060194.1	**Bacillus anthracis strain Kanchipuram chromosome**	Bacillus anthracis	32.2	32.2	72%	138	100.00%	5227419
.....	.....	.....	.....	.....	.....	.....	.....	.....
LR862364.1	***Pipistrellus pipistrellus***	common pipistrelle	32.2	32.2	90%	138	95.00%	83522093

1/2 2/2 3/2 3/3 3/5 6/1 7/2 9/1 10/1	***genome assembly, chromosome: 8***							
.....	.....	.....	.....	.....	.....	.....	.....	.....
CP050973.1	**Bacillus anthracis strain FDAARGOS_699 chromosome**	Bacillus anthracis	32.2	32.2	72%	138	100.00%	5227345
CP050972.1	**Bacillus anthracis strain FDAARGOS_697 chromosome**	Bacillus anthracis	32.2	32.2	72%	138	100.00%	5240765
.....	.....	.....	.....	.....	.....	.....	.....	.....
CP047111.1	**Bacillus anthracis strain FDAARGOS_704 chromosome, complete genome**	Bacillus anthracis	32.2	32.2	72%	138	100.00%	5227450
CP047107.1	**Bacillus anthracis strain FDAARGOS_705 chromosome, complete genome**	Bacillus anthracis	32.2	32.2	72%	138	100.00%	5227452
CP047098.1	**Bacillus anthracis strain FDAARGOS_701 chromosome, complete genome**	Bacillus anthracis	32.2	32.2	72%	138	100.00%	5227419

Pipistrellus pipistrellus genome assembly, chromosome: 8

GenBank: LR862364.1

FASTA Graphics

Go to:

LOCUS LR862364 83522093 bp DNA linear MAM **03-AUG-2020**
DEFINITION Pipistrellus pipistrellus genome assembly, chromosome: 8.
ACCESSION LR862364
VERSION LR862364.1
DBLINK BioProject: PRJEB39564
BioSample: SAMEA994724
KEYWORDS .
SOURCE Pipistrellus pipistrellus (common pipistrelle)
ORGANISM Pipistrellus pipistrellus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Laurasiatheria; Chiroptera; Microchiroptera;
Vespertilionidae; Pipistrellus.
REFERENCE 1
CONSRTM Wellcome Sanger Institute Data Sharing
TITLE Direct Submission
JOURNAL **Submitted (27-JUL-2020) SC, Wellcome Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, United Kingdom**
COMMENT The assembly mPipPip1.1 is based on 62x PacBio data, 34x 10X
Genomics Chromium data, and 58x Arima Hi-C data generated at the
Wellcome Sanger Institute. The Hi-C data was derived from a
separate sample, mPipPip2. The assembly process included the
following sequence of steps: initial PacBio assembly generation
with Falcon-unzip, retained haplotig separation with purge_dups,
10X based scaffolding with scaff10x, Hi-C based scaffolding with
SALSA2, Arrow polishing, two rounds of FreeBayes polishing. The

mitochondrial assembly was produced at The Rockefeller University using mitoVGP. Finally, the assembly was analysed and manually improved using gEVAL. Chromosome-scale scaffolds confirmed by the Hi-C data have been named in order of size.

Pipistrellus pipistrellus genome assembly, chromosome: 8

Sequence ID: LR862364.1 **Length:** 83522093 **Number of Matches:** 1

Range 1: 52733517 to 52733536 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
32.2 bits(16)	138	19/20(95%)	0/20(0%)	Plus/Minus

```
Query  3         ATTTCGGATGCACACAATAA  22
                 ||||| ||||||||||||||
Sbjct  52733536  ATTTCAGATGCACACAATAA  52733517
```

PRODUCTS OF SIGNIFICANT ALIGNMENTS (CDS FEATURE)

Bacillus anthracis strain FDAARGOS_695 chromosome

GenBank: CP054816.1

FASTA Graphics

Go to:

LOCUS CP054816 16 bp DNA linear **BCT 17-JUN-2020**

DEFINITION Bacillus anthracis strain FDAARGOS 695 chromosome.

ACCESSION CP054816 REGION: 460078..460093

VERSION CP054816.1

DBLINK BioProject: PRJNA231221
BioSample: SAMN11056410

KEYWORDS .

SOURCE Bacillus anthracis

ORGANISM Bacillus anthracis
Bacteria; Firmicutes; Bacilli; Bacillales; Bacillaceae; Bacillus;
Bacillus cereus group.

REFERENCE 1 (bases 1 to 16)

AUTHORS Sozhamannan,S., Rosenzweig,N., Tallon,L., Sadzewicz,L.,
Vavikolanu,K., Mehta,A., Aluvathingal,J., Nadendla,S., Yan,Y. and
Sichtig,H.

TITLE FDA dAtabase for Regulatory Grade micrObial Sequences (FDA-ARGOS):
Supporting development and validation of Infectious Disease Dx
tests

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 16)

AUTHORS Sozhamannan,S., Rosenzweig,N., Tallon,L., Sadzewicz,L.,
Vavikolanu,K., Mehta,A., Aluvathingal,J., Nadendla,S., Yan,Y. and
Sichtig,H.

TITLE Direct Submission

JOURNAL **Submitted (28-MAY-2020) Center for Devices and Radiological Health, US Food and Drug Administration, 10903 New Hampshire Avenue, Silver Spring, MD 20993-0002, USA**

COMMENT The annotation was added by the NCBI Prokaryotic Genome Annotation
Pipeline (PGAP). Information about PGAP can be found here:
https://www.ncbi.nlm.nih.gov/genome/annotation_prok/
##Genome-Assembly-Data-START##
Assembly Method :: canu v. 1.2
Genome Coverage :: 494.25x
Sequencing Technology :: Pacbio; Illumina
##Genome-Assembly-Data-END##
##Genome-Annotation-Data-START##
Annotation Provider :: NCBI

Annotation Date :: 06/10/2020 00:47:27
Annotation Pipeline :: NCBI Prokaryotic Genome Annotation Pipeline (PGAP)
Annotation Method :: Best-placed reference protein set; GeneMarkS-2+
Annotation Software revision :: 4.11
Features Annotated :: Gene; CDS; rRNA; tRNA; ncRNA; repeat_region
Genes (total) :: 5,442
CDSs (total) :: 5,309
Genes (coding) :: 5,022
CDSs (with protein) :: 5,022
Genes (RNA) :: 133
rRNAs :: 11, 11, 11 (5S, 16S, 23S)
complete rRNAs :: 11, 11, 11 (5S, 16S, 23S)
tRNAs :: 95
ncRNAs :: 5
Pseudo Genes (total) :: 287
CDSs (without protein) :: 287
Pseudo Genes (ambiguous residues) :: 0 of 287
Pseudo Genes (frameshifted) :: 206 of 287
Pseudo Genes (incomplete) :: 34 of 287
Pseudo Genes (internal stop) :: 71 of 287
Pseudo Genes (multiple problems) :: 20 of 287
##Genome-Annotation-Data-END##

FEATURES Location/Qualifiers
source 1..16
/organism="Bacillus anthracis"
/mol_type="genomic DNA"
/strain="FDAARGOS 695"
/culture_collection="FDA:FDAARGOS_695"
/db_xref="taxon:1392"
/country="USA: MD"
/collected_by="U.S. Army Edgewood Chemical Biological Center"
/note="genetically modified strain"
gene complement(<1..>16)
/locus tag="FOC03 02710"
CDS complement(<1..>16)
/locus_tag="FOC03_02710"
/inference="COORDINATES: similar to AA sequence:RefSeq:WP_000109344.1"
/note="Derived by automated computational analysis using gene prediction method: Protein Homology."

/codon_start=3

/transl table=11

/product="iron-sulfur cluster-binding protein"

/protein_id="QKU23764.1"

/translation="MSVAGELDFLKRVDKALHDPQMRSAVPFTQDRLRMGREQATHEL
GNMEEWREAASEIRMHTIENLDSYLEQLATNVRKNGGHVHFAFKAEDAVSYVSEIVRK
KGARSVIKSKSMVSEEIHLNKNLEEMGVKVVEYDLGEYIIQLAGETPSHLIAPAIHKT
KEQVAELFSKEAGRKIPVDTNQLLQFAREKLRKEFLEADIGISGCNFAVAESGSVVLI
SNEGNARLTTTLPKVHIAIMGMERLVPTWKDLDVVISMLTRSATGQKITSYVTGINGA
RLHEDMDGPEEFHLIILDNGRSDILGTAYEEVLKCIRCGACVNVCPVYRNIGGHAYGS
VYSGPIGAVLAPLLEGYEELKELPFASSLCGACTDVCPVKIPLHDLLIEHRKDVVEQG
FITKFEQLAFKGFTFMVSHPFVYEKAIKMAHTGLSIFAKDGYIANGPGPLKGWTESRD
FPKPAKQSFRDWWKKETRDS"

ORIGIN

1 attgtgtgca tccgaa

//

Bacillus anthracis strain FDAARGOS_695 chromosome

GenBank: CP054816.1

FASTA Graphics

LOCUS CP054816 1413 bp DNA linear BCT 17-JUN-2020

DEFINITION Bacillus anthracis strain FDAARGOS 695 chromosome.

ACCESSION CP054816 REGION: 458850..460262

VERSION CP054816.1

DBLINK BioProject: PRJNA231221

BioSample: SAMN11056410

KEYWORDS .

SOURCE Bacillus anthracis

ORGANISM Bacillus anthracis

Bacteria; Firmicutes; Bacilli; Bacillales; Bacillaceae; Bacillus;

Bacillus cereus group.

REFERENCE 1 (bases 1 to 1413)

AUTHORS Sozhamannan,S., Rosenzweig,N., Tallon,L., Sadzewicz,L.,

Vavikolanu,K., Mehta,A., Aluvathingal,J., Nadendla,S., Yan,Y. and

Sichtig,H.

TITLE FDA dAtabase for Regulatory Grade micrObial Sequences (FDA-ARGOS):

Supporting development and validation of Infectious Disease Dx

tests

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 1413)

AUTHORS Sozhamannan,S., Rosenzweig,N., Tallon,L., Sadzewicz,L.,

Vavikolanu,K., Mehta,A., Aluvathingal,J., Nadendla,S., Yan,Y. and

Sichtig,H.

TITLE Direct Submission

```
  JOURNAL   Submitted (28-MAY-2020) Center for Devices and Radiological Health,
            US Food and Drug Administration, 10903 New Hampshire Avenue, Silver
            Spring, MD 20993-0002, USA
COMMENT     The annotation was added by the NCBI Prokaryotic Genome Annotation
            Pipeline (PGAP). Information about PGAP can be found here:
            https://www.ncbi.nlm.nih.gov/genome/annotation_prok/


            ##Genome-Assembly-Data-START##
            Assembly Method       :: canu v. 1.2
            Genome Coverage       :: 494.25x
            Sequencing Technology :: Pacbio; Illumina
            ##Genome-Assembly-Data-END##

            ##Genome-Annotation-Data-START##
            Annotation Provider               :: NCBI
            Annotation Date                   :: 06/10/2020 00:47:27
            Annotation Pipeline               :: NCBI Prokaryotic Genome
                                                 Annotation Pipeline (PGAP)
            Annotation Method                 :: Best-placed reference protein
                                                 set; GeneMarkS-2+
            Annotation Software revision      :: 4.11
            Features Annotated                :: Gene; CDS; rRNA; tRNA; ncRNA;
                                                 repeat_region
            Genes (total)                     :: 5,442
            CDSs (total)                      :: 5,309
            Genes (coding)                    :: 5,022
            CDSs (with protein)               :: 5,022
            Genes (RNA)                       :: 133
            rRNAs                             :: 11, 11, 11 (5S, 16S, 23S)
            complete rRNAs                    :: 11, 11, 11 (5S, 16S, 23S)
            tRNAs                             :: 95
            ncRNAs                            :: 5
            Pseudo Genes (total)              :: 287
            CDSs (without protein)            :: 287
            Pseudo Genes (ambiguous residues) :: 0 of 287
            Pseudo Genes (frameshifted)       :: 206 of 287
            Pseudo Genes (incomplete)         :: 34 of 287
            Pseudo Genes (internal stop)      :: 71 of 287
            Pseudo Genes (multiple problems)  :: 20 of 287
            ##Genome-Annotation-Data-END##
FEATURES             Location/Qualifiers
     source          1..1413
                     /organism="Bacillus anthracis"
                     /mol_type="genomic DNA"
```

/strain="FDAARGOS_695"
/culture collection="FDA:FDAARGOS 695"
/db_xref="taxon:1392"
/country="USA: MD"
/collected by="U.S. Army Edgewood Chemical Biological Center"
/note="genetically modified strain"

gene complement(1..1413)
/locus_tag="FOC03_02710"

CDS complement(1..1413)
/locus_tag="FOC03_02710"
/inference="COORDINATES: similar to AA sequence:RefSeq:WP 000109344.1"
/note="Derived by automated computational analysis using gene prediction method: Protein Homology."
/codon_start=1
/transl_table=11
/product="iron-sulfur cluster-binding protein"
/protein_id="QKU23764.1"
/translation="MSVAGELDFLKRVDKALHDPQMRSAVPFTQDRLRMGREQATHEL
GNMEEWREAASE**IRMHTI**ENLDSYLEQLATNVRKNGGHVHFAFKAEDAVSYVSEIVRK
KGARSVIKSKSMVSEEIHLNKNLEEMGVKVVEYDLGEYIIQLAGETPSHLIAPAIHKT
KEQVAELFSKEAGRKIPVDTNQLLQFAREKLRKEFLEADIGISGCNFAVAESGSVVLI
SNEGNARLTTTLPKVHIAIMGMERLVPTWKDLDVVISMLTRSATGQKITSYVTGINGA
RLHEDMDGPEEFHLIILDNGRSDILGTAYEEVLKCIRCGACVNVCPVYRNIGGHAYGS
VYSGPIGAVLAPLLEGYEELKELPFASSLCGACTDVCPVKIPLHDLLIEHRKDVVEQG
FITKFEQLAFKGFTFMVSHPFVYEKAIKMAHTGLSIFAKDGYIANGPGPLKGWTESRD
FPKPAKQSFRDWWKKETRDS"

gene complement(1410..>1413)
/locus tag="FOC03 02715"

CDS complement(1410..>1413)
/locus_tag="FOC03_02715"
/inference="COORDINATES: similar to AA sequence:RefSeq:WP_017151663.1"
/note="Derived by automated computational analysis using gene prediction method: Protein Homology."
/codon_start=2
/transl table=11
/product="(Fe-S)-binding protein"
/protein_id="QKU23765.1"
/translation="MKVSLFITCLSDVFFPQVGKSVVEIMNQCGVELDFPEGQTCCGQ
PAYNSGYQEDAKLAAKQMIKAFEHSEYIVTPSGSCASMVHHYYKEMFKGDSEWYEKAV
HLADRTYELTDFVVNILGKNDWKSKLVEKAVFHQSCHMSRALGIKEEPLKLLSQVEGL
DIKELPYCQDCCGFGGTFAVKMSSISETMVDEKIKHIEATEANLLIGADMGCLMNIGG
RLRRENKNIQVLHVAEVLAKGLNK"

```
ORIGIN
        1 ttatgaatcc ctcgtttctt ttttccacca atcacggaat gattgtttcg caggttttgg
       61 aaaatctcta ctttctgtcc atccttttaa tggacctggg ccattcgcaa tataaccatc
      121 ttttgcaaat atcgataacc ctgtatgtgc cattttaatt gctttctcat aaacgaatgg
      181 atgactcacc ataaacgtga aacctttaaa ggcaagttgc tcaaatttcg ttataaaccc
      241 ttgttccacg acatctttac gatgttcaat gagaaggtca tgcaatggaa ttttcacagg
      301 acaaacatct gtacatgctc cgcacaaact tgatgcaaaa ggtaattctt ttaactcttc
      361 atatccttct aataaaggcg caagtaccgc tccaattggc ccgctataaa cacttccata
      421 agcatggcct ccaatattgc gatacacagg acaaacattt acacaagctc cgcaccttat
      481 acactttaat acttcttcat aagcagttcc taaaatatca gaccggccat tatctaaaat
      541 aattaaatga aattcttcag gtccatccat atcttcatgt aaacgcgccc cattaattcc
      601 tgttacataa gatgtaatct tttggcctgt agcacttctt gttaacatcg aaatgactac
      661 atcaagatct ttccatgtcg gaacaaggcg ttccatcccc ataatcgcaa tatgaacctt
      721 tggcaatgtc gtagtaaggc gagcatttcc ttcattactt ataagcacaa ctgatccaga
      781 ttccgctacc gcaaagttac aaccggaaat accaatatca gcctctaaaa actctttacg
      841 caatttctcc cttgcaaatt gtagtaattg attcgtatca actgggatct ttcttcctgc
      901 ttccttagaa aaaagctctg ctacttgttc cttcgtctta tgaatggccg gggcgattaa
      961 atgggatggc gtttctccag ctaattgaat aatatactcc ccaagatcat actctactac
     1021 ttttaccccc atctcttcta agtttttatt aagatgtatt tcttctgaaa ccatggattt
     1081 tgacttaatt actgatctcg ctccttttt tcttacgatt tcagaaacat atgatacagc
     1141 atcttccgct ttgaaagcga aatgcacatg ccccccattt ttccgaacat tagtagctaa
     1201 ttgttccaaa taactatcta aattttcgat tgtgtgcatc cgaatttcag atgctgcctc
     1261 gcgccattct tccatgttcc ctaattcatg tgtagcctgt tcccttccca tccgaagccg
     1321 atcttgtgta aacggaaccg cgcttctcat ctgcggatca tgcaacgctt tatctactct
     1381 tttcaaaaaa tctagttccc ccgccacact cat
//
```

iron-sulfur cluster-binding protein [Bacillus anthracis]

GenBank: QKU23764.1

Identical Proteins FASTA Graphics

Go to:

```
LOCUS       QKU23764                470 aa            linear   BCT 17-JUN-2020
DEFINITION  iron-sulfur cluster-binding protein [Bacillus anthracis].
ACCESSION   QKU23764
VERSION     QKU23764.1
DBLINK      BioProject: PRJNA231221
            BioSample: SAMN11056410
DBSOURCE    accession CP054816.1
KEYWORDS    .
SOURCE      Bacillus anthracis
  ORGANISM  Bacillus anthracis
            Bacteria; Firmicutes; Bacilli; Bacillales; Bacillaceae; Bacillus;
            Bacillus cereus group.
REFERENCE   1  (residues 1 to 470)
```

AUTHORS Sozhamannan,S., Rosenzweig,N., Tallon,L., Sadzewicz,L., Vavikolanu,K., Mehta,A., Aluvathingal,J., Nadendla,S., Yan,Y. and Sichtig,H.

TITLE FDA dAtabase for Regulatory Grade micrObial Sequences (FDA-ARGOS): Supporting development and validation of Infectious Disease Dx tests

JOURNAL Unpublished

REFERENCE 2 (residues 1 to 470)

AUTHORS Sozhamannan,S., Rosenzweig,N., Tallon,L., Sadzewicz,L., Vavikolanu,K., Mehta,A., Aluvathingal,J., Nadendla,S., Yan,Y. and Sichtig,H.

TITLE Direct Submission

JOURNAL Submitted (28-MAY-2020) Center for Devices and Radiological Health, US Food and Drug Administration, 10903 New Hampshire Avenue, Silver Spring, MD 20993-0002, USA

COMMENT The annotation was added by the NCBI Prokaryotic Genome Annotation Pipeline (PGAP). Information about PGAP can be found here: https://www.ncbi.nlm.nih.gov/genome/annotation_prok/

##Genome-Annotation-Data-START##

Annotation Provider	:: NCBI
Annotation Date	:: 06/10/2020 00:47:27
Annotation Pipeline	:: NCBI Prokaryotic Genome Annotation Pipeline (PGAP)
Annotation Method	:: Best-placed reference protein set; GeneMarkS-2+
Annotation Software revision	:: 4.11
Features Annotated	:: Gene; CDS; rRNA; tRNA; ncRNA; repeat_region
Genes (total)	:: 5,442
CDSs (total)	:: 5,309
Genes (coding)	:: 5,022
CDSs (with protein)	:: 5,022
Genes (RNA)	:: 133
rRNAs	:: 11, 11, 11 (5S, 16S, 23S)
complete rRNAs	:: 11, 11, 11 (5S, 16S, 23S)
tRNAs	:: 95
ncRNAs	:: 5
Pseudo Genes (total)	:: 287
CDSs (without protein)	:: 287
Pseudo Genes (ambiguous residues)	:: 0 of 287
Pseudo Genes (frameshifted)	:: 206 of 287
Pseudo Genes (incomplete)	:: 34 of 287
Pseudo Genes (internal stop)	:: 71 of 287
Pseudo Genes (multiple problems)	:: 20 of 287

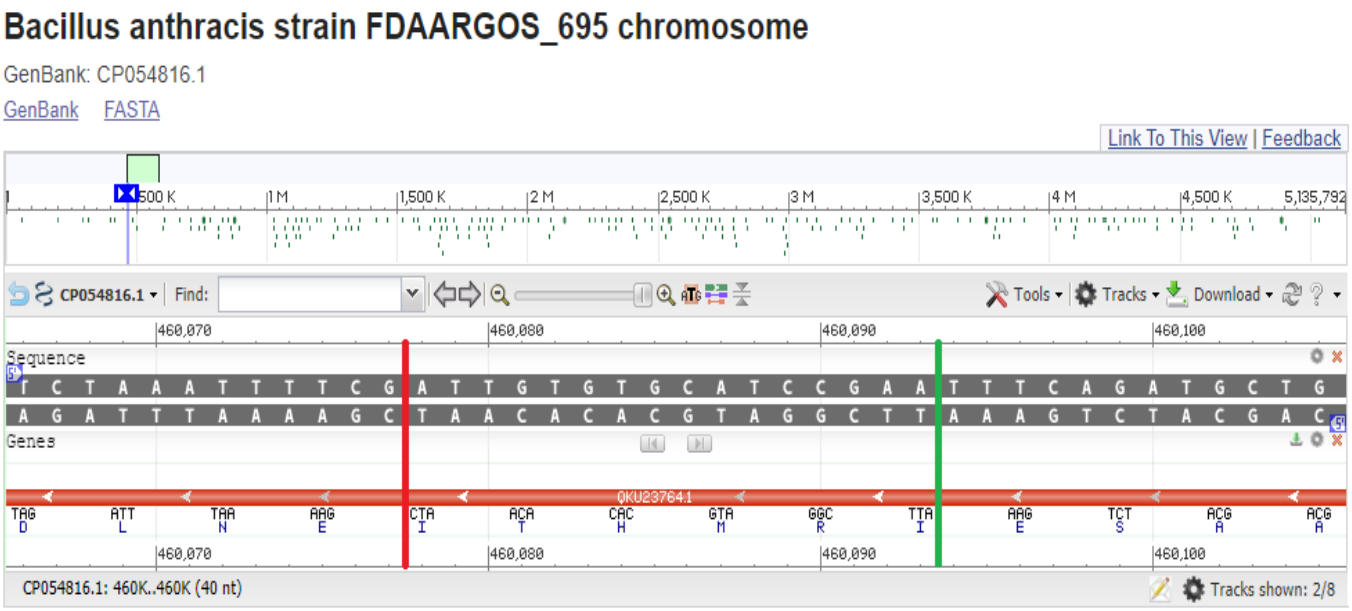
```
            ##Genome-Annotation-Data-END##
FEATURES             Location/Qualifiers
     source          1..470
                     /organism="Bacillus anthracis"
                     /strain="FDAARGOS 695"
                     /culture_collection="FDA:FDAARGOS_695"
                     /db xref="taxon:1392"
                     /country="USA: MD"
                     /collected_by="U.S. Army Edgewood Chemical Biological
                                    Center"
                     /note="genetically modified strain"
     Protein         1..470
                     /product="iron-sulfur cluster-binding protein"
     Region          8..462
                     /region_name="LutB"
                     /note="L-lactate utilization protein LutB, contains a
                            ferredoxin-type domain [Energy production and
                            conversion]; COG1139"
                     /db_xref="CDD:224062"
     CDS             1..470
                     /locus tag="FOC03 02710"
                     /coded_by="complement(CP054816.1:458850..460262)"
                     /inference="COORDINATES: similar to AA
                     sequence:RefSeq:WP_000109344.1"
                     /note="Derived by automated computational analysis using
                     gene prediction method: Protein Homology."
                     /transl_table=11
ORIGIN
        1 msvageldfl krvdkalhdp qmrsavpftq drlrmgreqa thelgnmeew reaase
       61 enldsyle qlatnvrkng ghvhfafkae davsyvseiv rkkgarsvik sksmvseeih
      121 lnknleemgv kvveydlgey iiqlagetps hliapaihkt keqvaelfsk eagrkipvdt
      181 nqllqfarek lrkefleadi gisgcnfava esgsvvlisn egnarltttl pkvhiaimgm
      241 erlvptwkdl dvvismltrs atgqkitsyv tgingarlhe dmdgpeefhl iildngrsdi
      301 lgtayeevlk circgacvnv cpvyrniggh aygsvysgpi gavlaplleg yeelkelpfa
      361 sslcgactdv cpvkiplhdl liehrkdvve qgfitkfeql afkgftfmvs hpfvyekaik
      421 mahtglsifa kdgyiangpg plkgwtesrd fpkpakqsfr dwwkketrds
//
```

Bacillus anthracis strain FDAARGOS_695 chromosome
Sequence ID: CP054816.1 **Length: 5135792 Number of Matches: 1**
Range 1: 460078 to 460093 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
32.2 bits(16)	138	16/16(100%)	0/16(0%)	Plus/Minus

Features:
iron-sulfur cluster-binding protein

```
CDS: Putative 1        1          R  M  H  T
Query                  5       TTCGGATGCACACAAT  20
                               ||||||||||||||||
Sbjct                  460093  TTCGGATGCACACAAT  460078
CDS:iron-sulfur clus   57      I  R  M  H  T  I
```



IRON-SULFUR CLUSTER-BINDING PROTEIN

JB Broderick , in Comprehensive Coordination Chemistry II , 2003 (https://doi.org/10.1016/B0-08-043748-6/08214-1) thus writes: "**Iron–sulfur clusters** are among the most ubiquitous and diverse metal-containing structures in biology and have been known for their role in electron transfer, including most notably in the *ferredoxins*, the *mitochondrial electron transport chain*, and *photosynthesis*….iron–sulfur clusters function (also) in regulatory roles, turning gene expression on or off in response to levels of iron, oxygen, or superoxide…. Evidence also points to an essential structural role for iron–sulfur clusters in several enzymes, including the DNA repair enzymes endonuclease III13 and MutY…. Iron–sulfur clusters can also be used directly in catalysis of redox chemistry on small molecules" (https://www.sciencedirect.com/topics/biochemistry-genetics-and-molecular-biology/iron-sulfur-cluster).
The protein identified by the **CP054816.1** sequence of **Bacillus Anthracis** is the **L-lactate utilization protein LutB, contains a *ferredoxin-type domain* [Energy production and conversion]** https://www.ncbi.nlm.nih.gov/Structure/cdd/cddsrv.cgi?uid=224062.

Bacillus anthracis strain FDAARGOS_703 chromosome

Sequence ID: CP054797.1 **Length:** 5261520 **Number of Matches:** 1

Range 1: 602816 to 602831 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
32.2 bits(16)	138	16/16(100%)	0/16(0%)	Plus/Minus

Features:
iron-sulfur cluster-binding protein

```
CDS: Putative 1       1          R  M  H  T
Query                 5       TTCGGATGCACACAAT  20
                              ||||||||||||||||
Sbjct                 602831  TTCGGATGCACACAAT  602816
CDS:iron-sulfur clus  57      I  R  M  H  T  I
```

Bacillus anthracis strain FDAARGOS_696 chromosome

Sequence ID: CP050971.1 **Length:** 5255974 **Number of Matches:** 1

Range 1: 597819 to 597834 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
32.2 bits(16)	138	16/16(100%)	0/16(0%)	Plus/Minus

Features:
iron-sulfur cluster-binding protein

```
CDS: Putative 1       1          R  M  H  T
Query                 5       TTCGGATGCACACAAT  20
                              ||||||||||||||||
Sbjct                 597834  TTCGGATGCACACAAT  597819
CDS:iron-sulfur clus  57      I  R  M  H  T  I
```

Bacillus anthracis strain FDAARGOS_694 chromosome

Sequence ID: CP050970.1 **Length:** 5251490 **Number of Matches:** 1

Range 1: 225247 to 225262 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
32.2 bits(16)	138	16/16(100%)	0/16(0%)	Plus/Minus

Features:
iron-sulfur cluster-binding protein

```
CDS: Putative 1       1          R  M  H  T
Query                 5       TTCGGATGCACACAAT  20
                              ||||||||||||||||
Sbjct                 225262  TTCGGATGCACACAAT  225247
CDS:iron-sulfur clus  57      I  R  M  H  T  I
```

Bacillus paranthracis strain NCCP 15910 chromosome, complete genome

Sequence ID: CP041981.1 **Length: 5208566 Number of Matches: 1**
Range 1: 3235474 to 3235489 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
32.2 bits(16)	138	16/16(100%)	0/16(0%)	Plus/Minus

```
CDS: Putative 1        1           R  M  H  T
Query                  5        TTCGGATGCACACAAT  20
                                |||||||||||||||
Sbjct                  3235489  TTCGGATGCACACAAT  3235474
CDS:iron-sulfur clus   57       I  R  M  H  T  I
```

Bacillus anthracis str. BF1 chromosome, complete genome

Sequence ID: CP047131.1 **Length: 5213532 Number of Matches: 1**
Range 1: 2995749 to 2995764 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
32.2 bits(16)	138	16/16(100%)	0/16(0%)	Plus/Minus

```
CDS: Putative 1        1           R  M  H  T
Query                  5        TTCGGATGCACACAAT  20
                                |||||||||||||||
Sbjct                  2995764  TTCGGATGCACACAAT  2995749
CDS:iron-sulfur clus   57       I  R  M  H  T  I
```

Bacillus anthracis strain FDAARGOS_706 chromosome, complete genome

Sequence ID: CP047104.1 **Length: 5227454 Number of Matches: 1**
Range 1: 767883 to 767898 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
32.2 bits(16)	138	16/16(100%)	0/16(0%)	Plus/Minus

```
CDS: Putative 1        1           R  M  H  T
Query                  5        TTCGGATGCACACAAT  20
                                |||||||||||||||
Sbjct                  767898   TTCGGATGCACACAAT  767883
CDS:iron-sulfur clus   57       I  R  M  H  T  I
```

Bacillus anthracis strain FDAARGOS_700 chromosome, complete genome

Sequence ID: CP047099.1 **Length: 5227419 Number of Matches: 1**
Range 1: 3386743 to 3386758 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
32.2 bits(16)	138	16/16(100%)	0/16(0%)	Plus/Minus

```
CDS: Putative 1        1           R  M  H  T
Query                  5        TTCGGATGCACACAAT  20
                                  ||||||||||||||||
Sbjct                  3386758  TTCGGATGCACACAAT  3386743
CDS:iron-sulfur clus   57       I  R  M  H  T  I
```

Bacillus anthracis strain FDAARGOS_698 chromosome, complete genome

Sequence ID: CP047097.1 **Length: 5225746 Number of Matches: 1**

Range 1: 2159321 to 2159336 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
32.2 bits(16)	138	16/16(100%)	0/16(0%)	Plus/Minus

```
CDS: Putative 1        1           R  M  H  T
Query                  5        TTCGGATGCACACAAT  20
                                  ||||||||||||||||
Sbjct                  2159336  TTCGGATGCACACAAT  2159321
CDS:iron-sulfur clus   57       I  R  M  H  T  I
```

Bacillus paranthracis strain Gxun-30 chromosome, complete genome

Sequence ID: CP065149.1 **Length: 5149464 Number of Matches: 1**

Range 1: 2963895 to 2963910 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
32.2 bits(16)	138	16/16(100%)	0/16(0%)	Plus/Minus

```
CDS: Putative 1        1           R  M  H  T
Query                  5        TTCGGATGCACACAAT  20
                                  ||||||||||||||||
Sbjct                  2963910  TTCGGATGCACACAAT  2963895
CDS:iron-sulfur clus   57       I  R  M  H  T  I
```

Bacillus anthracis strain Kanchipuram chromosome

Sequence ID: CP060194.1 **Length: 5227419 Number of Matches: 1**

Range 1: 3009241 to 3009256 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
32.2 bits(16)	138	16/16(100%)	0/16(0%)	Plus/Minus

```
Query  5         TTCGGATGCACACAAT  20
               ||||||||||||||||
Sbjct  3009256   TTCGGATGCACACAAT  3009241
```

Bacillus anthracis strain FDAARGOS_699 chromosome

Sequence ID: CP050973.1 **Length: 5227345 Number of Matches: 1**

Range 1: 546251 to 546266 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
32.2 bits(16)	138	16/16(100%)	0/16(0%)	Plus/Plus

Features:
iron-sulfur cluster-binding protein

```
CDS: Putative 1        1           R  M  H  T
Query                  5        TTCGGATGCACACAAT  20
                                ||||||||||||||||
Sbjct                  546251   TTCGGATGCACACAAT  546266
CDS:iron-sulfur clus   57       I  R  M  H  T  I
```

Bacillus anthracis strain FDAARGOS_697 chromosome

Sequence ID: CP050972.1 **Length: 5240765 Number of Matches: 1**

Range 1: 2510750 to 2510765 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
32.2 bits(16)	138	16/16(100%)	0/16(0%)	Plus/Plus

Features:
iron-sulfur cluster-binding protein

```
CDS: Putative 1        1           R  M  H  T
Query                  5        TTCGGATGCACACAAT  20
                                ||||||||||||||||
Sbjct                  2510750  TTCGGATGCACACAAT  2510765
CDS:iron-sulfur clus   57       I  R  M  H  T  I
```

Bacillus anthracis strain FDAARGOS_704 chromosome, complete genome

Sequence ID: CP047111.1 Length: 5227450 Number of Matches: 1

Range 1: 5091232 to 5091247 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
32.2 bits(16)	138	16/16(100%)	0/16(0%)	Plus/Plus

```
CDS: Putative 1       1           R  M  H  T
Query                 5        TTCGGATGCACACAAT  20
                               ||||||||||||||||
Sbjct                 5091232  TTCGGATGCACACAAT  5091247
CDS:iron-sulfur clus  57       I  R  M  H  T  I
```

Bacillus anthracis strain FDAARGOS_705 chromosome, complete genome

Sequence ID: CP047107.1 Length: 5227452 Number of Matches: 1

Range 1: 3977980 to 3977995 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
32.2 bits(16)	138	16/16(100%)	0/16(0%)	Plus/Plus

```
CDS: Putative 1       1           R  M  H  T
Query                 5        TTCGGATGCACACAAT  20
                               ||||||||||||||||
Sbjct                 3977980  TTCGGATGCACACAAT  3977995
CDS:iron-sulfur clus  57       I  R  M  H  T  I
```

Bacillus anthracis strain FDAARGOS_701 chromosome, complete genome

Sequence ID: CP047098.1 Length: 5227419 Number of Matches: 1

Range 1: 4610166 to 4610181 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
32.2 bits(16)	138	16/16(100%)	0/16(0%)	Plus/Plus

```
CDS: Putative 1       1           R  M  H  T
Query                 5        TTCGGATGCACACAAT  20
                               ||||||||||||||||
Sbjct                 4610166  TTCGGATGCACACAAT  4610181
CDS:iron-sulfur clus  57       I  R  M  H  T  I
```

CHAPTER XII

BLAST RESEARCHES OF OTHER MODIFIED SEQUENCES OF THE RNA SEQUENCE “GTGAAATGGTCATGTGTGGCGG” OF THE RdRP GENE (RdRp_SARSr-F)

GOALS OF CHAPTER XII

This chapter contains the results of BLAST research of other Modified Sequences of the RNA Sequence “GTGAAATGGTCATGTGTGGCGG” of the RdRp Gene (RdRp_SARSr-F). Modified sequences are sequences n° : **1/2**, **2/2**, **3/2**, **3/3**; **3/4**, **3/5**, **3/6**, **4/1**, **5/1**, **5/2**, **6/1**, **6/2**, **7/1**, **7/2**, **9/1**, and **10/1**.

WARNING: For the reasons already explained on page 437 (in “Objectives of chapter XI”) I am forced to leave scientists concerned with the arduous task (which I absolutely recommend to perform) to deepen the analysis and selection of the sequences considered relevant according to the indications proposed in the previous chapters.

SEQUENCE N° 1/2

(T=0; C=1; G=2; A=3)

SEQUENCE N° 1/2 (T=0; C=1; G=2; A=3)

Original Sequence	G	T	G	A	A	A	T	G	G	T	C	A	T	G	T	G	T	G	G	C	G	G
Trend n° 1 (T=0; C=1; G=2; A=3)	2	0	2	3	3	3	0	2	2	0	1	3	0	2	0	2	0	2	2	1	2	2
Trend n° 1 (I i D d)	-	D	I	I	i	i	D	I	i	D	I	I	D	I	D	I	D	I	i	D	I	i
Trend n° 1/2	1	0	1	2	2	2	1	3	3	0	2	3	1	3	1	3	1	3	3	0	1	1
Sequence n° 1/2 (86,36%)	C	T	C	G	G	G	C	A	A	T	G	A	C	A	C	A	C	A	A	T	C	C

Query: **CTCGGGCAATGACACACAATCC**

Search Parameters	
Program	blastn
Word size	7
Expect value	1000
Hitlist size	100
Match/Mismatch scores	1,-3
Gapcosts	5,2
Filter string	F
Genetic Code	1

100 sequences selected:

Accession Seq	Description	Comomm Name	Max Score	Total Score	Query C.	E value	Per. Id.	Acc. Len
CP022464.2	Enterocloster bolteae strain ATCC BAA-613 chromosome, complete genome	[Clostridium] bolteae	40.1	40.1	90%	0.56	100.00%	6570176
LR697106.1 2/2 3/1 3/3 3/4 8/1	Chanos chanos genome assembly, chromosome: 1	milkfish	38.2	100	86%	2.2	100.00%	63121898
LR633955.1 2/2 3/2 3/3	Gadus morhua genome assembly, chromosome: 13	Atlantic cod	36.2	66.4	81%	8.7	100.00%	28829685
XM_023272289.2	PREDICTED: Amphiprion ocellaris transmembrane protein 182-like (LOC111569869), mRNA	clown anemonefish	36.2	36.2	81%	8.7	100.00%	2086
AC225726.1 2/2 3/2 3/3 5/1 6/1 10/1	Homo sapiens FOSMID clone ABC14-50121100H1 from chromosome 1, complete sequence	human	36.2	36.2	81%	8.7	100.00%	38081
LR131992.1 3/3 3/4 3/6	Gouania willdenowi genome assembly, chromosome: 16	blunt-snouted clingfish	36.2	66.4	81%	8.7	100.00%	40234402
CP027092.1 3/6 7/1 10/1	Bos mutus isolate yakQH1 chromosome 24	wild yak	36.2	66.4	90%	8.7	100.00%	61837517
AL445218.12 3/3 3/5	Human DNA sequence from clone RP11-14O19 on chromosome 1, complete sequence	human	36.2	36.2	81%	8.7	100.00%	44203
LR812132.1 3/2 5/1	Erithacus rubecula genome assembly, chromosome: 28	European robin	34.2	34.2	77%	35	100.00%	5344592
CP054177.1	Bacillus subtilis strain JCL16 chromosome, complete genome	Bacillus subtilis	34.2	34.2	77%	35	100.00%	4101682
CP053102.1	Bacillus subtilis subsp. subtilis str. 168 chromosome, complete genome	Bacillus subtilis subsp. subtilis str. 168	34.2	34.2	77%	35	100.00%	4316079
CP052842.1	Bacillus subtilis subsp. subtilis str. 168 chromosome, complete genome	Bacillus subtilis subsp. subtilis str. 168	34.2	34.2	77%	35	100.00%	4398844
CP051466.1	Bacillus subtilis subsp. subtilis strain UCMB5021 chromosome, complete genome	Bacillus subtilis subsp. subtilis	34.2	34.2	77%	35	100.00%	4060035
CP051465.1	Bacillus subtilis subsp. subtilis strain UCMB5121 chromosome, complete genome	Bacillus subtilis subsp. subtilis	34.2	34.2	77%	35	100.00%	4059834
CP051462.1	Bacillus subtilis strain At3 chromosome, complete genome	Bacillus subtilis	34.2	34.2	77%	35	100.00%	4197841
CP051306.1	Bacillus subtilis strain RI4914	Bacillus subtilis	34.2	34.2	77%	35	100.00%	4100930

	chromosome, complete genome								
CP050532.1	Bacillus subtilis subsp. subtilis str. SMY chromosome, complete genome	Bacillus subtilis subsp. subtilis str. SMY	34.2	34.2	77%	35	100.00%	4212427
CP026662.1	Bacillus subtilis strain H1 chromosome, complete genome	Bacillus subtilis	34.2	34.2	77%	35	100.00%	4356270
CP046860.1	Bacillus subtilis strain RS10 chromosome	Bacillus subtilis	34.2	34.2	77%	35	100.00%	4457201
CP045672.1	Bacillus subtilis strain 2014-3557 chromosome, complete genome	Bacillus subtilis	34.2	34.2	77%	35	100.00%	4240660
CP031693.1	Bacillus subtilis strain SRCM101393 chromosome, complete genome	Bacillus subtilis	34.2	34.2	77%	35	100.00%	4089165
CP028215.1	Bacillus subtilis strain SRCM102750 chromosome, complete genome	Bacillus subtilis	34.2	34.2	77%	35	100.00%	4098278
CP028201.1	Bacillus subtilis strain SRCM102753 chromosome, complete genome	Bacillus subtilis	34.2	34.2	77%	35	100.00%	4025706
CP047564.1								
2/2 3/2 3/3 3/5 6/1 8/1 9/1 10/1	Solanum pinnatisectum cultivar CGN17745 chromosome 8	tansyleaf nightshade	34.2	64.4	81%	35	100.00%	56908857
CP047485.1	Bacillus subtilis strain BJQ0005 chromosome, complete genome	Bacillus subtilis	34.2	34.2	77%	35	100.00%	4112497
CP023409.1	Bacillus subtilis strain 7PJ-16 chromosome, complete genome	Bacillus subtilis	34.2	34.2	77%	35	100.00%	4209045
CP047325.1	Bacillus subtilis strain GOT9 chromosome, complete genome	Bacillus subtilis	34.2	34.2	77%	35	100.00%	4119195
XM_023214463.2 **9/1**	PREDICTED: Piliocolobus tephrosceles adaptor related protein complex 3 subunit mu 2 (AP3M2), transcript variant X1, mRNA	Ugandan red Colobus	34.2	34.2	77%	35	100.00%	3551
CP045811.1	Bacillus subtilis strain P9_B1 chromosome, complete genome	Bacillus subtilis	34.2	34.2	77%	35	100.00%	4063468
CP045812.1	Bacillus subtilis strain P8_B3 chromosome, complete genome	Bacillus subtilis	34.2	34.2	77%	35	100.00%	4215511
CP045922.1	Bacillus subtilis strain P8_B1 chromosome, complete genome	Bacillus subtilis	34.2	34.2	77%	35	100.00%	4215512
CP045816.1	Bacillus subtilis strain P5_B2 chromosome, complete genome	Bacillus subtilis	34.2	34.2	77%	35	100.00%	4103324
CP045817.1	Bacillus subtilis strain P5_B1 chromosome, complete genome	Bacillus subtilis	34.2	34.2	77%	35	100.00%	4083248
CP045818.1	Bacillus subtilis strain MB9_B6 chromosome, complete genome	Bacillus subtilis	34.2	34.2	77%	35	100.00%	4087720
CP045819.1	Bacillus subtilis strain MB9_B4 chromosome, complete genome	Bacillus subtilis	34.2	34.2	77%	35	100.00%	4105407
CP045820.1	Bacillus subtilis strain MB9_B1 chromosome, complete genome	Bacillus subtilis	34.2	34.2	77%	35	100.00%	4263919
CP045821.1	Bacillus subtilis strain MB8_B7 chromosome, complete genome	Bacillus subtilis	34.2	34.2	77%	35	100.00%	4191568
CP045823.1	Bacillus subtilis strain MB8_B1 chromosome, complete genome	Bacillus subtilis	34.2	34.2	77%	35	100.00%	4221278
CP045824.1	Bacillus subtilis strain MB8_B10 chromosome, complete genome	Bacillus subtilis	34.2	34.2	77%	35	100.00%	4225362
CP045825.1	Bacillus subtilis strain 75 chromosome, complete genome	Bacillus subtilis	34.2	34.2	77%	35	100.00%	4156459
CP045826.1	Bacillus subtilis strain 73 chromosome, complete genome	Bacillus subtilis	34.2	34.2	77%	35	100.00%	4166516
CP031784.1	Bacillus subtilis strain HMNig-2 chromosome, complete genome	Bacillus subtilis	34.2	34.2	77%	35	100.00%	4178124
CP031783.1	Bacillus subtilis strain MENO2 chromosome, complete genome	Bacillus subtilis	34.2	34.2	77%	35	100.00%	4083694
CP045425.1	Bacillus subtilis strain JAAA chromosome, complete genome	Bacillus subtilis	34.2	34.2	77%	35	100.00%	4217124
CP044498.1	Bacillus subtilis strain ms-2 chromosome, complete genome	Bacillus subtilis	34.2	34.2	77%	35	100.00%	4126713
XM_030153818.1 **2/2 3/1 3/5 3/6**	PREDICTED: Sphaeramia orbicularis Alport syndrome, mental retardation, midface hypoplasia and elliptocytosis chromosomal region gene 1 (ammecr1), mRNA	orbiculate cardinalfish	34.2	34.2	77%	35	100.00%	6788
CP041757.1 **3/1 8/1**	Bacillus sp. KBS0812 chromosome, complete genome	Bacillus sp. KBS0812	34.2	34.2	77%	35	100.00%	4215631
CP041360.1 **1/1**	Spirosoma sp. KCTC 42546 chromosome, complete genome	Spirosoma sp. KCTC 42546	34.2	34.2	77%	35	100.00%	8664091
LR606196.1	Aquila chrysaetos chrysaetos genome assembly, chromosome:	Aquila chrysaetos chrysaetos	34.2	34.2	95%	35	95.24%	30610376

16 3/2 3/3 3/5 9/1								
LR597471.1	Sphaeramia orbicularis genome assembly, chromosome: 14	orbiculate cardinalfish	34.2	34.2	77%	35	100.00%	55997211
LR597461.1 2/2 3/1 3/5 3/6	Sphaeramia orbicularis genome assembly, chromosome: 4	orbiculate cardinalfish	34.2	34.2	77%	35	100.00%	60192552
AP019714.1	Bacillus subtilis subsp. subtilis NBRC 13719 DNA, complete genome	Bacillus subtilis subsp. subtilis	34.2	34.2	77%	35	100.00%	4211090
CP034484.1	Bacillus subtilis subsp. subtilis NCIB 3610 = ATCC 6051 strain NCIB 3610 chromosome, complete genome	Bacillus subtilis subsp. subtilis NCIB 3610 = ATCC 6051	34.2	34.2	77%	35	100.00%	4211262
CP040528.1	Bacillus subtilis strain PR10 chromosome, complete genome	Bacillus subtilis	34.2	34.2	77%	35	100.00%	4073059
CP039935.1	Bacillus subtilis strain H19 chromosome, complete genome	Bacillus subtilis	34.2	34.2	77%	35	100.00%	4063004
CP033205.1	Bacillus subtilis strain MBI 600 chromosome, complete genome	Bacillus subtilis	34.2	34.2	77%	35	100.00%	4076736
LR584416.1 2/2 3/3 3/4 3/6 8/1	Salmo trutta genome assembly, chromosome: 3	river trout	34.2	157	77%	35	100.00%	74753489
CP029609.1	Bacillus subtilis subsp. subtilis strain G7 chromosome, complete genome	Bacillus subtilis subsp. subtilis	34.2	34.2	77%	35	100.00%	4216133
CP024767.1 5/1 7/1 7/2 9/1	Pseudomonas arsenicoxydans strain ACM1 chromosome	Pseudomonas arsenicoxydans	34.2	34.2	77%	35	100.00%	6620935
CP035414.1	Bacillus subtilis strain SRCM103637 chromosome, complete genome	Bacillus subtilis	34.2	34.2	77%	35	100.00%	4027238
CP035413.1	Bacillus subtilis strain SRCM103629 chromosome, complete genome	Bacillus subtilis	34.2	34.2	77%	35	100.00%	4129860
CP035411.1	Bacillus subtilis strain SRCM103622 chromosome, complete genome	Bacillus subtilis	34.2	34.2	77%	35	100.00%	4125217
CP035403.1	Bacillus subtilis strain SRCM103581 chromosome, complete genome	Bacillus subtilis	34.2	34.2	77%	35	100.00%	4126329
CP035406.1	Bacillus subtilis strain SRCM103612 chromosome, complete genome	Bacillus subtilis	34.2	34.2	77%	35	100.00%	4228232
CP035397.1	Bacillus subtilis strain SRCM103773 chromosome, complete genome	Bacillus subtilis	34.2	34.2	77%	35	100.00%	4052954
CP035400.1	Bacillus subtilis strain SRCM103835 chromosome, complete genome	Bacillus subtilis	34.2	34.2	77%	35	100.00%	4143632
CP046592.1	Bacillus subtilis strain TR21 chromosome, complete genome	Bacillus subtilis	34.2	34.2	77%	35	100.00%	4105857
CP062497.1	Bacillus subtilis strain CV16 chromosome, complete genome	Bacillus subtilis	34.2	34.2	77%	35	100.00%	4200880
CP060710.1	Bacillus subtilis subsp. subtilis strain DSM 10 chromosome, complete genome	Bacillus subtilis subsp. subtilis	34.2	34.2	77%	35	100.00%	4215636
CP060417.1	Bacillus subtilis subsp. subtilis strain ONU 559 chromosome, complete genome	Bacillus subtilis subsp. subtilis	34.2	34.2	77%	35	100.00%	4004302
CP049924.1	Bacillus subtilis strain So1b chromosome, complete genome	Bacillus subtilis	34.2	34.2	77%	35	100.00%	4070574
CP058242.1	Bacillus subtilis strain SP1 chromosome, complete genome	Bacillus subtilis	34.2	34.2	77%	35	100.00%	4215613
CP051860.1	Bacillus subtilis subsp. subtilis str. 168 chromosome, complete genome	Bacillus subtilis subsp. subtilis str. 168	34.2	34.2	77%	35	100.00%	4227167
CP050582.1 2/2 3/3 3/4 3/5 3/6 5/1 10/1	Canis lupus familiaris breed Labrador retriever chromosome 27a	dog	34.2	34.2	77%	35	100.00%	46050380
CP050640.1 2/2 3/3 3/4 3/5 3/6 5/1 10/1	Canis lupus familiaris breed Labrador retriever chromosome 27b	dog	34.2	34.2	77%	35	100.00%	46054595
CP050319.1	Bacillus subtilis strain AMR1 chromosome	Bacillus subtilis	34.2	34.2	77%	35	100.00%	4142132
LR778280.1 2/2 3/3 8/1 9/1	Coregonus sp. 'balchen' genome assembly, chromosome: 28	Coregonus sp. 'balchen'	34.2	34.2	77%	35	100.00%	46671285
LR778256.1 2/2 3/3 8/1 9/1	Coregonus sp. 'balchen' genome assembly, chromosome: 4	Coregonus sp. 'balchen'	34.2	64.4	81%	35	100.00%	92224161

XM_032439116.1	PREDICTED: Drosophila virilis octopamine receptor Oamb (LOC6629914), transcript variant X2, mRNA	Drosophila virilis	34.2	34.2	77%	35	100.00%	4221
XM_032439114.1	PREDICTED: Drosophila virilis octopamine receptor Oamb (LOC6629914), transcript variant X1, mRNA	Drosophila virilis	34.2	34.2	77%	35	100.00%	5158
CP028218.1	Bacillus subtilis strain SRCM102756 chromosome, complete genome	Bacillus subtilis	34.2	34.2	77%	35	100.00%	4145782
CP028217.1	Bacillus subtilis strain SRCM102751 chromosome, complete genome	Bacillus subtilis	34.2	34.2	77%	35	100.00%	4047680
CP028213.1	Bacillus subtilis strain SRCM102749 chromosome, complete genome	Bacillus subtilis	34.2	34.2	77%	35	100.00%	4192708
CP028212.1	Bacillus subtilis strain SRCM102748 chromosome, complete genome	Bacillus subtilis	34.2	34.2	77%	35	100.00%	4210797
CP028209.1	Bacillus subtilis strain SRCM102745 chromosome, complete genome	Bacillus subtilis	34.2	34.2	77%	35	100.00%	4102601
CP028202.1	Bacillus subtilis strain SRCM102754 chromosome, complete genome	Bacillus subtilis	34.2	34.2	77%	35	100.00%	4281863
CP046448.1	Bacillus subtilis strain ZD01 chromosome, complete genome	Bacillus subtilis	34.2	34.2	77%	35	100.00%	4015360
CP046047.1	Bacillus subtilis strain MSP5 chromosome	Bacillus subtilis	34.2	34.2	77%	35	100.00%	4049479
CP045006.1	Bacillus subtilis strain MSP1 chromosome	Bacillus subtilis	34.2	34.2	77%	35	100.00%	4049437
XM_031106775.1	PREDICTED: Quercus lobata **F-box** protein At3g07870-like (LOC115983915), mRNA	Quercus lobata	34.2	34.2	77%	35	100.00%	1990
LR697108.1 **2/2 3/1 3/3 3/4 8/1**	**Chanos chanos genome assembly, chromosome: 3**	milkfish	34.2	64.4	86%	35	100.00%	58651890
LR606189.1 **3/2 3/3 3/5 9/1**	Aquila chrysaetos chrysaetos genome assembly, chromosome: 9	Aquila chrysaetos chrysaetos	34.2	34.2	77%	35	100.00%	45241448
CP021123.1	Bacillus subtilis strain SEM-9 chromosome, complete genome	Bacillus subtilis	34.2	34.2	77%	35	100.00%	4121982
CP025987.1	Marinomonas sp. FW-1 chromosome, complete genome	Marinomonas sp. FW-1	34.2	34.2	77%	35	100.00%	3964876
LR584417.1 **2/2 3/3 3/4 3/6 8/1**	Salmo trutta genome assembly, chromosome: 37	river trout	34.2	64.4	77%	35	100.00%	35205783
LR535845.1	Mastacembelus armatus genome assembly, chromosome: 13	zig-zag eel	34.2	64.4	77%	35	100.00%	27498257
LR877217.1	**Acomys russatus genome assembly, chromosome: 6**	golden spiny mouse	34.2	64.4	77%	35	100.00%	89442464
LR862365.1 **2/2 3/2 3/3 3/5 6/1 7/2 8/1 9/1 10/1**	*Pipistrellus pipistrellus genome assembly, chromosome: 9*	common pipistrelle	34.2	68.4	77%	35	100.00%	72542072
LR828303.1	Ananas comosus genome assembly, chromosome: 23	pineapple	34.2	34.2	77%	35	100.00%	7979763
CP028812.1	Bacillus subtilis strain BL-01 chromosome, complete genome	Bacillus subtilis	34.2	34.2	77%	35	100.00%	4242126

Pipistrellus pipistrellus genome assembly, chromosome: 9

GenBank: LR862365.1

FASTA Graphics

```
LOCUS       LR862365                17 bp    DNA     linear   MAM 03-AUG-2020
DEFINITION  Pipistrellus pipistrellus genome assembly, chromosome: 9.
ACCESSION   LR862365 REGION: 22466815..22466831
VERSION     LR862365.1
DBLINK      BioProject: PRJEB39564
            BioSample: SAMEA994724
KEYWORDS    .
SOURCE      Pipistrellus pipistrellus (common pipistrelle)
  ORGANISM  Pipistrellus pipistrellus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Laurasiatheria; Chiroptera; Microchiroptera;
            Vespertilionidae; Pipistrellus.
REFERENCE   1
  CONSRTM   Wellcome Sanger Institute Data Sharing
  TITLE     Direct Submission
  JOURNAL   Submitted (27-JUL-2020) SC, Wellcome Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, United Kingdom
COMMENT     The assembly mPipPip1.1 is based on 62x PacBio data, 34x 10X
            Genomics Chromium data, and 58x Arima Hi-C data generated at the
            Wellcome Sanger Institute. The Hi-C data was derived from a
            separate sample, mPipPip2. The assembly process included the
            following sequence of steps: initial PacBio assembly generation
            with Falcon-unzip, retained haplotig separation with purge_dups,
            10X based scaffolding with scaff10x, Hi-C based scaffolding with
            SALSA2, Arrow polishing, two rounds of FreeBayes polishing. The
            mitochondrial assembly was produced at The Rockefeller University
            using mitoVGP. Finally, the assembly was analysed and manually
            improved using gEVAL. Chromosome-scale scaffolds confirmed by the
            Hi-C data have been named in order of size.
FEATURES             Location/Qualifiers
     source          1..17
                     /organism="Pipistrellus pipistrellus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:59474"
                     /chromosome="9"
ORIGIN
        1 gcaatgacac acaatcc
//
```

Pipistrellus pipistrellus genome assembly, chromosome: 9
Sequence ID: LR862365.1 Length: 72542072 Number of Matches: 2
Range 1: 22466815 to 22466831 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
34.2 bits(17)	35	17/17(100%)	0/17(0%)	Plus/Plus

```
Query  6         GCAATGACACACAATCC  22
               |||||||||||||||||
Sbjct  22466815  GCAATGACACACAATCC  22466831
```

Range 2: 22505624 to 22505640 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
34.2 bits(17)	35	17/17(100%)	0/17(0%)	Plus/Plus

```
Query  6         GCAATGACACACAATCC  22
               |||||||||||||||||
Sbjct  22505624  GCAATGACACACAATCC  22505640
```

ALIGNMENT VIEW FLAT QUERY-ANCHORED WITH LETTERS FOR IDENTITIES

Query range 1: 1 to 22

Query	**1**	**CTCGGGCAATGACACACAATCC**	**22**
CP022464.2	5016426	CGGGCAATGACACACAATCC	5016445
LR697106.1	56765126	GGGCAATGACACACAATCC	56765144
LR697106.1	15185948	GGCAATGACACACAAT	15185933
LR697106.1	19801905	GGCAATGACACACAA	19801919
LR633955.1	10851675	GGCAATGACACACAATCC	10851658
LR633955.1	19246919	CAATGACACACAATC	19246905
XM_023272289.2	2019	CGGGCAATGACACACAAT	2002
AC225726.1	5695	GGCAATGACACACAATCC	5678
LR131992.1	30835046	TCGGGCAATGACACACAA	30835063
LR131992.1	2784243	GGCAATGACACACAA	2784257
CP027092.1	45563484	CGGGCAATGACACACAAT	45563501
CP027092.1	57838314	AATGACACACAATCC	57838328
AL445218.12	9837	GGCAATGACACACAATCC	9854
LR812132.1	2930334	TCGGGCAATGACACACA	2930318
CP054177.1	1149598	GCAATGACACACAATCC	1149582
CP053102.1	1170504	GCAATGACACACAATCC	1170488
CP052842.1	1172842	GCAATGACACACAATCC	1172826
CP051466.1	1148553	GCAATGACACACAATCC	1148537
CP051465.1	1148304	GCAATGACACACAATCC	1148288
CP051462.1	1286758	GCAATGACACACAATCC	1286742
CP051306.1	1149953	GCAATGACACACAATCC	1149937

CP050532.1	1163761	GCAATGACACACAATCC	1163745
CP026662.1	3670433	GCAATGACACACAATCC	3670417
CP046860.1	220600	GCAATGACACACAATCC	220584
CP045672.1	1149573	GCAATGACACACAATCC	1149557
CP031693.1	1784581	GCAATGACACACAATCC	1784565
CP028215.1	3409502	GCAATGACACACAATCC	3409486
CP028201.1	958052	GCAATGACACACAATCC	958036
CP047564.1	55777957	GGGCAATGACACACAAT	55777941
CP047564.1	127201	CAATGACACACAATC	127187
CP047485.1	1188146	GCAATGACACACAATCC	1188130
CP023409.1	1217994	GCAATGACACACAATCC	1217978
CP047325.1	2666125	GCAATGACACACAATCC	2666109
XM_023214463.2	2390	GCAATGACACACAATCC	2374
CP045811.1	1149504	GCAATGACACACAATCC	1149488
CP045812.1	1170071	GCAATGACACACAATCC	1170055
CP045922.1	1170071	GCAATGACACACAATCC	1170055
CP045816.1	1179248	GCAATGACACACAATCC	1179232
CP045817.1	1136046	GCAATGACACACAATCC	1136030
CP045818.1	1138918	GCAATGACACACAATCC	1138902
CP045819.1	1157077	GCAATGACACACAATCC	1157061
CP045820.1	1138568	GCAATGACACACAATCC	1138552
CP045821.1	1149568	GCAATGACACACAATCC	1149552
CP045823.1	1138556	GCAATGACACACAATCC	1138540
CP045824.1	1148638	GCAATGACACACAATCC	1148622
CP045825.1	1183154	GCAATGACACACAATCC	1183138
CP045826.1	1214198	GCAATGACACACAATCC	1214182
CP031784.1	218244	GCAATGACACACAATCC	218228
CP031783.1	1658707	GCAATGACACACAATCC	1658691
CP045425.1	1137129	GCAATGACACACAATCC	1137113
CP044498.1	1123295	GCAATGACACACAATCC	1123279
XM_030153818.1	4465	GGGCAATGACACACAAT	4449
CP041757.1	1170071	GCAATGACACACAATCC	1170055
CP041360.1	3508421	GGCAATGACACACAATC	3508405
LR606196.1	9886770	TCGGGCAATGACTCACAATCC	9886750
LR597471.1	22884527	GGGCAATGACACACAAT	22884511
LR597461.1	29961218	GCAATGACACACAATCC	29961202
AP019714.1	1170075	GCAATGACACACAATCC	1170059
CP034484.1	1170078	GCAATGACACACAATCC	1170062
CP040528.1	1145059	GCAATGACACACAATCC	1145043
CP039935.1	1153914	GCAATGACACACAATCC	1153898
CP033205.1	1149966	GCAATGACACACAATCC	1149950
LR584416.1	58542907	GCAATGACACACAATCC	58542891
LR584416.1	50619263	CAATGACACACAATCC	50619248
LR584416.1	36250074	CAATGACACACAATC	36250060
LR584416.1	45280676	CAATGACACACAATC	45280662
LR584416.1	41598409	AATGACACACAATCC	41598423
CP029609.1	1187677	GCAATGACACACAATCC	1187661
CP024767.1	4158911	CTCGGGCAATGACACAC	4158895
CP035414.1	1160645	GCAATGACACACAATCC	1160629
CP035413.1	1186728	GCAATGACACACAATCC	1186712
CP035411.1	1144667	GCAATGACACACAATCC	1144651
CP035403.1	1233149	GCAATGACACACAATCC	1233133
CP035406.1	1181052	GCAATGACACACAATCC	1181036
CP035397.1	1163029	GCAATGACACACAATCC	1163013
CP035400.1	1160029	GCAATGACACACAATCC	1160013
CP046592.1	1187398	GCAATGACACACAATCC	1187382
CP062497.1	1163796	GCAATGACACACAATCC	1163780
CP060710.1	1170515	GCAATGACACACAATCC	1170499
CP060417.1	1135143	GCAATGACACACAATCC	1135127
CP049924.1	1128410	GCAATGACACACAATCC	1128394

CP058242.1	1170479	GCAATGACACACAATCC	1170463
CP051860.1	4093524	GCAATGACACACAATCC	4093540
CP050582.1	4860768	CTCGGGCAATGACACAC	4860784
CP050640.1	4858108	CTCGGGCAATGACACAC	4858124
CP050319.1	3061266	GCAATGACACACAATCC	3061282
LR778280.1	30594062	GCAATGACACACAATCC	30594078
LR778256.1	2397833	GGCAATGACACACAATC	2397849
LR778256.1	55782119	GGGCAATGACACACA	55782133
XM_032439116.1	354	CTCGGGCAATGACACAC	370
XM_032439114.1	333	CTCGGGCAATGACACAC	349
CP028218.1	903496	GCAATGACACACAATCC	903512
CP028217.1	3858288	GCAATGACACACAATCC	3858304
CP028213.1	2531942	GCAATGACACACAATCC	2531958
CP028212.1	1598342	GCAATGACACACAATCC	1598358
CP028209.1	3607020	GCAATGACACACAATCC	3607036
CP028202.1	1844442	GCAATGACACACAATCC	1844458
CP046448.1	3051899	GCAATGACACACAATCC	3051915
CP046047.1	2932105	GCAATGACACACAATCC	2932121
CP045006.1	2934342	GCAATGACACACAATCC	2934358
XM_031106775.1	1434	GGGCAATGACACACAAT	1450
LR697108.1	33470624	GCAATGACACACAATCC	33470640
LR697108.1	52963245	GGGCAATGACACACA	52963231
LR606189.1	2584263	GCAATGACACACAATCC	2584279
CP021123.1	2980154	GCAATGACACACAATCC	2980170
CP025987.1	2995424	GGCAATGACACACAATC	2995440
LR584417.1	17130184	GCAATGACACACAATCC	17130200
LR584417.1	25903436	AATGACACACAATCC	25903422
LR535845.1	6973460	GGCAATGACACACAATC	6973476
LR535845.1	8298486	GCAATGACACACAAT	8298472
LR877217.1	81321755	GGGCAATGACACACAAT	81321771
LR877217.1	64453847	GCAATGACACACAAT	64453833
LR862365.1	22466815	GCAATGACACACAATCC	22466831
LR862365.1	22505624	GCAATGACACACAATCC	22505640
LR828303.1	653605	CGGGCAATGACACACAA	653621
CP028812.1	2442752	GCAATGACACACAATCC	2442768

PRODUCTS OF SIGNIFICANT ALIGNMENTS (CDS FEATURE)

- GGDEF domain-containing protein

https://www.ncbi.nlm.nih.gov/nucleotide/CP022464.2?report=gbwithparts&from=5015702&to=5017594&RID=W5C1YPBB013

- Adenylyl-sulfate kinase
- Sulfate adenylyltransferase

https://www.ncbi.nlm.nih.gov/nucleotide/CP054177.1?report=genbank&log$=nuclalign&blast_rank=10&RID=W5C1YPBB013&from=1149582&to=1149598
https://www.ncbi.nlm.nih.gov/nucleotide/CP053102.1?report=genbank&log$=nuclalign&blast_rank=11&RID=W5C1YPBB013&from=1170488&to=1170504
https://www.ncbi.nlm.nih.gov/nucleotide/CP052842.1?report=genbank&log$=nuclalign&blast_rank=12&RID=W5C1YPBB013&from=1172826&to=1172842
https://www.ncbi.nlm.nih.gov/nucleotide/CP051466.1?report=genbank&log$=nuclalign&blast_rank=13&RID=W5C1YPBB013&from=1148537&to=1148553
https://www.ncbi.nlm.nih.gov/nucleotide/CP051465.1?report=genbank&log$=nuclalign&blast_rank=14&RID=W5C1YPBB013&from=1148288&to=1148304
https://www.ncbi.nlm.nih.gov/nucleotide/CP051462.1?report=genbank&log$=nuclalign&blast_rank=15&RID=W5C1YPBB013&from=1286742&to=1286758
https://www.ncbi.nlm.nih.gov/nucleotide/CP051306.1?report=genbank&log$=nuclalign&blast_rank=16&RID=W5C1YPBB013&from=1149937&to=1149953
https://www.ncbi.nlm.nih.gov/nucleotide/CP050532.1?report=genbank&log$=nuclalign&blast_rank=17&RID=W5C1YPBB013&from=1163745&to=1163761
https://www.ncbi.nlm.nih.gov/nucleotide/CP046860.1?report=genbank&log$=nuclalign&blast_rank=19&RID=W5C1YPBB013&from=220584&to=220600
https://www.ncbi.nlm.nih.gov/nucleotide/CP031693.1?report=genbank&log$=nuclalign&blast_rank=21&RID=W5C1YPBB013&from=1784565&to=1784581
https://www.ncbi.nlm.nih.gov/nucleotide/CP028215.1?report=genbank&log$=nuclalign&blast_rank=22&RID=W5C1YPBB013&from=3409486&to=3409502
https://www.ncbi.nlm.nih.gov/nucleotide/CP028201.1?report=genbank&log$=nuclalign&blast_rank=23&RID=W5C1YPBB013&from=958036&to=958052
https://www.ncbi.nlm.nih.gov/nucleotide/CP047485.1?report=genbank&log$=nuclalign&blast_rank=25&RID=W5C1YPBB013&from=1188130&to=1188146
https://www.ncbi.nlm.nih.gov/nucleotide/CP023409.1?report=genbank&log$=nuclalign&blast_rank=26&RID=W5C1YPBB013&from=1217978&to=1217994
https://www.ncbi.nlm.nih.gov/nucleotide/CP047325.1?report=genbank&log$=nuclalign&blast_rank=27&RID=W5C1YPBB013&from=2666109&to=2666125
https://www.ncbi.nlm.nih.gov/nucleotide/CP045811.1?report=genbank&log$=nuclalign&blast_rank=29&RID=W5C1YPBB013&from=1149488&to=1149504
https://www.ncbi.nlm.nih.gov/nucleotide/CP045812.1?report=genbank&log$=nuclalign&blast_rank=30&RID=W5C1YPBB013&from=1170055&to=1170071
https://www.ncbi.nlm.nih.gov/nucleotide/CP045922.1?report=genbank&log$=nuclalign&blast_rank=31&RID=W5C1YPBB013&from=1170055&to=1170071
https://www.ncbi.nlm.nih.gov/nucleotide/CP045816.1?report=genbank&log$=nuclalign&blast_rank=32&RID=W5C1YPBB013&from=1179232&to=1179248
https://www.ncbi.nlm.nih.gov/nucleotide/CP045817.1?report=genbank&log$=nuclalign&blast_rank=33&RID=W5C1YPBB013&from=1136030&to=1136046
https://www.ncbi.nlm.nih.gov/nucleotide/CP045818.1?report=genbank&log$=nuclalign&blast_rank=34&RID=W5C1YPBB013&from=1138902&to=1138918
https://www.ncbi.nlm.nih.gov/nucleotide/CP045819.1?report=genbank&log$=nuclalign&blast_rank=35&RID=W5C1YPBB013&from=1157061&to=1157077
https://www.ncbi.nlm.nih.gov/nucleotide/CP045820.1?report=genbank&log$=nuclalign&blast_rank=36&RID=W5C1YPBB013&from=1138552&to=1138568
https://www.ncbi.nlm.nih.gov/nucleotide/CP045821.1?report=genbank&log$=nuclalign&blast_rank=37&RID=W5C1YPBB013&from=1149552&to=1149568
https://www.ncbi.nlm.nih.gov/nucleotide/CP045823.1?report=genbank&log$=nuclalign&blast_rank=38&RID=W5C1YPBB013&from=1138540&to=1138556
https://www.ncbi.nlm.nih.gov/nucleotide/CP045824.1?report=genbank&log$=nuclalign&blast_rank=39&RID=W5C1YPBB013&from=1148622&to=1148638
https://www.ncbi.nlm.nih.gov/nucleotide/CP045825.1?report=genbank&log$=nuclalign&blast_rank=40&RID=W5C1YPBB013&from=1183138&to=1183154
https://www.ncbi.nlm.nih.gov/nucleotide/CP045826.1?report=genbank&log$=nuclalign&blast_rank=41&RID=W5C1YPBB013&from=1214182&to=1214198
https://www.ncbi.nlm.nih.gov/nucleotide/CP031784.1?report=genbank&log$=nuclalign&blast_rank=42&RID=W5C1YPBB013&from=218228&to=218244
https://www.ncbi.nlm.nih.gov/nucleotide/CP031783.1?report=genbank&log$=nuclalign&blast_rank=43&RID=W5C1YPBB013&from=1658691&to=1658707
https://www.ncbi.nlm.nih.gov/nucleotide/CP041757.1?report=genbank&log$=nuclalign&blast_rank=47&RID=W5C1YPBB013&from=1170055&to=1170071
https://www.ncbi.nlm.nih.gov/nucleotide/AP019714.1?report=genbank&log$=nuclalign&blast_rank=52&RID=W5C1YPBB013&from=1170059&to=1170075
https://www.ncbi.nlm.nih.gov/nucleotide/CP034484.1?report=genbank&log$=nuclalign&blast_rank=53&RID=W5C1YPBB013&from=1170062&to=1170078
https://www.ncbi.nlm.nih.gov/nucleotide/CP040528.1?report=genbank&log$=nuclalign&blast_rank=54&RID=W5C1YPBB013&from=1145043&to=1145059

https://www.ncbi.nlm.nih.gov/nucleotide/CP033205.1?report=genbank&log$=nuclalign&blast_rank=56&RID=W5C1YPBB013&from=1149950&to=1149966
https://www.ncbi.nlm.nih.gov/nucleotide/CP029609.1?report=genbank&log$=nuclalign&blast_rank=58&RID=W5C1YPBB013&from=1187661&to=1187677
https://www.ncbi.nlm.nih.gov/nucleotide/CP035414.1?report=genbank&log$=nuclalign&blast_rank=60&RID=W5C1YPBB013&from=1160629&to=1160645
https://www.ncbi.nlm.nih.gov/nucleotide/CP035413.1?report=genbank&log$=nuclalign&blast_rank=61&RID=W5C1YPBB013&from=1186712&to=1186728
https://www.ncbi.nlm.nih.gov/nucleotide/CP035411.1?report=genbank&log$=nuclalign&blast_rank=62&RID=W5C1YPBB013&from=1144651&to=1144667
https://www.ncbi.nlm.nih.gov/nucleotide/CP035403.1?report=genbank&log$=nuclalign&blast_rank=63&RID=W5C1YPBB013&from=1233133&to=1233149
https://www.ncbi.nlm.nih.gov/nucleotide/CP035406.1?report=genbank&log$=nuclalign&blast_rank=64&RID=W5C1YPBB013&from=1181036&to=1181052
https://www.ncbi.nlm.nih.gov/nucleotide/CP035397.1?report=genbank&log$=nuclalign&blast_rank=65&RID=W5C1YPBB013&from=1163013&to=1163029
https://www.ncbi.nlm.nih.gov/nucleotide/CP035400.1?report=genbank&log$=nuclalign&blast_rank=66&RID=W5C1YPBB013&from=1160013&to=1160029
https://www.ncbi.nlm.nih.gov/nucleotide/CP046592.1?report=genbank&log$=nuclalign&blast_rank=67&RID=W5C1YPBB013&from=1187382&to=1187398
https://www.ncbi.nlm.nih.gov/nucleotide/CP062497.1?report=genbank&log$=nuclalign&blast_rank=68&RID=W5C1YPBB013&from=1163780&to=1163796
https://www.ncbi.nlm.nih.gov/nucleotide/CP060710.1?report=genbank&log$=nuclalign&blast_rank=69&RID=W5C1YPBB013&from=1170499&to=1170515
https://www.ncbi.nlm.nih.gov/nucleotide/CP060417.1?report=genbank&log$=nuclalign&blast_rank=70&RID=W5C1YPBB013&from=1135127&to=1135143
https://www.ncbi.nlm.nih.gov/nucleotide/CP049924.1?report=genbank&log$=nuclalign&blast_rank=71&RID=W5C1YPBB013&from=1128394&to=1128410
https://www.ncbi.nlm.nih.gov/nucleotide/CP058242.1?report=genbank&log$=nuclalign&blast_rank=72&RID=W5C1YPBB013&from=1170463&to=1170479
https://www.ncbi.nlm.nih.gov/nucleotide/CP051860.1?report=genbank&log$=nuclalign&blast_rank=73&RID=W5C1YPBB013&from=4093524&to=4093540
https://www.ncbi.nlm.nih.gov/nucleotide/CP050319.1?report=genbank&log$=nuclalign&blast_rank=76&RID=W5C1YPBB013&from=3061266&to=3061282
https://www.ncbi.nlm.nih.gov/nucleotide/CP028218.1?report=genbank&log$=nuclalign&blast_rank=81&RID=W5C1YPBB013&from=903496&to=903512
https://www.ncbi.nlm.nih.gov/nucleotide/CP028217.1?report=genbank&log$=nuclalign&blast_rank=82&RID=W5C1YPBB013&from=3858288&to=3858304
https://www.ncbi.nlm.nih.gov/nucleotide/CP028213.1?report=genbank&log$=nuclalign&blast_rank=83&RID=W5C1YPBB013&from=2531942&to=2531958
https://www.ncbi.nlm.nih.gov/nucleotide/CP028212.1?report=genbank&log$=nuclalign&blast_rank=84&RID=W5C1YPBB013&from=1598342&to=1598358
https://www.ncbi.nlm.nih.gov/nucleotide/CP028209.1?report=genbank&log$=nuclalign&blast_rank=85&RID=W5C1YPBB013&from=3607020&to=3607036
https://www.ncbi.nlm.nih.gov/nucleotide/CP028202.1?report=genbank&log$=nuclalign&blast_rank=86&RID=W5C1YPBB013&from=1844442&to=1844458
https://www.ncbi.nlm.nih.gov/nucleotide/CP046448.1?report=genbank&log$=nuclalign&blast_rank=87&RID=W5C1YPBB013&from=3051899&to=3051915
https://www.ncbi.nlm.nih.gov/nucleotide/CP046047.1?report=genbank&log$=nuclalign&blast_rank=88&RID=W5C1YPBB013&from=2932105&to=2932121
https://www.ncbi.nlm.nih.gov/nucleotide/CP045006.1?report=genbank&log$=nuclalign&blast_rank=89&RID=W5C1YPBB013&from=2934342&to=2934358
https://www.ncbi.nlm.nih.gov/nucleotide/XM_031106775.1?report=genbank&log$=nuclalign&blast_rank=90&RID=W5C1YPBB013&from=1434&to=1450
https://www.ncbi.nlm.nih.gov/nucleotide/CP021123.1?report=genbank&log$=nuclalign&blast_rank=93&RID=W5C1YPBB013&from=2980154&to=2980170

- Mannuronan epimerase

https://www.ncbi.nlm.nih.gov/nucleotide/CP024767.1?report=gbwithparts&from=4157795&to=4163413&RID=W5C1YPBB013

SEQUENCE N° 2/2

(T=3; C=2; G=1; A=0)

SEQUENCE N° 2/2 (T=3; C=2; G=1; A=0)

Original Sequence	G	T	G	A	A	A	T	G	G	T	C	A	T	G	T	G	T	G	G	C	G	G
Trend n° 2 (T=3; C=2; G=1; A=0)	1	3	1	0	0	0	3	1	1	3	2	0	3	1	3	1	3	1	1	2	1	1
Trend n° 2 (I i D d)	-	I	D	D	d	d	I	D	d	I	D	D	I	D	I	D	I	D	d	I	D	d
Trend n° 2/2	0	3	2	1	1	1	3	0	0	3	1	0	2	0	2	0	2	0	0	3	0	0
Sequence n° 2/2 (81,82%)	A	T	C	G	G	G	T	C	C	T	G	A	C	A	C	A	C	A	A	G	A	A

Query: **ATCGGGTCCTGACACACAAGAA**

Search Parameters	
Program	blastn
Word size	7
Expect value	1000
Hitlist size	100
Match/Mismatch scores	1,-3
Gapcosts	5,2
Filter string	F
Genetic Code	1

99 sequences selected:

Accession Seq	Description	Comomm Name	Max Score	Total Score	Query C.	E value	Per. Id.	Acc. Len
CP034495.1	Eukaryotic synthetic construct chromosome 17	eukaryotic synthetic construct	38.2	128	95%	2.2	100.00%	88299790
AP023477.1 1/2 3/2 3/3 5/1 6/1 10/1	Homo sapiens DNA, chromosome 17, nearly complete genome	human	38.2	98.6	86%	2.2	100.00%	80688777
AC005411.1 1/2 3/2 3/3 5/1 6/1 10/1	Homo sapiens chromosome 17, clone hRPK.721_K_1, complete sequence	human	38.2	38.2	86%	2.2	100.00%	163908
LR584236.2	Takifugu rubripes genome assembly, chromosome: 20	torafugu	36.2	36.2	81%	8.7	100.00%	18239318
LR584433.1 1/2 3/3 3/4 3/6 8/1	Salmo trutta genome assembly, chromosome: 5	river trout	36.2	36.2	81%	8.7	100.00%	67763595
CP025669.1	**Ipomoea triloba cultivar NCNSP0323 chromosome 10**	trilobed morning glory	36.2	36.2	81%	8.7	100.00%	29809665
CP025666.1	**Ipomoea triloba cultivar NCNSP0323 chromosome 7**	trilobed morning glory	36.2	36.2	81%	8.7	100.00%	28894297
LN590708.1	Cyprinus carpio genome assembly common carp genome, scaffold: LG39, chromosome: 39	common carp	36.2	66.4	81%	8.7	100.00%	16255708
LR862367.1 1/2 3/2 3/3 3/5 6/1 7/2 8/1 9/1 10/1	*Pipistrellus pipistrellus genome assembly, chromosome: 11*	common pipistrelle	36.2	36.2	81%	8.7	100.00%	69386144
XR_004133838.1	PREDICTED: Camelus dromedarius uncharacterized LOC116150175 (LOC116150175), transcript variant X2, ncRNA	Arabian camel	36.2	36.2	81%	8.7	100.00%	622
XM_031265003.1	**PREDICTED: Ipomoea triloba zinc finger BED domain-containing protein RICESLEEPER 2-like (LOC116024098), mRNA**	trilobed morning glory	36.2	36.2	81%	8.7	100.00%	1971
LR633957.1 1/2 3/2 3/3	Gadus morhua genome assembly, chromosome: 15	Atlantic cod	36.2	36.2	81%	8.7	100.00%	28657694
LR597469.1 1/2 3/1 3/5 3/6	Sphaeramia orbicularis genome assembly, chromosome: 12	orbiculate cardinalfish	36.2	102	86%	8.7	100.00%	81199652

CP025664.1	**Ipomoea triloba cultivar NCNSP0323 chromosome 5**	trilobed morning glory	36.2	36.2	81%	8.7	100.00%	32570749
XM_018226941.1	PREDICTED: Xenopus laevis catenin beta interacting protein 1 S homeolog (ctnnbip1.S), transcript variant X1, mRNA	African clawed frog	36.2	36.2	81%	8.7	100.00%	2264
LK066360.1	Apteryx australis mantelli genome assembly AptMant0, scaffold scaffold1686	Apteryx australis mantelli	36.2	36.2	81%	8.7	100.00%	207642
LR812500.1	Danio aesculapii genome assembly, chromosome: 4	Danio aesculapii	34.2	102	77%	35	100.00%	60163563
LR812084.1	Danio rerio genome assembly, chromosome: 22	zebrafish	34.2	64.4	77%	35	100.00%	38893734
XM_035010550.1	PREDICTED: Mirounga leonina centromere protein F (CENPF), mRNA	Southern elephant seal	34.2	34.2	77%	35	100.00%	10433
CP051060.1	Aspergillus flavus strain A1 chromosome 2	Aspergillus flavus	34.2	34.2	95%	35	95.24%	6587329
CP051036.1	Aspergillus flavus strain A9 chromosome 2	Aspergillus flavus	34.2	34.2	95%	35	95.24%	6230940
CP051020.1	Aspergillus flavus strain AF36 chromosome 2	Aspergillus flavus	34.2	34.2	95%	35	95.24%	6512809
CP051068.1	Aspergillus flavus strain Afla-Guard chromosome 2	Aspergillus flavus	34.2	34.2	95%	35	95.24%	6149748
CP051076.1	Aspergillus flavus strain K49 chromosome 2	Aspergillus flavus	34.2	34.2	95%	35	95.24%	6357944
CP051084.1	Aspergillus flavus strain K54A chromosome 2	Aspergillus flavus	34.2	34.2	95%	35	95.24%	5455511
CP051044.1	Aspergillus flavus strain Tox4 chromosome 2	Aspergillus flavus	34.2	34.2	95%	35	95.24%	6383110
CP051092.1	Aspergillus flavus strain VCG1 chromosome 2	Aspergillus flavus	34.2	34.2	95%	35	95.24%	6266380
CP051052.1	Aspergillus flavus strain VCG4 chromosome 2	Aspergillus flavus	34.2	34.2	95%	35	95.24%	6159567
CP051028.1	Aspergillus parasiticus strain NRRL 2999 chromosome 2	Aspergillus parasiticus	34.2	34.2	95%	35	95.24%	6231755
CP051010.1	Streptomyces sp. S1D4-11 chromosome, complete genome	Streptomyces sp. S1D4-11	34.2	34.2	77%	35	100.00%	12276515
CP050974.1	Streptomyces sp. RLB1-33 chromosome, complete genome	Streptomyces sp. RLB1-33	34.2	34.2	77%	35	100.00%	12127650
CP041650.2	Streptomyces sp. RLB1-8 chromosome, complete genome	Streptomyces sp. RLB1-8	34.2	34.2	77%	35	100.00%	11765395
XM_032393171.1	PREDICTED: Phoca vitulina centromere protein F (CENPF), mRNA	harbor seal	34.2	34.2	77%	35	100.00%	10367
CP041612.2	Streptomyces sp. S1A1-8 chromosome, complete genome	Streptomyces sp. S1A1-8	34.2	34.2	77%	35	100.00%	12037094
CP041607.2	Streptomyces sp. S1D4-14 chromosome, complete genome	Streptomyces sp. S1D4-14	34.2	34.2	77%	35	100.00%	11723552
CP041609.2	Streptomyces sp. S1D4-20 chromosome	Streptomyces sp. S1D4-20	34.2	34.2	77%	35	100.00%	11851273
CP041613.2	Streptomyces sp. S1D4-23 chromosome	Streptomyces sp. S1D4-23	34.2	34.2	77%	35	100.00%	12057750
CP041604.2	Streptomyces sp. S1A1-7 chromosome	Streptomyces sp. S1A1-7	34.2	34.2	77%	35	100.00%	11713216
CP041602.2	Streptomyces sp. RLB3-6 chromosome	Streptomyces sp. RLB3-6	34.2	34.2	77%	35	100.00%	12338170
CP041610.2	Streptomyces sp. RLB3-17 chromosome, complete genome	Streptomyces sp. RLB3-17	34.2	34.2	77%	35	100.00%	12022941
XM_032318161.1	PREDICTED: Mustela erminea centromere protein F (CENPF), transcript variant X4, mRNA	ermine	34.2	34.2	77%	35	100.00%	10189
XM_032318160.1	PREDICTED: Mustela erminea centromere protein F (CENPF), transcript variant X3, mRNA	ermine	34.2	34.2	77%	35	100.00%	10216
XM_032318159.1	PREDICTED: Mustela erminea centromere protein F (CENPF), transcript variant X2, mRNA	ermine	34.2	34.2	77%	35	100.00%	10368
XM_032318158.1	PREDICTED: Mustela erminea centromere protein F (CENPF), transcript variant X1, mRNA	ermine	34.2	34.2	77%	35	100.00%	10231
CP047558.1								
1/2 3/2 3/3 3/5 6/1 8/1 9/1 10/1	Solanum pinnatisectum cultivar CGN17745 chromosome 12	tansyleaf nightshade	34.2	64.4	77%	35	100.00%	61096591
LR738417.1	Lutra lutra genome assembly, chromosome: 15	Eurasian river otter	34.2	98.6	86%	35	100.00%	69992071
CP047250.1	Aspergillus flavus strain SU-16 chromosome 2	Aspergillus flavus	34.2	34.2	95%	35	95.24%	6246212
CP044622.1	Aspergillus flavus strain NRRL 3357 chromosome 2	Aspergillus flavus	34.2	34.2	95%	35	95.24%	6314854

XM_031040735.1	PREDICTED: Leptonychotes weddellii centromere protein F (CENPF), mRNA	Weddell seal	34.2	34.2	77%	35	100.00%	10315
CP035529.1	Aspergillus sojae strain SMF134 chromosome 2	Aspergillus sojae	34.2	34.2	95%	35	95.24%	6530912
LR699164.1	Geotrypetes seraphini genome assembly, chromosome: 19	Geotrypetes seraphini	34.2	34.2	77%	35	100.00%	41053462
LR697115.1 1/2 3/1 3/3 3/4 3/1	Chanos chanos genome assembly, chromosome: 10	milkfish	34.2	96.6	86%	35	100.00%	40347274
LR537134.1	Sparus aurata genome assembly, chromosome: 14	gilthead seabream	34.2	34.2	77%	35	100.00%	25861820
CP041651.1	Streptomyces sp. RLB3-5 chromosome	Streptomyces sp. RLB3-5	34.2	34.2	77%	35	100.00%	11898970
CP041611.1	Streptomyces sp. S1A1-3 chromosome	Streptomyces sp. S1A1-3	34.2	34.2	77%	35	100.00%	12042091
CP041654.1	Streptomyces sp. RLB1-9 chromosome, complete genome	Streptomyces sp. RLB1-9	34.2	34.2	77%	35	100.00%	11940408
XM_029457131.1	PREDICTED: Cottoperca gobio mesenchyme homeobox 1 (meox1), mRNA	Cottoperca gobio	34.2	34.2	77%	35	100.00%	1833
LR594556.1	Streptopelia turtur genome assembly, chromosome: 5	Streptopelia turtur	34.2	64.4	95%	35	100.00%	70854002
LR132013.2	Betta splendens genome assembly, chromosome: 6	Siamese fighting fish	34.2	34.2	77%	35	100.00%	19902465
XM_028096253.1	PREDICTED: Eumetopias jubatus centromere protein F-like (LOC114202864), partial mRNA	Steller sea lion	34.2	34.2	77%	35	100.00%	9034
XM_036099982.1	PREDICTED: Halichoerus grypus centromere protein F (CENPF), mRNA	gray seal	34.2	34.2	77%	35	100.00%	10101
CP059859.1	Aspergillus flavus strain AF13 chromosome 2	Aspergillus flavus	34.2	34.2	95%	35	95.24%	6263604
CP059867.1	Aspergillus flavus NRRL3357 chromosome 2	Aspergillus flavus NRRL3357	34.2	34.2	95%	35	95.24%	6246150
XM_027612786.2	PREDICTED: Zalophus californianus centromere protein F (CENPF), transcript variant X3, mRNA	California sea lion	34.2	34.2	77%	35	100.00%	10189
LR812059.1	Danio rerio genome assembly, chromosome: 22	zebrafish	34.2	34.2	77%	35	100.00%	39020267
LR812590.1	Danio rerio strain Nadia (NA) genome assembly, chromosome: 22	zebrafish	34.2	64.4	77%	35	100.00%	37117486
LR812565.1	Danio rerio strain Cooch Behar (CB) genome assembly, chromosome: 22	zebrafish	34.2	34.2	77%	35	100.00%	32929056
AP023164.1	Felis catus Senzu DNA, chromosome: D4, American Shorthair breed	domestic cat	34.2	124	86%	35	100.00%	96896805
XM_034761652.1	PREDICTED: Trachemys scripta elegans KAT8 regulatory NSL complex subunit 3 (KANSL3), transcript variant X6, mRNA	Trachemys scripta elegans	34.2	34.2	77%	35	100.00%	2471
XM_034761651.1	PREDICTED: Trachemys scripta elegans KAT8 regulatory NSL complex subunit 3 (KANSL3), transcript variant X5, mRNA	Trachemys scripta elegans	34.2	34.2	77%	35	100.00%	3847
XM_034761650.1	PREDICTED: Trachemys scripta elegans KAT8 regulatory NSL complex subunit 3 (KANSL3), transcript variant X4, mRNA	Trachemys scripta elegans	34.2	34.2	77%	35	100.00%	2645
XR_004644615.1	PREDICTED: Trachemys scripta elegans KAT8 regulatory NSL complex subunit 3 (KANSL3), transcript variant X3, misc_RNA	Trachemys scripta elegans	34.2	34.2	77%	35	100.00%	3648
XM_034761648.1	PREDICTED: Trachemys scripta elegans KAT8 regulatory NSL complex subunit 3 (KANSL3), transcript variant X2, mRNA	Trachemys scripta elegans	34.2	34.2	77%	35	100.00%	3033
XM_034761647.1	PREDICTED: Trachemys scripta elegans KAT8 regulatory NSL complex subunit 3 (KANSL3), transcript variant X1, mRNA	Trachemys scripta elegans	34.2	34.2	77%	35	100.00%	3111
LR792560.1	Pseudochaenichthys georgianus genome assembly, chromosome: 15	South Georgia icefish	34.2	64.4	86%	35	100.00%	39528010
CP050600.1 1/2 3/3 3/4 3/5 3/6 5/1 10/1	Canis lupus familiaris breed Labrador retriever chromosome 11a	dog	34.2	94.6	81%	35	100.00%	73584800

CP050630.1 1/2 3/3 3/4 3/5 3/6 5/1 10/1	Canis lupus familiaris breed Labrador retriever chromosome 11b	dog	34.2	94.6	81%	35	100.00%	73564375
XM_032790304.1	PREDICTED: Chelonoidis abingdonii KAT8 regulatory NSL complex subunit 3 (KANSL3), transcript variant X5, mRNA	Abingdon island giant tortoise	34.2	34.2	77%	35	100.00%	2950
XM_032790303.1	PREDICTED: Chelonoidis abingdonii KAT8 regulatory NSL complex subunit 3 (KANSL3), transcript variant X4, mRNA	Abingdon island giant tortoise	34.2	34.2	77%	35	100.00%	4335
XM_032790302.1	PREDICTED: Chelonoidis abingdonii KAT8 regulatory NSL complex subunit 3 (KANSL3), transcript variant X3, mRNA	Abingdon island giant tortoise	34.2	34.2	77%	35	100.00%	3102
XM_032790301.1	PREDICTED: Chelonoidis abingdonii KAT8 regulatory NSL complex subunit 3 (KANSL3), transcript variant X2, mRNA	Abingdon island giant tortoise	34.2	34.2	77%	35	100.00%	3656
XM_032790300.1	PREDICTED: Chelonoidis abingdonii KAT8 regulatory NSL complex subunit 3 (KANSL3), transcript variant X1, mRNA	Abingdon island giant tortoise	34.2	34.2	77%	35	100.00%	3734
LR778253.1 1/2 3/3 8/1 9/1	Coregonus sp. 'balchen' genome assembly, chromosome: 1	Coregonus sp. 'balchen'	34.2	34.2	77%	35	100.00%	93459789
XM_030551832.1	PREDICTED: Gopherus evgoodei KAT8 regulatory NSL complex subunit 3 (KANSL3), transcript variant X6, mRNA	Goodes thornscrub tortoise	34.2	34.2	77%	35	100.00%	2963
XM_030551831.1	PREDICTED: Gopherus evgoodei KAT8 regulatory NSL complex subunit 3 (KANSL3), transcript variant X5, mRNA	Goodes thornscrub tortoise	34.2	34.2	77%	35	100.00%	4104
XM_030551829.1	PREDICTED: Gopherus evgoodei KAT8 regulatory NSL complex subunit 3 (KANSL3), transcript variant X4, mRNA	Goodes thornscrub tortoise	34.2	34.2	77%	35	100.00%	3117
XM_030551828.1	PREDICTED: Gopherus evgoodei KAT8 regulatory NSL complex subunit 3 (KANSL3), transcript variant X3, mRNA	Goodes thornscrub tortoise	34.2	34.2	77%	35	100.00%	3671
XM_030551827.1	PREDICTED: Gopherus evgoodei KAT8 regulatory NSL complex subunit 3 (KANSL3), transcript variant X2, mRNA	Goodes thornscrub tortoise	34.2	34.2	77%	35	100.00%	3772
XM_030551826.1	PREDICTED: Gopherus evgoodei KAT8 regulatory NSL complex subunit 3 (KANSL3), transcript variant X1, mRNA	Goodes thornscrub tortoise	34.2	34.2	77%	35	100.00%	3749
XM_030394229.1	**PREDICTED: Sparus aurata cytohesin 2 (cyth2), transcript variant X2, mRNA**	gilthead seabream	34.2	34.2	77%	35	100.00%	4687
XM_030394228.1	**PREDICTED: Sparus aurata cytohesin 2 (cyth2), transcript variant X1, mRNA**	gilthead seabream	34.2	34.2	77%	35	100.00%	4697
LR537137.1	**Sparus aurata genome assembly, chromosome: 17**	gilthead seabream	34.2	64.4	77%	35	100.00%	37948182
LR597560.1	**Myripristis murdjan genome assembly, chromosome: 11**	pinecone soldierfish	34.2	34.2	77%	35	100.00%	31823041
LR597445.1	Salarias fasciatus genome assembly, chromosome: 10	jewelled blenny	34.2	98.6	77%	35	100.00%	28124721
LR535816.1	Denticeps clupeoides genome assembly, chromosome: 4	denticle herring	34.2	34.2	77%	35	100.00%	35364899
CP062029.1	Macrobrachium nipponense isolate FS-2020 chromosome 28	Macrobrachium nipponense	34.2	34.2	77%	35	100.00%	75602129
CP061805.1	Aspergillus flavus strain CA14 chromosome 2	Aspergillus flavus	34.2	34.2	95%	35	95.24%	6306092
LR877219.1	**Acomys russatus genome assembly, chromosome: 8**	golden spiny mouse	34.2	96.6	86%	35	100.00%	77196262
CP046697.1 1/2 3/2 3/3 3/5 6/1 8/1 9/1 10/1	Solanum tuberosum cultivar MSH/14-112 chromosome 12	potato	30.2	30.2	77%	540	94.12%	61134531

Pipistrellus pipistrellus genome assembly, chromosome: 11

GenBank: LR862367.1

FASTA Graphics

Go to:

LOCUS LR862367 18 bp DNA linear MAM **03-AUG-2020**

DEFINITION Pipistrellus pipistrellus genome assembly, chromosome: 11.

ACCESSION LR862367 REGION: 15305412..15305429

VERSION LR862367.1

DBLINK BioProject: PRJEB39564

BioSample: SAMEA994724

KEYWORDS .

SOURCE Pipistrellus pipistrellus (common pipistrelle)

ORGANISM Pipistrellus pipistrellus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Laurasiatheria; Chiroptera; Microchiroptera; Vespertilionidae; Pipistrellus.

REFERENCE 1

CONSRTM Wellcome Sanger Institute Data Sharing

TITLE Direct Submission

JOURNAL Submitted (27-JUL-2020) SC, Wellcome Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, United Kingdom

COMMENT The assembly mPipPip1.1 is based on 62x PacBio data, 34x 10X Genomics Chromium data, and 58x Arima Hi-C data generated at the Wellcome Sanger Institute. The Hi-C data was derived from a separate sample, mPipPip2. The assembly process included the following sequence of steps: initial PacBio assembly generation with Falcon-unzip, retained haplotig separation with purge_dups, 10X based scaffolding with scaff10x, Hi-C based scaffolding with SALSA2, Arrow polishing, two rounds of FreeBayes polishing. The mitochondrial assembly was produced at The Rockefeller University using mitoVGP. Finally, the assembly was analysed and manually improved using gEVAL. Chromosome-scale scaffolds confirmed by the Hi-C data have been named in order of size.

FEATURES Location/Qualifiers

source 1..18

/organism="**Pipistrellus pipistrellus**"

/mol_type="genomic DNA"

/db_xref="taxon:59474"

/chromosome="11"

ORIGIN

1 **ttcttgtgtg tcaggacc**

//

Pipistrellus pipistrellus genome assembly, chromosome: 11
Sequence ID: LR862367.1 **Length: 69386144 Number of Matches: 1**
Range 1: 15305412 to 15305429 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
36.2 bits(18)	8.9	18/18(100%)	0/18(0%)	Plus/Minus

```
Query  5         GGTCCTGACACACAAGAA  22
                 ||||||||||||||||||
Sbjct  15305429  GGTCCTGACACACAAGAA  15305412
```

ALIGNMENT VIEW FLAT QUERY-ANCHORED WITH LETTERS FOR IDENTITIES

Query range 1: 1 to 22

Query	**1**	**ATCGGGTCCTGACACACAAGAA**	**22**
CP034495.1	13466263	ATCGGGTCCTGACACACAA	13466245
CP034495.1	22049998	TCCTGACACACAAGA	22049984
CP034495.1	76750192	GGTCCTGACACACAA	76750178
CP034495.1	63678704	GGTCCTGACACACAA	63678718
AP023477.1	13650426	ATCGGGTCCTGACACACAA	13650408
AP023477.1	76248445	GGTCCTGACACACAA	76248431
AP023477.1	63152054	GGTCCTGACACACAA	63152068
AC005411.1	33193	ATCGGGTCCTGACACACAA	33175
LR584236.2	6619499	GGGTCCTGACACACAAGA	6619482
LR584433.1	30548470	ATCGGGTCCTGACACACA	30548453
CP025669.1	15580543	GGTCCTGACACACAAGAA	15580526
CP025666.1	10467775	GGTCCTGACACACAAGAA	10467758
LN590708.1	10032142	GGTCCTGACACACAAGAA	10032125
LN590708.1	13202056	TCCTGACACACAAGA	13202070
LR862367.1	15305429	GGTCCTGACACACAAGAA	15305412
XR_004133838.1	4	CGGGTCCTGACACACAAG	21
XM_031265003.1	133	GGTCCTGACACACAAGAA	150
LR633957.1	13512425	ATCGGGTCCTGACACACA	13512442
LR597469.1	79933437	CGGGTCCTGACACACAAG	79933454
LR597469.1	80104901	CGGGTCCTGACACACAAG	80104918
LR597469.1	65336798	TCCTGACACACAAGA	65336812
CP025664.1	9561951	GGTCCTGACACACAAGAA	9561968
XM_018226941.1	2179	GGGTCCTGACACACAAGA	2196
LK066360.1	175561	CGGGTCCTGACACACAAG	175578
LR812500.1	11950907	GGTCCTGACACACAAGA	11950891
LR812500.1	21562119	GGTCCTGACACACAAGA	21562103
LR812500.1	20641898	GGTCCTGACACACAAGA	20641914
LR812084.1	22351656	GTCCTGACACACAAGAA	22351640
LR812084.1	12278486	CCTGACACACAAGAA	12278472
XM_035010550.1	9541	GGGTCCTGACACACAAG	9525

CP051060.1	1764302	TCGGGTGCTGACACACAAGAA	1764282
CP051036.1	1808656	TCGGGTGCTGACACACAAGAA	1808636
CP051020.1	2070075	TCGGGTGCTGACACACAAGAA	2070055
CP051068.1	1763679	TCGGGTGCTGACACACAAGAA	1763659
CP051076.1	1905409	TCGGGTGCTGACACACAAGAA	1905389
CP051084.1	1826983	TCGGGTGCTGACACACAAGAA	1826963
CP051044.1	1474962	TCGGGTGCTGACACACAAGAA	1474942
CP051092.1	1840194	TCGGGTGCTGACACACAAGAA	1840174
CP051052.1	1741091	TCGGGTGCTGACACACAAGAA	1741071
CP051028.1	1806024	TCGGGTGCTGACACACAAGAA	1806004
CP051010.1	8085892	TCGGGTCCTGACACACA	8085876
CP050974.1	7984726	TCGGGTCCTGACACACA	7984710
CP041650.2	8366161	TCGGGTCCTGACACACA	8366145
XM_032393171.1	9671	GGGTCCTGACACACAAG	9655
CP041612.2	8160571	TCGGGTCCTGACACACA	8160555
CP041607.2	7937840	TCGGGTCCTGACACACA	7937824
CP041609.2	8065585	TCGGGTCCTGACACACA	8065569
CP041613.2	8368507	TCGGGTCCTGACACACA	8368491
CP041604.2	8289159	TCGGGTCCTGACACACA	8289143
CP041602.2	8020607	TCGGGTCCTGACACACA	8020591
CP041610.2	7993768	TCGGGTCCTGACACACA	7993752
XM_032318161.1	9488	GGGTCCTGACACACAAG	9472
XM_032318160.1	9515	GGGTCCTGACACACAAG	9499
XM_032318159.1	9667	GGGTCCTGACACACAAG	9651
XM_032318158.1	9530	GGGTCCTGACACACAAG	9514
CP047558.1	13102812	TCGGGTCCTGACACACA	13102796
CP047558.1	23959045	TCGGGTCCTGACACA	23959059
LR738417.1	15156877	GGGTCCTGACACACAAG	15156861
LR738417.1	58724401	GTCCTGACACACAAGAA	58724417
LR738417.1	61670400	GGGTCCTGACACACA	61670414
CP047250.1	1740771	TCGGGTGCTGACACACAAGAA	1740751
CP044622.1	1812873	TCGGGTGCTGACACACAAGAA	1812853
XM_031040735.1	9464	GGGTCCTGACACACAAG	9448
CP035529.1	1961507	TCGGGTGCTGACACACAAGAA	1961487
LR699164.1	34911047	GGGTCCTGACACACAAG	34911031
LR697115.1	32421543	TCGGGTCCTGACACACA	32421527
LR697115.1	27036679	GGTCCTGACACACAAG	27036664
LR697115.1	4290476	GGTCCTGACACACAA	4290462
LR537134.1	2921913	GGTCCTGACACACAAGA	2921897
CP041651.1	7902720	TCGGGTCCTGACACACA	7902704
CP041611.1	8165294	TCGGGTCCTGACACACA	8165278
CP041654.1	8164184	TCGGGTCCTGACACACA	8164168
XM_029457131.1	140	TCGGGTCCTGACACACA	124
LR594556.1	53940900	GGGTCCTGACAAACAAGAA	53940882
LR594556.1	1399214	TCGGGTCCTGACACACA	1399198
LR132013.2	14345330	CGGGTCCTGACACACAA	14345314
XM_028096253.1	8289	GGGTCCTGACACACAAG	8273
XM_036099982.1	9366	GGGTCCTGACACACAAG	9350
CP059859.1	1812250	TCGGGTGCTGACACACAAGAA	1812230
CP059867.1	1812830	TCGGGTGCTGACACACAAGAA	1812810
XM_027612786.2	9360	GGGTCCTGACACACAAG	9344
LR812059.1	22791268	GTCCTGACACACAAGAA	22791252
LR812590.1	17194139	GTCCTGACACACAAGAA	17194123
LR812590.1	13780417	CCTGACACACAAGAA	13780431
LR812565.1	18830476	GTCCTGACACACAAGAA	18830460
AP023164.1	71550743	GGTCCTGACACACAAGA	71550759
AP023164.1	43772683	GGTCCTGACACACAA	43772669
AP023164.1	76006964	CCTGACACACAAGAA	76006950
AP023164.1	89775056	GGGTCCTGACACACA	89775070
XM_034761652.1	1347	GGGTCCTGACACACAAG	1363

XM_034761651.1	1347	GGGTCCTGACACACAAG	1363
XM_034761650.1	1347	GGGTCCTGACACACAAG	1363
XR_004644615.1	1347	GGGTCCTGACACACAAG	1363
XM_034761648.1	1347	GGGTCCTGACACACAAG	1363
XM_034761647.1	1347	GGGTCCTGACACACAAG	1363
LR792560.1	5337604	CGGGTCCTGACACACAA	5337620
LR792560.1	25042884	TCCTGACACACAAGA	25042870
CP050600.1	63813944	GGTCCTGACACACAAGA	63813960
CP050600.1	6594778	GTCCTGACACACAAG	6594764
CP050600.1	7332222	GGGTCCTGACACACA	7332236
CP050630.1	63791832	GGTCCTGACACACAAGA	63791848
CP050630.1	6585343	GTCCTGACACACAAG	6585329
CP050630.1	7323174	GGGTCCTGACACACA	7323188
XM_032790304.1	1858	GGGTCCTGACACACAAG	1874
XM_032790303.1	1858	GGGTCCTGACACACAAG	1874
XM_032790302.1	1858	GGGTCCTGACACACAAG	1874
XM_032790301.1	1858	GGGTCCTGACACACAAG	1874
XM_032790300.1	1858	GGGTCCTGACACACAAG	1874
LR778253.1	74667960	GTCCTGACACACAAGAA	74667976
XM_030551832.1	1874	GGGTCCTGACACACAAG	1890
XM_030551831.1	1874	GGGTCCTGACACACAAG	1890
XM_030551829.1	1874	GGGTCCTGACACACAAG	1890
XM_030551828.1	1874	GGGTCCTGACACACAAG	1890
XM_030551827.1	1897	GGGTCCTGACACACAAG	1913
XM_030551826.1	1874	GGGTCCTGACACACAAG	1890
XM_030394229.1	1874	GGGTCCTGACACACAAG	1890
XM_030394228.1	1884	GGGTCCTGACACACAAG	1900
LR537137.1	23280262	GGGTCCTGACACACAAG	23280278
LR537137.1	15586339	GTCCTGACACACAAG	15586325
LR597560.1	7549408	CGGGTCCTGACACACAA	7549424
LR597445.1	4090641	GTCCTGACACACAAGAA	4090657
LR597445.1	21828624	TCCTGACACACAAGAA	21828609
LR597445.1	24913684	TCCTGACACACAAGAA	24913699
LR535816.1	2764452	GGGTCCTGACACACAAG	2764468
CP062029.1	57872644	CGGGTCCTGACACACAA	57872660
CP061805.1	4502060	TCGGGTGCTGACACACAAGAA	4502080
LR877219.1	73395407	GGGTGCTGACACACAAGAA	73395389
LR877219.1	67213109	GGGTCCTGACACACAAG	67213125
LR877219.1	71189442	GTCCTGACACACAAGA	71189427
CP046697.1	13106661	TCGGGTCCTGAYACACA	13106645

PRODUCTS OF SIGNIFICANT ALIGNMENTS (CDS FEATURE)

- Zinc finger BED domain-containing protein RICESLEEPER 2-like

https://www.ncbi.nlm.nih.gov/nucleotide/XM_031265003.1?report=genbank&log$=nuclalign&blast_rank=11&RID=W6XMJ10P013&from=133&to=150

- KAT8 regulatory NSL complex subunit 3

https://www.ncbi.nlm.nih.gov/nucleotide/XM_030551832.1?report=genbank&log$=nuclalign&blast_rank=84&RID=W6XMJ10P013&from=1874&to=1890
https://www.ncbi.nlm.nih.gov/nucleotide/XM_034761652.1?report=genbank&log$=nuclalign&blast_rank=69&RID=W6XMJ10P013&from=1347&to=1363
https://www.ncbi.nlm.nih.gov/nucleotide/XM_032790304.1?report=genbank&log$=nuclalign&blast_rank=78&RID=W6XMJ10P013&from=1858&to=1874
https://www.ncbi.nlm.nih.gov/nucleotide/XM_030551831.1?report=genbank&log$=nuclalign&blast_rank=85&RID=W6XMJ10P013&from=1874&to=1890
https://www.ncbi.nlm.nih.gov/nucleotide/XM_034761651.1?report=genbank&log$=nuclalign&blast_rank=70&RID=W6XMJ10P013&from=1347&to=1363
https://www.ncbi.nlm.nih.gov/nucleotide/XM_032790303.1?report=genbank&log$=nuclalign&blast_rank=79&RID=W6XMJ10P013&from=1858&to=1874
https://www.ncbi.nlm.nih.gov/nucleotide/XM_030551829.1?report=genbank&log$=nuclalign&blast_rank=86&RID=W6XMJ10P013&from=1874&to=1890
https://www.ncbi.nlm.nih.gov/nucleotide/XM_034761650.1?report=genbank&log$=nuclalign&blast_rank=71&RID=W6XMJ10P013&from=1347&to=1363
https://www.ncbi.nlm.nih.gov/nucleotide/XM_032790302.1?report=genbank&log$=nuclalign&blast_rank=80&RID=W6XMJ10P013&from=1858&to=1874
https://www.ncbi.nlm.nih.gov/nucleotide/XM_030551828.1?report=genbank&log$=nuclalign&blast_rank=87&RID=W6XMJ10P013&from=1874&to=1890
https://www.ncbi.nlm.nih.gov/nucleotide/XM_034761648.1?report=genbank&log$=nuclalign&blast_rank=73&RID=W6XMJ10P013&from=1347&to=1363
https://www.ncbi.nlm.nih.gov/nucleotide/XM_032790301.1?report=genbank&log$=nuclalign&blast_rank=81&RID=W6XMJ10P013&from=1858&to=1874
https://www.ncbi.nlm.nih.gov/nucleotide/XM_030551827.1?report=genbank&log$=nuclalign&blast_rank=88&RID=W6XMJ10P013&from=1897&to=1913
https://www.ncbi.nlm.nih.gov/nucleotide/XM_034761647.1?report=genbank&log$=nuclalign&blast_rank=74&RID=W6XMJ10P013&from=1347&to=1363
https://www.ncbi.nlm.nih.gov/nucleotide/XM_032790300.1?report=genbank&log$=nuclalign&blast_rank=82&RID=W6XMJ10P013&from=1858&to=1874
https://www.ncbi.nlm.nih.gov/nucleotide/XM_030551826.1?report=genbank&log$=nuclalign&blast_rank=89&RID=W6XMJ10P013&from=1874&to=1890

SEQUENCE N° 3/2

(T=2; C=0; G=3; A=1)

SEQUENCE N° 3/2 (T=2; C=0; G=3; A=1)

Original Sequence	**G**	**T**	**G**	**A**	**A**	**A**	**T**	**G**	**G**	**T**	**C**	**A**	**T**	**G**	**T**	**G**	**T**	**G**	**G**	**C**	**G**	**G**
Trend n° 3 (T=2; C=0; G=3; A=1)	**3**	**2**	**3**	**1**	**1**	**1**	**2**	**3**	**3**	**2**	**0**	**1**	**2**	**3**	**2**	**3**	**2**	**3**	**3**	**0**	**3**	**3**
Trend n° 3 (I i D d)	**-**	**D**	**I**	**D**	**d**	**d**	**I**	**I**	**i**	**D**	**D**	**I**	**I**	**I**	**D**	**I**	**D**	**I**	**i**	**D**	**I**	**i**
Trend n° 3/2	**2**	**1**	**2**	**0**	**0**	**0**	**1**	**2**	**2**	**1**	**0**	**1**	**2**	**3**	**0**	**2**	**0**	**1**	**1**	**0**	**1**	**1**
Sequence n° 3/2 (77,27%)	**T**	**A**	**T**	**C**	**C**	**C**	**A**	**T**	**T**	**A**	**C**	**A**	**T**	**G**	**C**	**T**	**C**	**A**	**A**	**C**	**A**	**A**

Query: **TATCCCATTACATGCTCAACAA**

Search Parameters	
Program	blastn
Word size	7
Expect value	1000
Hitlist size	100
Match/Mismatch scores	1,-3
Gapcosts	5,2
Filter string	F
Genetic Code	1

100 sequences selected:

Accession Seq	Description	Comomm Name	Max Score	Total Score	Query C.	E value	Per. Id.	Acc. Len
CP023962.1	Yersinia frederiksenii strain FDAARGOS_417 chromosome, complete genome	Yersinia frederiksenii	40.1	40.1	90%	0.57	100.00%	5000678
LR597563.1	Myripristis murdjan genome assembly, chromosome: 14	pinecone soldierfish	38.2	38.2	86%	2.2	100.00%	39034921
CP032589.1	Lateolabrax maculatus linkage group 15 sequence	spotted sea bass	38.2	38.2	86%	2.2	100.00%	23369250
LR738416.1	Lutra lutra genome assembly, chromosome: 14	Eurasian river otter	38.2	68.4	90%	2.2	100.00%	89080780
XM_030069313.1	PREDICTED: Myripristis murdjan adhesion G-protein coupled receptor G2-like (LOC115371788), transcript variant X2, mRNA	pinecone soldierfish	38.2	38.2	86%	2.2	100.00%	3419
XM_030069312.1	PREDICTED: Myripristis murdjan adhesion G-protein coupled receptor G2-like (LOC115371788), transcript variant X1, mRNA	pinecone soldierfish	38.2	38.2	86%	2.2	100.00%	4348
CP027273.1	Lateolabrax maculatus chromosome Lm12	spotted sea bass	38.2	38.2	86%	2.2	100.00%	23523986
KX538963.1	Zea mays ABI3 mRNA, complete cds	Zea mays	38.2	38.2	86%	2.2	100.00%	1026
XM_035959857.1	PREDICTED: Zea mays B3 domain-containing protein IDEF1 (LOC103628923), transcript variant X4, mRNA	Zea mays	38.2	38.2	86%	2.2	100.00%	1726
XM_020539753.3	PREDICTED: Zea mays B3 domain-containing protein IDEF1 (LOC103628923), transcript variant X3, mRNA	Zea mays	38.2	38.2	86%	2.2	100.00%	1802
XM_020539752.3	PREDICTED: Zea mays B3 domain-containing protein IDEF1 (LOC103628923), transcript variant X2, mRNA	Zea mays	38.2	38.2	86%	2.2	100.00%	1784

XM_020539751.2	PREDICTED: Zea mays B3 domain-containing protein IDEF1 (LOC103628923), transcript variant X1, mRNA	Zea mays	38.2	38.2	86%	2.2	100.00%	1862
EU971515.1	Zea mays clone 366455 mRNA sequence	Zea mays	38.2	38.2	86%	2.2	100.00%	1523
CP032675.1	Rhodococcus rhodochrous strain ATCC BAA870 chromosome, complete genome	Rhodococcus rhodochrous	36.2	36.2	81%	8.9	100.00%	5370536
CP032582.1	Gossypium turneri isolate D10-2 chromosome D10_12	Gossypium turneri	36.2	66.4	95%	8.9	100.00%	54556719
CP020623.1	Oryzias latipes strain HSOK chromosome 3	Japanese medaka	36.2	36.2	81%	8.9	100.00%	39097270
CP020781.1	Oryzias latipes strain HNI chromosome 3	Japanese medaka	36.2	66.4	81%	8.9	100.00%	35737580
CP020667.1	Oryzias latipes strain Hd-rR chromosome 3 sequence	Japanese medaka	36.2	66.4	81%	8.9	100.00%	38248663
HF933209.1	TPA: Oryzias latipes strain Hd-rR, complete genome assembly, chromosome 3	Japanese medaka	36.2	66.4	81%	8.9	100.00%	36623554
XM_009376990.1	PREDICTED: Pyrus x bretschneideri 50S ribosomal protein L7/L12-like (LOC103964099), mRNA	Chinese white pear	36.2	36.2	81%	8.9	100.00%	1128
LR880663.1	Poecilia reticulata genome assembly, chromosome: 19	guppy	36.2	36.2	81%	8.9	100.00%	25553809
AC087190.5 **1/2 2/2 3/3 5/1 6/1 10/1**	Homo sapiens chromosome 16 clone RP11-473I1, complete sequence	human	36.2	36.2	81%	8.9	100.00%	159712
LR812129.1 **1/2 5/1**	Erithacus rubecula genome assembly, chromosome: 26	European robin	36.2	66.4	81%	8.9	100.00%	6516792
AP022681.1	Epinephelus fuscoguttatus DNA, LG7, complete sequence	brown-marbled grouper	36.2	36.2	81%	8.9	100.00%	47170333
CP032564.1	Gossypium raimondii isolate D5-4 chromosome D5_12	Gossypium raimondii	36.2	66.4	95%	8.9	100.00%	57421043
LR584085.1	Scleropages formosus genome assembly, chromosome: 20	Asian bonytongue	36.2	126	86%	8.9	100.00%	25884345
LR584069.1	Scleropages formosus genome assembly, chromosome: 4	Asian bonytongue	36.2	98.6	95%	8.9	100.00%	39064107
CP034494.1	Eukaryotic synthetic construct chromosome 16	eukaryotic synthetic construct	36.2	96.6	95%	8.9	100.00%	98200793
XM_020714013.2	PREDICTED: Oryzias latipes TEA domain transcription factor 1 (tead1), transcript variant X8, mRNA	Japanese medaka	36.2	36.2	81%	8.9	100.00%	1446
XM_023953470.1	PREDICTED: Oryzias latipes TEA domain transcription factor 1 (tead1), transcript variant X7, mRNA	Japanese medaka	36.2	36.2	81%	8.9	100.00%	1458
XM_011489176.3	PREDICTED: Oryzias latipes TEA domain transcription factor 1 (tead1), transcript variant X6, mRNA	Japanese medaka	36.2	36.2	81%	8.9	100.00%	1738
XM_023953469.1	PREDICTED: Oryzias latipes TEA domain transcription factor 1 (tead1), transcript variant X5, mRNA	Japanese medaka	36.2	36.2	81%	8.9	100.00%	1750
XM_004066996.4	PREDICTED: Oryzias latipes TEA domain transcription factor 1 (tead1), transcript variant X4, mRNA	Japanese medaka	36.2	36.2	81%	8.9	100.00%	1750
XM_023953468.1	PREDICTED: Oryzias latipes TEA domain transcription factor 1 (tead1), transcript variant X3, mRNA	Japanese medaka	36.2	36.2	81%	8.9	100.00%	1802
XM_023953467.1	PREDICTED: Oryzias latipes TEA domain transcription factor 1 (tead1), transcript variant X2, mRNA	Japanese medaka	36.2	36.2	81%	8.9	100.00%	4586
XM_023953466.1	PREDICTED: Oryzias latipes TEA domain transcription factor 1 (tead1), transcript variant X1, mRNA	Japanese medaka	36.2	36.2	81%	8.9	100.00%	1792
XM_016879281.1	PREDICTED: Gossypium hirsutum transcription factor bHLH25-like (LOC107945325), mRNA	cotton	36.2	36.2	81%	8.9	100.00%	1309
XM_012581734.1	PREDICTED: Gossypium raimondii transcription factor	Gossypium raimondii	36.2	36.2	81%	8.9	100.00%	1622

	bHLH19-like (LOC105763496), mRNA								
CP009355.1	Vibrio tubiashii ATCC 19109 chromosome 2, complete sequence	Vibrio tubiashii ATCC 19109	36.2	36.2	81%	8.9	100.00%	1766582
CP061205.1	Kordiimonas pumila strain N18 chromosome, complete genome	Kordiimonas pumila	36.2	36.2	81%	8.9	100.00%	4041658
CP027557.1	Rhodococcus rhodochrous strain BX2 chromosome, complete genome	Rhodococcus rhodochrous	36.2	36.2	81%	8.9	100.00%	5459556
XM_036210237.1	PREDICTED: Oryzias melastigma TEA domain family member 1b (tead1b), transcript variant X4, mRNA	Indian medaka	36.2	36.2	81%	8.9	100.00%	2041
XM_024295636.2	PREDICTED: Oryzias melastigma TEA domain family member 1b (tead1b), transcript variant X3, mRNA	Indian medaka	36.2	36.2	81%	8.9	100.00%	2053
XM_024295635.2	PREDICTED: Oryzias melastigma TEA domain family member 1b (tead1b), transcript variant X2, mRNA	Indian medaka	36.2	36.2	81%	8.9	100.00%	2106
XM_024295634.2	PREDICTED: Oryzias melastigma TEA domain family member 1b (tead1b), transcript variant X1, mRNA	Indian medaka	36.2	36.2	81%	8.9	100.00%	4885
AP023476.1								
1/2 2/2 3/2 3/3 6/1 10/1	Homo sapiens DNA, chromosome 16, nearly complete genome	human	36.2	96.6	95%	8.9	100.00%	94690957
LR812108.1 **1/2 5/1**	Erithacus rubecula genome assembly, chromosome: 6	European robin	34.2	94.6	86%	35	100.00%	60684490
LR812106.1 **1/2 5/1**	Erithacus rubecula genome assembly, chromosome: 4	European robin	34.2	64.4	95%	35	95.24%	68599267
XM_034203981.1	PREDICTED: Gymnodraco acuticeps sacsin-like (LOC117538367), mRNA	Gymnodraco acuticeps	34.2	34.2	77%	35	100.00%	9925
XM_034203980.1	PREDICTED: Gymnodraco acuticeps sacsin-like (LOC117538366), mRNA	Gymnodraco acuticeps	34.2	34.2	77%	35	100.00%	11751
XM_034203979.1	PREDICTED: Gymnodraco acuticeps sacsin (si:dkeyp-118h9.7), transcript variant X2, mRNA	Gymnodraco acuticeps	34.2	34.2	77%	35	100.00%	14578
XM_034203978.1	PREDICTED: Gymnodraco acuticeps sacsin (si:dkeyp-118h9.7), transcript variant X1, mRNA	Gymnodraco acuticeps	34.2	34.2	77%	35	100.00%	14704
XM_034134622.1	PREDICTED: Trematomus bernacchii sacsin-like (LOC117485916), mRNA	emerald rockcod	34.2	34.2	77%	35	100.00%	5584
XM_034134621.1	PREDICTED: Trematomus bernacchii sacsin-like (LOC117485915), mRNA	emerald rockcod	34.2	34.2	77%	35	100.00%	12460
XM_034134161.1	PREDICTED: Trematomus bernacchii sacsin-like (LOC117485571), mRNA	emerald rockcod	34.2	34.2	77%	35	100.00%	14753
XM_034110707.1	PREDICTED: Pseudochaenichthys georgianus sacsin (si:dkeyp-118h9.7), mRNA	South Georgia icefish	34.2	34.2	77%	35	100.00%	14745
LR792566.1	Pseudochaenichthys georgianus genome assembly, chromosome: 21	South Georgia icefish	34.2	34.2	77%	35	100.00%	39591329
CP048847.1	Crassostrea gigas strain QD chromosome 9	Pacific oyster	34.2	64.4	77%	35	100.00%	55104829
AP022682.1	Epinephelus fuscoguttatus DNA, LG8, complete sequence	brown-marbled grouper	34.2	34.2	77%	35	100.00%	47003518
XM_031810749.1	PREDICTED: Oncorhynchus kisutch transmembrane protein 71 (tmem71), transcript variant X6, mRNA	coho salmon	34.2	34.2	77%	35	100.00%	3513
XM_031810748.1	PREDICTED: Oncorhynchus kisutch transmembrane protein 71 (tmem71), transcript variant X5, mRNA	coho salmon	34.2	34.2	77%	35	100.00%	3709
XM_020467159.2	PREDICTED: Oncorhynchus kisutch transmembrane protein 71 (tmem71), transcript variant X4, mRNA	coho salmon	34.2	34.2	77%	35	100.00%	3638

XM_020467158.2	PREDICTED: Oncorhynchus kisutch transmembrane protein 71 (tmem71), transcript variant X3, mRNA	coho salmon	34.2	34.2	77%	35	100.00%	3627
XM_020467155.2	PREDICTED: Oncorhynchus kisutch transmembrane protein 71 (tmem71), transcript variant X2, mRNA	coho salmon	34.2	34.2	77%	35	100.00%	3523
XM_020467160.2	PREDICTED: Oncorhynchus kisutch transmembrane protein 71 (tmem71), transcript variant X1, mRNA	coho salmon	34.2	34.2	77%	35	100.00%	3738
CP032578.1	Gossypium turneri isolate D10-2 chromosome D10_08	Gossypium turneri	34.2	64.4	90%	35	100.00%	61617684
LR633960.1 1/2 2/2 3/2	Gadus morhua genome assembly, chromosome: 18	Atlantic cod	34.2	68.4	77%	35	100.00%	24902675
LR633953.1 1/2 2/2 3/2	Gadus morhua genome assembly, chromosome: 11	Atlantic cod	34.2	34.2	77%	35	100.00%	30713045
LR597560.1	Myripristis murdjan genome assembly, chromosome: 11	pinecone soldierfish	34.2	34.2	95%	35	95.24%	31823041
XM_029459375.1	PREDICTED: Cottoperca gobio sacsin-like (LOC115026523), mRNA	Cottoperca gobio	34.2	34.2	77%	35	100.00%	10623
CP032560.1	Gossypium raimondii isolate D5-4 chromosome D5_08	Gossypium raimondii	34.2	34.2	77%	35	100.00%	61765408
LR584231.2	Takifugu rubripes genome assembly, chromosome: 1	torafugu	34.2	34.2	77%	35	100.00%	29232231
CP028923.1	Roseivirga pacifica strain 9dcg1 chromosome, complete genome	Fabibacter pacificus	34.2	34.2	77%	35	100.00%	4661901
CP041165.1	Sulfurimonas sp. B2 chromosome, complete genome	Sulfurimonas sp. B2	34.2	34.2	77%	35	100.00%	2261034
CP060294.1	Pedobacter sp. PAMC26386 chromosome, complete genome	Pedobacter sp. PAMC26386	34.2	34.2	77%	35	100.00%	4796773
LR877197.1	Onychomys torridus genome assembly, chromosome: 10	southern grasshopper mouse	34.2	185	95%	35	100.00%	98623089
LR862394.1	Arvicola amphibius genome assembly, chromosome: 14	Eurasian water mole	34.2	34.2	77%	35	100.00%	63161238
XM_034535788.1	PREDICTED: Cyclopterus lumpus E3 ubiquitin-protein ligase TRIM21-like (LOC117732666), transcript variant X6, mRNA	lumpfish	34.2	34.2	95%	35	95.24%	3388
XM_034535769.1	PREDICTED: Cyclopterus lumpus E3 ubiquitin-protein ligase TRIM21-like (LOC117732666), transcript variant X4, mRNA	lumpfish	34.2	34.2	95%	35	95.24%	3566
LR792831.1	Digitaria exilis annotation	Digitaria exilis	34.2	34.2	77%	35	100.00%	27194159
LR761616.1	Digitaria exilis genome assembly, chromosome: 6B	Digitaria exilis	34.2	34.2	77%	35	100.00%	27194159
CP047344.1	Proteus vulgaris strain ZN3 chromosome, complete genome	Proteus vulgaris	34.2	64.4	81%	35	100.00%	4310124
AP019652.1	Vibrio taketomensis DNA, chromosome 2, complete sequence	Vibrio taketomensis	34.2	34.2	77%	35	100.00%	1343375
AP019650.1	Vibrio taketomensis DNA, chromosome 2, complete sequence	Vibrio taketomensis	34.2	34.2	77%	35	100.00%	1341331
LR738604.1	Sciurus carolinensis genome assembly, chromosome: 14	gray squirrel	34.2	34.2	95%	35	95.24%	88654151
AP018694.1	Prolixibacteraceae bacterium MeG22 DNA, complete genome	Prolixibacteraceae bacterium MeG22	34.2	34.2	77%	35	100.00%	5712279
CP045339.1	Vibrio sp. THAF190c plasmid pTHAF190c_a, complete sequence	Vibrio sp. THAF190c	34.2	34.2	77%	35	100.00%	1722714
LR698965.1	Planococcus massiliensis isolate MGYG-HGUT-01482 genome assembly, chromosome: 1	Planococcus massiliensis	34.2	34.2	77%	35	100.00%	3357017
LR633950.1 1/2 2/2 3/2	Gadus morhua genome assembly, chromosome: 8	Atlantic cod	34.2	100	86%	35	100.00%	29710654
LR606197.1 1/2 3/3 3/5 9/1	Aquila chrysaetos chrysaetos genome assembly, chromosome: 17	Aquila chrysaetos chrysaetos	34.2	34.2	95%	35	95.24%	29699254
LR597448.1	Salarias fasciatus genome assembly, chromosome: 13	jewelled blenny	34.2	34.2	77%	35	100.00%	29904995
AP019297.1	Prunus dulcis DNA, pseudomolecule Pd01	almond	34.2	94.6	86%	35	100.00%	43698516
CP064939.1	Pedobacter sp. JBR3-12 chromosome, complete genome	Pedobacter sp. JBR3-12	34.2	34.2	77%	35	100.00%	5426070

LR813497.2	Saccharomyces cerevisiae strain HN1 genome assembly, chromosome: 15	baker's yeast	34.2	34.2	77%	35	100.00%	1079300
LR884475.1	Acanthopagrus latus isolate v.2019 genome assembly, chromosome: 16	yellowfin seabream	34.2	34.2	77%	35	100.00%	25616698
CP051488.1	Kosakonia sp. SMBL-WEM22 chromosome, complete genome	Kosakonia sp. SMBL-WEM22	34.2	34.2	77%	35	100.00%	5037617
LR862150.1	Ananas comosus var. bracteatus genome assembly, chromosome: 22	Ananas comosus var. bracteatus	34.2	34.2	77%	35	100.00%	17269086
AP022274.1	Enterobacter cloacae DNA, complete genome, strain: WP8-W19-CRE-02	Enterobacter cloacae	34.2	34.2	77%	35	100.00%	4891055
LR812551.1	Danio rerio strain Cooch Behar (CB) genome assembly, chromosome: 8	zebrafish	34.2	66.4	81%	35	100.00%	53882377
CP046702.1								
1/2 2/2 3/3 3/5 6/1 8/1 9/1 10/1	Solanum tuberosum cultivar MSH/14-112 chromosome 1	potato	30.2	60.5	90%	547	100.00%	88648484

ALIGNMENT VIEW FLAT QUERY-ANCHORED WITH LETTERS FOR IDENTITIES

Query range 1: 1 to 22

Query	1	TATCCCATTACATGCTCAACAA	22
CP023962.1	1745724	TCCCATTACATGCTCAACAA	1745705
LR597563.1	17250998	TATCCCATTACATGCTCAA	17250980
CP032589.1	13422761	TATCCCATTACATGCTCAA	13422743
LR738416.1	24420922	TCCCATTACATGCTCAACA	24420940
LR738416.1	15575990	ATCCCATTACATGCT	15575976
XM_030069313.1	2276	TATCCCATTACATGCTCAA	2294
XM_030069312.1	3205	TATCCCATTACATGCTCAA	3223
CP027273.1	2644037	TATCCCATTACATGCTCAA	2644055
KX538963.1	329	CCCATTACATGCTCAACAA	347
XM_035959857.1	735	CCCATTACATGCTCAACAA	753
XM_020539753.3	730	CCCATTACATGCTCAACAA	748
XM_020539752.3	723	CCCATTACATGCTCAACAA	741
XM_020539751.2	742	CCCATTACATGCTCAACAA	760
EU971515.1	457	CCCATTACATGCTCAACAA	475
CP032675.1	1712571	CCCATTACATGCTCAACA	1712554
CP032582.1	16487747	ATCCCATTACATGCTCAA	16487730
CP032582.1	9499005	TTACATGCTCAACAA	9499019
CP020623.1	19671468	CCATTACATGCTCAACAA	19671451
CP020781.1	18753876	CCATTACATGCTCAACAA	18753859
CP020781.1	16681984	TTACATGCTCAACAA	16681998
CP020667.1	20228307	CCATTACATGCTCAACAA	20228290
CP020667.1	18168716	TTACATGCTCAACAA	18168730
HF933209.1	18717783	CCATTACATGCTCAACAA	18717766
HF933209.1	16656976	TTACATGCTCAACAA	16656990
XM_009376990.1	355	CCCATTACATGCTCAACA	338
LR880663.1	20588898	CCCATTACATGCTCAACA	20588881
AC087190.5	95049	TATCCCATTACATGCTCA	95032
LR812129.1	6141125	TATCCCATTACATGCTCA	6141142
LR812129.1	4095966	TCCCATTACATGCTC	4095980
AP022681.1	42443338	CCATTACATGCTCAACAA	42443355
CP032564.1	37529827	ATCCCATTACATGCTCAA	37529844

CP032564.1	44871891	TTACATGCTCAACAA	44871877
LR584085.1	17507700	TATCCCATTACATGCTCA	17507717
LR584085.1	8204435	CCATTACATGCTCAA	8204421
LR584085.1	14388452	TCCCATTACATGCTC	14388466
LR584085.1	14575511	TCCCATTACATGCTC	14575525
LR584069.1	15589628	ATCCCATTACATGCTAAACA	15589609
LR584069.1	36872323	CCATTACATGCTCAACAA	36872340
LR584069.1	24491114	TCCCATTACATGCTC	24491100
CP034494.1	52008196	TCCCATGACATGCTCAACA	52008178
CP034494.1	9113208	TATCCCATTACATGCTCA	9113225
CP034494.1	66183523	CCATTACATGCTCAA	66183537
XM_020714013.2	1121	CCATTACATGCTCAACAA	1138
XM_023953470.1	1133	CCATTACATGCTCAACAA	1150
XM_011489176.3	1413	CCATTACATGCTCAACAA	1430
XM_023953469.1	1425	CCATTACATGCTCAACAA	1442
XM_004066996.4	1425	CCATTACATGCTCAACAA	1442
XM_023953468.1	1477	CCATTACATGCTCAACAA	1494
XM_023953467.1	1488	CCATTACATGCTCAACAA	1505
XM_023953466.1	1467	CCATTACATGCTCAACAA	1484
XM_016879281.1	349	ATCCCATTACATGCTCAA	366
XM_012581734.1	662	ATCCCATTACATGCTCAA	679
CP009355.1	1269741	CCATTACATGCTCAACAA	1269758
CP061205.1	3194946	CCATTACATGCTCAACAA	3194963
CP027557.1	4946095	CCCATTACATGCTCAACA	4946112
XM_036210237.1	1692	CCATTACATGCTCAACAA	1709
XM_024295636.2	1704	CCATTACATGCTCAACAA	1721
XM_024295635.2	1757	CCATTACATGCTCAACAA	1774
XM_024295634.2	1766	CCATTACATGCTCAACAA	1783
AP023476.1	56430623	TCCCATGACATGCTCAACA	56430605
AP023476.1	9156872	TATCCCATTACATGCTCA	9156889
AP023476.1	70607256	CCATTACATGCTCAA	70607270
LR812108.1	14024173	TATCCCATTACATGCTC	14024157
LR812108.1	55816198	CCATTACATGCTCAA	55816184
LR812108.1	25243419	ATCCCATTACATGCT	25243433
LR812106.1	17540369	ATCCCATTACTTGCTCAACAA	17540349
LR812106.1	1724166	TTACATGCTCAACAA	1724152
XM_034203981.1	1725	CATTACATGCTCAACAA	1709
XM_034203980.1	3355	CATTACATGCTCAACAA	3339
XM_034203979.1	3340	CATTACATGCTCAACAA	3324
XM_034203978.1	3340	CATTACATGCTCAACAA	3324
XM_034134622.1	3322	CATTACATGCTCAACAA	3306
XM_034134621.1	257	CATTACATGCTCAACAA	241
XM_034134161.1	3353	CATTACATGCTCAACAA	3337
XM_034110707.1	3324	CATTACATGCTCAACAA	3308
LR792566.1	22736379	CATTACATGCTCAACAA	22736363
CP048847.1	30598354	TCCCATTACATGCTCAA	30598338
CP048847.1	34594943	CCATTACATGCTCAA	34594929
AP022682.1	44113831	CATTACATGCTCAACAA	44113815
XM_031810749.1	2203	CCATTACATGCTCAACA	2187
XM_031810748.1	2399	CCATTACATGCTCAACA	2383
XM_020467159.2	2328	CCATTACATGCTCAACA	2312
XM_020467158.2	2317	CCATTACATGCTCAACA	2301
XM_020467155.2	2213	CCATTACATGCTCAACA	2197
XM_020467160.2	2428	CCATTACATGCTCAACA	2412
CP032578.1	31033463	TCCCATTACATGCTCAA	31033447
CP032578.1	53587352	TTACATGCTCAACAA	53587338
LR633960.1	22633765	TCCCATTACATGCTCAA	22633749
LR633960.1	22634484	TCCCATTACATGCTCAA	22634468
LR633953.1	19265064	CCATTACATGCTCAACA	19265048
LR597560.1	14470542	ATCCCATTACAAGCTCAACAA	14470522

XM_029459375.1	135	CATTACATGCTCAACAA	119
CP032560.1	23750225	TCCCATTACATGCTCAA	23750209
LR584231.2	1752792	CATTACATGCTCAACAA	1752776
CP028923.1	2912772	CATTACATGCTCAACAA	2912756
CP041165.1	2188370	TCCCATTACATGCTCAA	2188354
CP060294.1	4655806	CCATTACATGCTCAACA	4655790
LR877197.1	75781442	CATTACATGCTCAACAA	75781426
LR877197.1	55851315	ATCCCATTACATGCT	55851301
LR877197.1	78953222	ATCCCATTACATGCT	78953208
LR877197.1	45008148	CATTACATGCTCAAC	45008162
LR877197.1	50164947	ATCCCATTACATGCT	50164961
LR877197.1	76807168	ATCCCATTACATGCT	76807182
LR862394.1	2205206	TCCCATTACATGCTCAA	2205190
XM_034535788.1	2631	TATCCCATTACATGGTCAACA	2651
XM_034535769.1	2828	TATCCCATTACATGGTCAACA	2848
LR792831.1	16950994	CATTACATGCTCAACAA	16951010
LR761616.1	16950994	CATTACATGCTCAACAA	16951010
CP047344.1	2305404	CCATTACATGCTCAACA	2305420
CP047344.1	2022003	TTACATGCTCAACAA	2022017
AP019652.1	979168	CATTACATGCTCAACAA	979184
AP019650.1	977013	CATTACATGCTCAACAA	977029
LR738604.1	43729402	TATCCCATTACATGTTCAACA	43729422
AP018694.1	1891606	CCCATTACATGCTCAAC	1891622
CP045339.1	404180	CATTACATGCTCAACAA	404196
LR698965.1	1848004	TCCCATTACATGCTCAA	1848020
LR633950.1	27856735	TCCCATTACATGCTCAA	27856751
LR633950.1	27857454	TCCCATTACATGCTCAA	27857470
LR633950.1	10352973	TATCCCATTACATGCT	10352988
LR606197.1	6047717	ATCCCATTACTTGCTCAACAA	6047737
LR597448.1	19915008	CATTACATGCTCAACAA	19915024
AP019297.1	4751659	CATTACATGCTCAACAA	4751675
AP019297.1	30436690	CCCATTACATGCTCA	30436676
AP019297.1	17094636	TTACATGCTCAACAA	17094650
CP064939.1	3184051	CCATTACATGCTCAACA	3184067
LR813497.2	978326	CCCATTACATGCTCAAC	978342
LR884475.1	22370334	CATTACATGCTCAACAA	22370350
CP051488.1	4062423	CCATTACATGCTCAACA	4062439
LR862150.1	14830253	CATTACATGCTCAACAA	14830269
AP022274.1	3949516	CCATTACATGCTCAACA	3949532
LR812551.1	30495017	CATTACATGCTCAACAA	30495033
LR812551.1	20403021	CCATTACATGCTCAAC	20403036
CP046702.1	2199171	CATYACATGCTCAACAA	2199187
CP046702.1	35161657	TCCCATTACATGCTC	35161643

PRODUCTS OF SIGNIFICANT ALIGNMENTS (CDS FEATURE)

- **D-amino acid dehydrogenase small subunit** GenBank
- **Adhesion G-protein coupled receptor G2-like** GenBank; GenBank
- **ZmABI3; B3 domain-containing transcription factor ABI3** GenBank
- **B3 domain-containing protein IDEF1** GenBank; GenBank; GenBank; GenBank
- **Formate dehydrogenase** GenBank; GenBank
- **Uncharacterized protein LOC103964099** GenBank
- **Transcriptional enhancer factor TEF-1** GenBank; GenBank; GenBank; GenBank; GenBank; GenBank GenBank; GenBank; GenBank; GenBank; GenBank; GenBank
- **Transcription factor bHLH25-like** GenBank
- **Transcription factor bHLH19-like** GenBank
- **Hypothetical protein** GenBank; GenBank; GenBank
- **LOW QUALITY PROTEIN: sacsin-like** GenBank; GenBank; GenBank; GenBank; GenBank; GenBank; GenBank
- **Sacsin** GenBank; GenBank
- **Peptidase** GenBank
- **Methionine synthase** GenBank
- **Bifunctional UDP-3-O-[3-hydroxymyristoyl]N-acetylglucosamine deacetylase/3-hydroxyacyl-ACP dehydratase** GenBank; GenBank
- **Fatty acid metabolism transcriptional regulator FadR** GenBank
- **Glycogen synthase GlgA** GenBank
- **HN1_G0040890.mRNA.1.CDS.1** GenBank
- **Carbohydrate ABC transporter permease** GenBank
- **Sugar ABC transporter permease** GenBank

SEQUENCE N° 3/2 (T=2; C=0; G=3; A=1)

Query: **TATCCCATTACATGATATTATT**

Search Parameters	
Program	blastn
Word size	7
Expect value	1000
Hitlist size	250
Match/Mismatch scores	1,-3
Gapcosts	5,2
Filter string	F
Genetic Code	1

250 sequences selected:

Accession Seq	Description	Comomm Name	Max Score	Total Score	Query C.	E value	Per. Id.	Acc. Len
.....	.....	.....	.....	.....	.....	.....	.....	.....
LR862368.1 1/2 2/2 3/3 3/5 6/1 7/2 8/1 9/1 10/1	***Pipistrellus pipistrellus genome assembly, chromosome: 12***	common pipistrelle	32.2	124	90%	138	100.00%	58291996
LR862363.1 1/2 2/2 3/3 3/5 6/1 7/2 8/1 9/1 10/1	***Pipistrellus pipistrellus genome assembly, chromosome: 7***	common pipistrelle	32.2	96.6	95%	138	100.00%	84548104
.....	.....	.....	.....	.....	.....	.....	.....	.....
LR862367.1 1/2 2/2 3/3 3/5 6/1 7/2 8/1 9/1 10/1	***Pipistrellus pipistrellus genome assembly, chromosome: 11***	common pipistrelle	32.2	62.4	90%	138	100.00%	69386144
LR862366.1 1/2 2/2 3/3 3/5 6/1 7/2 8/1 9/1 10/1	***Pipistrellus pipistrellus genome assembly, chromosome: 10***	common pipistrelle	32.2	62.4	72%	138	100.00%	71504835
LR862362.1 1/2 2/2 3/3 3/5 6/1 7/2 8/1 9/1 10/1	***Pipistrellus pipistrellus genome assembly, chromosome: 6***	common pipistrelle	32.2	62.4	95%	138	100.00%	85793147

SEQUENCE N° 3/3

(T=2; C=0; G=3; A=1)

SEQUENCE N° 3/3 (T=2; C=0; G=3; A=1)

Original Sequence	**G**	**T**	**G**	**A**	**A**	**A**	**T**	**G**	**G**	**T**	**C**	**A**	**T**	**G**	**T**	**G**	**T**	**G**	**G**	**C**	**G**	**G**
Trend n° 3 (T=2; C=0; G=3; A=1)	**3**	**2**	**3**	**1**	**1**	**1**	**2**	**3**	**3**	**2**	**0**	**1**	**2**	**3**	**2**	**3**	**2**	**3**	**3**	**0**	**3**	**3**
Trend n° 3 (I i D d)	**-**	**D**	**I**	**D**	**d**	**d**	**I**	**I**	**i**	**D**	**D**	**I**	**I**	**I**	**D**	**I**	**D**	**I**	**i**	**D**	**I**	**i**
Trend n° 3/3	**1**	**0**	**2**	**0**	**0**	**0**	**1**	**2**	**2**	**1**	**0**	**1**	**2**	**3**	**0**	**2**	**1**	**2**	**2**	**0**	**1**	**1**
Sequence n° 3/3 (77,27%)	**A**	**C**	**T**	**C**	**C**	**C**	**A**	**T**	**T**	**A**	**C**	**A**	**T**	**G**	**C**	**T**	**A**	**T**	**T**	**C**	**A**	**A**

Query: **ACTCCCATTACATGCTATTCAA**

Search Parameters	
Program	blastn
Word size	7
Expect value	1000
Hitlist size	100
Match/Mismatch scores	1,-3
Gapcosts	5,2
Filter string	F
Genetic Code	1

97 sequences selected:

Accession Seq	Description	Comomm Name	Max Score	Total Score	Query C.	E value	Per. Id.	Acc. Len
CP017762.1	Virgibacillus sp. 6R, complete genome	Virgibacillus sp. 6R	40.1	40.1	90%	0.57	100.00%	4749385
CP027633.1	Linum usitatissimum chromosome Lu9	flax	38.2	38.2	86%	2.2	100.00%	22091576
LT972194.1	Larimichthys crocea genome assembly, chromosome: V	large yellow croaker	38.2	38.2	86%	2.2	100.00%	34017524
OC164279.1	2_Tsi_b3v08	Timema shepardi	38.2	38.2	86%	2.2	100.00%	371
HE601627.2	Schistosoma mansoni strain Puerto Rico genome assembly, chromosome: 4	Schistosoma mansoni	38.2	38.2	86%	2.2	100.00%	47279781
AP022721.1	Plectropomus leopardus DNA, chromosome 22, nearly complete sequence	leopard coralgrouper	36.2	36.2	81%	8.9	100.00%	25156635
CP047562.1 1/2 2/2 3/2 3/5 6/1 8/1 9/1 10/1	Solanum pinnatisectum cultivar CGN17745 chromosome 7	tansyleaf nightshade	36.2	68.4	81%	8.9	100.00%	56727470
CP046689.1 1/2 2/2 3/2 3/5 6/1 8/1 9/1 10/1	Solanum tuberosum cultivar P8 chromosome 7	potato	36.2	68.4	81%	8.9	100.00%	56755552
CP046700.1 1/2 2/2 3/2 3/5 6/1 8/1 9/1 10/1	Solanum tuberosum cultivar MSH/14-112 chromosome 7	potato	36.2	68.4	81%	8.9	100.00%	56756616
CP039345.1	Vigna unguiculata cultivar Xiabao 2 chromosome Vu06	cowpea	36.2	36.2	81%	8.9	100.00%	45383982
XR_003327131.1	PREDICTED: Papaver somniferum uncharacterized LOC113282717 (LOC113282717), transcript variant X4, ncRNA	opium poppy	36.2	36.2	81%	8.9	100.00%	3718
XR_003327129.1	PREDICTED: Papaver somniferum uncharacterized LOC113282717 (LOC113282717), transcript variant X2, ncRNA	opium poppy	36.2	36.2	81%	8.9	100.00%	4091
XR_003327128.1	PREDICTED: Papaver somniferum uncharacterized LOC113282717 (LOC113282717), transcript variant X1, ncRNA	opium poppy	36.2	36.2	81%	8.9	100.00%	4371
CP025675.1	Ipomoea triloba cultivar NCNSP0323 chromosome 16	trilobed morning glory	36.2	36.2	81%	8.9	100.00%	18351525
CP025662.1	Ipomoea triloba cultivar NCNSP0323 chromosome 3	trilobed morning glory	36.2	36.2	81%	8.9	100.00%	33448119
CP025646.1	Ipomoea trifida cultivar NCNSP0306 chromosome 3	Ipomoea trifida	36.2	66.4	81%	8.9	100.00%	28619729
CP034499.1	Eukaryotic synthetic construct	eukaryotic	36.2	96.6	86%	8.9	100.00%	68480253

	chromosome 20	synthetic construct						
CP023762.1 1/2 2/2 3/2 3/5 6/1 8/1 9/1 10/1	Solanum lycopersicum cultivar I-3 chromosome 6	tomato	36.2	36.2	81%	8.9	100.00%	49756038
XR_470386.2	PREDICTED: Monodelphis domestica uncharacterized LOC103105194 (LOC103105194), ncRNA	gray short-tailed opossum	36.2	36.2	81%	8.9	100.00%	530
LK065388.1	Apteryx australis mantelli genome assembly AptMant0, scaffold scaffold806	Apteryx australis mantelli	36.2	36.2	81%	8.9	100.00%	1316800
HG975518.1	Solanum lycopersicum chromosome ch06, complete genome	tomato	36.2	36.2	81%	8.9	100.00%	46045610
HG975445.1 1/2 2/2 3/2 3/5 6/1 8/1 9/1 10/1	Solanum pennellii chromosome ch06, complete genome	Solanum pennellii	36.2	36.2	81%	8.9	100.00%	60730942
XM_036911866.1	**PREDICTED: Manis pentadactyla xin actin-binding repeat-containing protein 2-like (LOC118925732), partial mRNA**	**Chinese pangolin**	36.2	36.2	81%	8.9	100.00%	5303
AP023480.1 1/2 2/2 3/2 5/1 6/1 10/1	Homo sapiens DNA, chromosome 20, nearly complete genome	human	36.2	66.4	86%	8.9	100.00%	62398354
LR812537.1	Danio kyathit genome assembly, chromosome: 19	Danio kyathit	36.2	36.2	81%	8.9	100.00%	59732453
AC166324.1	Mus musculus BAC clone RP24-500J7 from chromosome 9, complete sequence	house mouse	36.2	36.2	81%	8.9	100.00%	168018
AL365401.20 1/2 3/5	**Human** DNA sequence from clone RP11-489M19 on chromosome 20, complete sequence	human	36.2	36.2	81%	8.9	100.00%	108291
CT030659.8	Mouse DNA sequence from clone RP23-125F8 on chromosome 9, complete sequence	house mouse	36.2	36.2	81%	8.9	100.00%	199662
LR812501.1	Danio aesculapii genome assembly, chromosome: 21	Danio aesculapii	34.2	34.2	77%	35	100.00%	46978008
LR812497.1	Danio aesculapii genome assembly, chromosome: 14	Danio aesculapii	34.2	64.4	77%	35	100.00%	56859170
AP023162.1	Felis catus Senzu DNA, chromosome: D2, American Shorthair breed	domestic cat	34.2	64.4	77%	35	100.00%	90643714
CP050569.1 1/2 2/2 3/4 3/5 3/6 5/1 10/1	Canis lupus familiaris breed Labrador retriever chromosome 15a	dog	34.2	94.6	100%	35	100.00%	64284575
CP050620.1 1/2 2/2 3/4 3/5 3/6 5/1 10/1	Canis lupus familiaris breed Labrador retriever chromosome 15b	dog	34.2	94.6	100%	35	100.00%	64292917
LR778278.1 1/2 2/2 8/1 9/1	Coregonus sp. 'balchen' genome assembly, chromosome: 26	Coregonus sp. 'balchen'	34.2	34.2	77%	35	100.00%	50922480
AP022706.1	Plectropomus leopardus DNA, chromosome 7, nearly complete sequence	leopard coralgrouper	34.2	64.4	86%	35	100.00%	36063594
XM_001953727.3	PREDICTED: Drosophila ananassae coiled-coil domain-containing protein 63 (LOC6499857), mRNA	Drosophila ananassae	34.2	34.2	77%	35	100.00%	1787
CP047567.1 1/2 2/2 3/2 3/5 6/1 8/1 9/1 10/1	Solanum pinnatisectum cultivar CGN17745 chromosome 1	tansyleaf nightshade	34.2	34.2	77%	35	100.00%	88611892
CP047558.1 1/2 2/2 3/2 3/5 6/1 8/1 9/1 10/1	Solanum pinnatisectum cultivar CGN17745 chromosome 12	tansyleaf nightshade	34.2	66.4	77%	35	100.00%	61096591
CP046684.1 1/2 2/2 3/2 3/5 6/1 8/1 9/1 10/1	Solanum tuberosum cultivar P8 chromosome 11	potato	34.2	66.4	77%	35	100.00%	45472623
CP046681.1	Solanum tuberosum cultivar P8 chromosome 12	potato	34.2	98.6	77%	35	100.00%	61163424

1/2 2/2 3/2 3/5 6/1 8/1 9/1 10/1									
CP046702.1									
1/2 2/2 3/2 3/5 6/1 8/1 9/1 10/1	Solanum tuberosum cultivar MSH/14-112 chromosome 1	potato	34.2	34.2	77%	35	100.00%	88648484	
CP046699.1 1/2 2/2 3/2 3/5 6/1 8/1 9/1 10/1	Solanum tuberosum cultivar MSH/14-112 chromosome 4	potato	34.2	34.2	77%	35	100.00%	72170609	
CP046697.1									
1/2 2/2 3/2 3/5 6/1 8/1 9/1 10/1	Solanum tuberosum cultivar MSH/14-112 chromosome 12	potato	34.2	66.4	77%	35	100.00%	61134531	
CP046691.1 1/2 2/2 3/2 3/5 6/1 8/1 9/1 10/1	Solanum tuberosum cultivar MSH/14-112 chromosome 11	potato	34.2	66.4	77%	35	100.00%	45472774	
LR738626.1	Sciurus vulgaris genome assembly, chromosome: 15	Eurasian red squirrel	34.2	124	100%	35	100.00%	90997183	
LR738604.1	Sciurus carolinensis genome assembly, chromosome: 14	gray squirrel	34.2	155	100%	35	100.00%	88654151	
CP045896.1	Caligus rogercresseyi isolate FCH chromosome 7	Caligus rogercresseyi	34.2	94.6	100%	35	100.00%	29872077	
CP045894.1	Caligus rogercresseyi isolate FCH chromosome 5	Caligus rogercresseyi	34.2	102	95%	35	95.24%	30612593	
XM_031403826.1	PREDICTED: Pistacia vera ADP-ribosylation factor GTPase-activating protein AGD2-like (LOC116117819), mRNA	Pistacia vera	34.2	34.2	77%	35	100.00%	614	
XM_031412472.1	PREDICTED: Pistacia vera ADP-ribosylation factor GTPase-activating protein AGD4-like (LOC116126793), mRNA	Pistacia vera	34.2	34.2	77%	35	100.00%	2644	
CP032575.1	Gossypium turneri isolate D10-2 chromosome D10_05	Gossypium turneri	34.2	94.6	100%	35	100.00%	60464062	
LR697107.1 1/2 2/2 3/1 3/4 8/1	**Chanos chanos genome assembly, chromosome: 2**	milkfish	34.2	94.6	100%	35	100.00%	59736555	
CP035805.1	Haemonchus contortus strain NZ_Hco_NP chromosome 1	barber pole worm	34.2	124	86%	35	100.00%	83295456	
LR633958.1	Gadus morhua genome assembly, chromosome: 16	Atlantic cod	34.2	94.6	86%	35	100.00%	34794352	
LR633957.1 1/2 2/2 3/2	Gadus morhua genome assembly, chromosome: 15	Atlantic cod	34.2	34.2	77%	35	100.00%	28657694	
LR633944.1 1/2 2/2 3/2	Gadus morhua genome assembly, chromosome: 2	Atlantic cod	34.2	98.6	90%	35	100.00%	28732775	
LR606182.1 1/2 3/2 3/5 9/1	Aquila chrysaetos chrysaetos genome assembly, chromosome: 2	Aquila chrysaetos chrysaetos	34.2	96.6	90%	35	100.00%	83001952	
LR597457.1	Salarias fasciatus genome assembly, chromosome: 23	jewelled blenny	34.2	94.6	77%	35	100.00%	41705463	
CP040924.1	Clostridium thermarum strain SYSU GA15002 chromosome, complete genome	Clostridium sp. SYSU GA15002	34.2	34.2	77%	35	100.00%	3805212	
CP032557.1	Gossypium raimondii isolate D5-4 chromosome D5_05	Gossypium raimondii	34.2	94.6	100%	35	100.00%	56796726	
LR594556.1	Streptopelia turtur genome assembly, chromosome: 5	Streptopelia turtur	34.2	64.4	86%	35	100.00%	70854002	
LR594589.1	Syngnathus acus genome assembly, chromosome: 3	greater pipefish	34.2	94.6	77%	35	100.00%	21527711	
CP033012.1	Buchnera aphidicola (Anoecia oenotherae) strain Aoe chromosome	Buchnera aphidicola (Anoecia oenotherae)	34.2	34.2	77%	35	100.00%	548691	
XM_028731005.1	PREDICTED: Podarcis muralis integral membrane protein 2C (ITM2C), mRNA	Common wall lizard	34.2	34.2	77%	35	100.00%	3624	
LR584444.1 1/2 2/2 3/4 3/6 8/1	Salmo trutta genome assembly, chromosome: 35	river trout	34.2	34.2	77%	35	100.00%	41918599	
LR584420.1 1/2 2/2 3/4 3/6 8/1	Salmo trutta genome assembly, chromosome: 4	river trout	34.2	34.2	77%	35	100.00%	73168960	
LR584403.1	Rhinatrema bivittatum genome assembly, chromosome: 17	two-lined caecilian	34.2	187	100%	35	100.00%	63936646	
XM_028449733.1	PREDICTED: Gouania willdenowi	blunt-snouted	34.2	34.2	77%	35	100.00%	6817	

1/2 3/4 3/6	SCY1 like pseudokinase 2 (scyl2), transcript variant X2, mRNA	clingfish						
XM_028449732.1 1/2 3/4 3/6	PREDICTED: Gouania willdenowi SCY1 like pseudokinase 2 (scyl2), transcript variant X1, mRNA	blunt-snouted clingfish	34.2	34.2	77%	35	100.00%	6826
LR535852.1	Mastacembelus armatus genome assembly, chromosome: 20	zig-zag eel	34.2	34.2	77%	35	100.00%	20063462
LR132035.1	Anabas testudineus genome assembly, chromosome: 5	climbing perch	34.2	64.4	77%	35	100.00%	26719370
LR132000.1 1/2 3/4 3/6	Gouania willdenowi genome assembly, chromosome: 6	blunt-snouted clingfish	34.2	66.4	90%	35	100.00%	72451976
LR131993.1 1/2 3/4 3/6	Gouania willdenowi genome assembly, chromosome: 5	blunt-snouted clingfish	34.2	64.4	95%	35	95.24%	42802790
AP019308.1	Paenibacillus baekrokdamisoli KCTC 33723 DNA, complete genome	Paenibacillus baekrokdamisoli	34.2	34.2	77%	35	100.00%	7261317
XM_026092999.1	PREDICTED: Dromaius novaehollandiae Rap guanine nucleotide exchange factor 4 (RAPGEF4), transcript variant X3, mRNA	emu	34.2	34.2	77%	35	100.00%	3874
XM_026092992.1	PREDICTED: Dromaius novaehollandiae Rap guanine nucleotide exchange factor 4 (RAPGEF4), transcript variant X2, mRNA	emu	34.2	34.2	77%	35	100.00%	4111
XM_026092982.1	PREDICTED: Dromaius novaehollandiae Rap guanine nucleotide exchange factor 4 (RAPGEF4), transcript variant X1, mRNA	emu	34.2	34.2	77%	35	100.00%	4381
XM_026042941.1	PREDICTED: Nothoprocta perdicaria Rap guanine nucleotide exchange factor 4 (RAPGEF4), transcript variant X2, mRNA	Nothoprocta perdicaria	34.2	34.2	77%	35	100.00%	3708
XM_026042863.1	PREDICTED: Nothoprocta perdicaria Rap guanine nucleotide exchange factor 4 (RAPGEF4), transcript variant X1, mRNA	Nothoprocta perdicaria	34.2	34.2	77%	35	100.00%	3762
XM_014902763.2	PREDICTED: Brachypodium distachyon disease resistance RPP13-like protein 4 (LOC100838112), transcript variant X4, mRNA	stiff brome	34.2	34.2	77%	35	100.00%	2198
CP023118.1	Lupinus angustifolius cultivar Tanjil chromosome LG-06	narrow-leaved blue lupine	34.2	34.2	77%	35	100.00%	40902325
CP023115.1	Lupinus angustifolius cultivar Tanjil chromosome LG-03	narrow-leaved blue lupine	34.2	34.2	77%	35	100.00%	38743218
CP020637.1	Oryzias latipes strain HSOK chromosome 17	Japanese medaka	34.2	34.2	77%	35	100.00%	33291814
CP020795.1	Oryzias latipes strain HNI chromosome 17	Japanese medaka	34.2	34.2	77%	35	100.00%	28809336
CP020681.1	Oryzias latipes strain Hd-rR chromosome 17 sequence	Japanese medaka	34.2	34.2	77%	35	100.00%	31792230
XM_021018677.1	PREDICTED: Arabidopsis lyrata subsp. lyrata uncharacterized LOC9305153 (LOC9305153), transcript variant X2, mRNA	Arabidopsis lyrata subsp. lyrata	34.2	34.2	77%	35	100.00%	1088
XM_002867023.2	PREDICTED: Arabidopsis lyrata subsp. lyrata uncharacterized LOC9305153 (LOC9305153), transcript variant X1, mRNA	Arabidopsis lyrata subsp. lyrata	34.2	34.2	77%	35	100.00%	1239
CP019645.1	Helicobacter bilis strain AAQJH, complete genome	Helicobacter bilis	34.2	34.2	77%	35	100.00%	2620399
XM_017981523.1	PREDICTED: Drosophila busckii protein lethal(2)k10201 (LOC108596103), mRNA	Drosophila busckii	34.2	34.2	77%	35	100.00%	676
HF933223.1	TPA: Oryzias latipes strain Hd-rR, complete genome assembly, chromosome 17	Japanese medaka	34.2	34.2	77%	35	100.00%	31848461
KU359346.1	Uncultured archaeon isolate DGGE gel band AOA-a ammonia monooxygenase gene, partial cds	uncultured archaeon	34.2	34.2	77%	35	100.00%	592
CP062013.1	Macrobrachium nipponense isolate FS-2020 chromosome 22	Macrobrachium nipponense	34.2	64.4	86%	35	100.00%	83304076
CP055245.1								
1/2 2/2 3/2 6/1 8/1 9/1 10/1	Solanum tuberosum cultivar Solyntus chromosome 12	potato	34.2	66.4	77%	35	100.00%	68020514

CP055244.1 1/2 2/2 3/2 6/1 8/1 9/1 10/1	Solanum tuberosum cultivar Solyntus chromosome 11	potato	34.2	66.4	77%	35	100.00%	57022023
CP055241.1 1/2 2/2 3/2 6/1 8/1 9/1 10/1	Solanum tuberosum cultivar Solyntus chromosome 8	potato	34.2	34.2	77%	35	100.00%	67850527
CP055237.1 1/2 2/2 3/2 6/1 8/1 9/1 10/1	Solanum tuberosum cultivar Solyntus chromosome 4	potato	34.2	34.2	77%	35	100.00%	72008707
LR877218.1	Acomys russatus genome assembly, chromosome: 7	golden spiny mouse	34.2	34.2	77%	35	100.00%	69034057

PREDICTED: Manis pentadactyla xin actin-binding repeat-containing protein 2-like (LOC118925732), partial mRNA

NCBI Reference Sequence: XM_036911866.1

FASTA Graphics

Go to:

LOCUS XM_036911866 5303 bp mRNA linear MAM **28-OCT-2020**

DEFINITION PREDICTED: Manis pentadactyla xin actin-binding repeat-containing protein 2-like (LOC118925732), partial mRNA.

ACCESSION XM_036911866

VERSION XM_036911866.1

DBLINK BioProject: PRJNA666705

KEYWORDS RefSeq; corrected model; includes ab initio.

SOURCE **Manis pentadactyla (Chinese pangolin)**

ORGANISM Manis pentadactyla

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Laurasiatheria; Pholidota; Manidae; Manis.

COMMENT MODEL REFSEQ: This record is predicted by automated computational analysis. This record is derived from a genomic sequence (NW_023455251.1) annotated using gene prediction method: Gnomon.

Also see:

Documentation of NCBI's Annotation Process

##Genome-Annotation-Data-START##

Annotation Provider :: NCBI

Annotation Status :: Full annotation

Annotation Name :: Manis pentadactyla Annotation Release 100

```
            Annotation Version          :: 100
            Annotation Pipeline         :: NCBI eukaryotic genome annotation
                                           pipeline
            Annotation Software Version :: 8.5
            Annotation Method           :: Best-placed RefSeq; Gnomon
            Features Annotated          :: Gene; mRNA; CDS; ncRNA
            ##Genome-Annotation-Data-END##
            ##RefSeq-Attributes-START##
            ab initio            :: 1% of CDS bases
            frameshifts          :: corrected 4 indels
            internal stop codons :: corrected 2 genomic stop codons
            ##RefSeq-Attributes-END##
            COMPLETENESS: incomplete on the 5' end.
PRIMARY     REFSEQ_SPAN          PRIMARY_IDENTIFIER PRIMARY_SPAN        COMP
            1-621                VBRX01004258.1     15246-15866         c
            622-622              "N"                1-1
            623-2829             VBRX01004258.1     13039-15245         c
            2830-2830            "N"                1-1
            2831-3300            VBRX01004258.1     12569-13038         c
            3301-3376            VBRX01004258.1     12492-12567         c
            3377-3494            VBRX01004258.1     10316-10433         c
            3495-4041            VBRX01004258.1     7334-7880           c
            4042-4042            "N"                1-1
            4043-5303            VBRX01004258.1     6073-7333           c
FEATURES             Location/Qualifiers
     source          1..5303
                     /organism="Manis pentadactyla"
                     /mol_type="mRNA"
                     /isolate="MP20"
                     /db_xref="taxon:143292"
                     /chromosome="Unknown"
                     /dev_stage="adult"
                     /collection_date="2017"
     gene            <1..5303
                     /gene="LOC118925732"
                     /note="The sequence of the model RefSeq transcript was
                     modified relative to its source genomic sequence to
                     represent the inferred CDS: inserted 3 bases in 3 codons;
                     deleted 1 base in 1 codon; Derived by automated
                     computational analysis using gene prediction method:
                     Gnomon. Supporting evidence includes similarity to: 2
                     Proteins, and 78% coverage of the annotated genomic
                     feature by RNAseq alignments"
                     /db_xref="GeneID:118925732"
     CDS             <1..4941
```

/gene="LOC118925732"
/note="The sequence of the model RefSeq protein was modified relative to its source genomic sequence to represent the inferred CDS: inserted 3 bases in 3 codons; deleted 1 base in 1 codon; substituted 2 bases at 2 genomic stop codons"
/codon_start=1
/transl_except=(pos:2755..2757,aa:OTHER)
/transl_except=(pos:4057..4059,aa:OTHER)
/product="**LOW QUALITY PROTEIN: xin actin-binding repeat-containing protein 2-like**"
/protein_id="XP_036767761.1"
/db_xref="GeneID:118925732"
/translation="IGSDYEETLKPSSIVGTSSTSCTPISQRKETSTRRSSDHRATSS
TLVQDNTVPSGGTEEFPPPPPSSLQTPIDVMAFSQSPEFPSPPRMPPVSKELCSKQRN
LYELNHLYKHIHPELRKNLEKDYISQVSEIVSSQVNPSSSVSADVQQARYVFENTNIS
CPKGLNSEREYLECNEILKGEVQSMRWIFENQPLDAINNGSPEEGNTXKGIADQEIIA
VGDVKYTTWMFETQPIDMLGDHSSGTEENAEKIPELARGDVCTARWMFETKPLDSMNK
MHPSQEEPGVSSIKDIIGGDVKTVRYMFETLHLDQLGQFHTVDQVHLLQLRSELQEIK
GNVKQSIKCFETQPFCVIRDGLGQTLEIKTVHREDVEKGDVKTAQGLETQPLDTINKD
ITEIKVIQEISMEEKVKGGVSRAKWLFETQPLEKIKESEEDMIEKETIMGTDVSRKCW
MFETQPSDILKEVPDADSVRSEEIIGGDVQTTKPLFQTLPIEALKDSPDVGKLQKITT
SEEEKGDVRHQKWIFETQPLEEIREDKKEYIRTVKLEEVDRGDIRNYTNTFESNNLIK
FDASHKIEVKGVTRGAVELNKSLFETTPLHAIQDHLGKCHQVKTVQQEEILRGEVRSC
RWLFETRPIDQFDESIHNFQIIRGISAQDIQTGNVKSAKWLFETQPLDSIKYFSNMEV
ESKIEQATDIIKGNVKTCSQPMESLYEKVSLMTDSEEIPKGDIKACTWLFETRPLDTI
KDDSEETVKSQAVKQEEINGVDVCSACFLFETENLDSIKGEEGKEIKAVEVDIQAGDV
SSMRHKFENQSLDSIRSSEEVLRKITSPATLCRQIKIETRRRDSPPIITIPVSVNHVD
SSSFRESMEAQEEVRKVGKRATYIHKDRISSSGCIGPDTECFDAVEIIHKVEGPLRSE
HTQRXEAANQTVQMAENFMNDCENEINRWXREFEDGPVLEAKSKRRVYVNGEANHNIT
LESHTFCKEEFGLTSSENSSFTGFSGNHPRELQEKIPVKQPSICSETRYLSEHFSGVD
AFESQVVGLKMTVSSPHNSEAGKSGFDFKHAPPTYEDVIAGHILDVSDSPKELRRNFQ
EMWQESQRVFKKLGYATSDSSATEMRTAFQEESAFISDKEICILCQKTVYPEECLIAD
MQIFHESCFRCHNCDSKLSLGYYASLHGQIYCKPHFEQLFKSKGNYDEGFGHKQHKDR
WNCKIQSSSVDCIPNEAPNMCKNSIENTPTLRDLNKHLDAENSKGQRDELRKVRERRK
LKIIWPPSRKLPEKTFPLEEELKMSKPKWPPEMTAPASAEFKDELREHNNTLENKGQE
QDTSLLETYLQHIHMCQKDVTGIQEMEXYKAKXDEKEGSKNVPDNLNEAEEIKSKRKS
EMDVSKNNIGVVQIAEEKNEPGNELDDAEVIQVTDTDDEVVPENHKENLNKKNNNYVA
VSDLNNGRQKTSILEFPNLLPLSSEANCVASDYVIENLESVPRISEETYSRNVPAMAC
KKQAGRATAGSPVQSAPKSVLSSSLRGKRNSSMASPPNTNTLNIKENHSNNKNLCFFF
SNTVKIAAFSKKNEDPDLIESVDPIKNMPCLYLRELGKNIKHWHGEAGAAGNTSFDAL
SSECATKPSFPRVEDQFEQLTVEEQIKRNRCYSETE"

ORIGIN

```
        1 attggaagtg actatgaaga gactcttaag ccatcatcca ttgtgggtac ctcttcgact
       61 tcttgtactc caatcagcca gaggaaggaa acatcgacca gacgatctag tgaccacagg
```

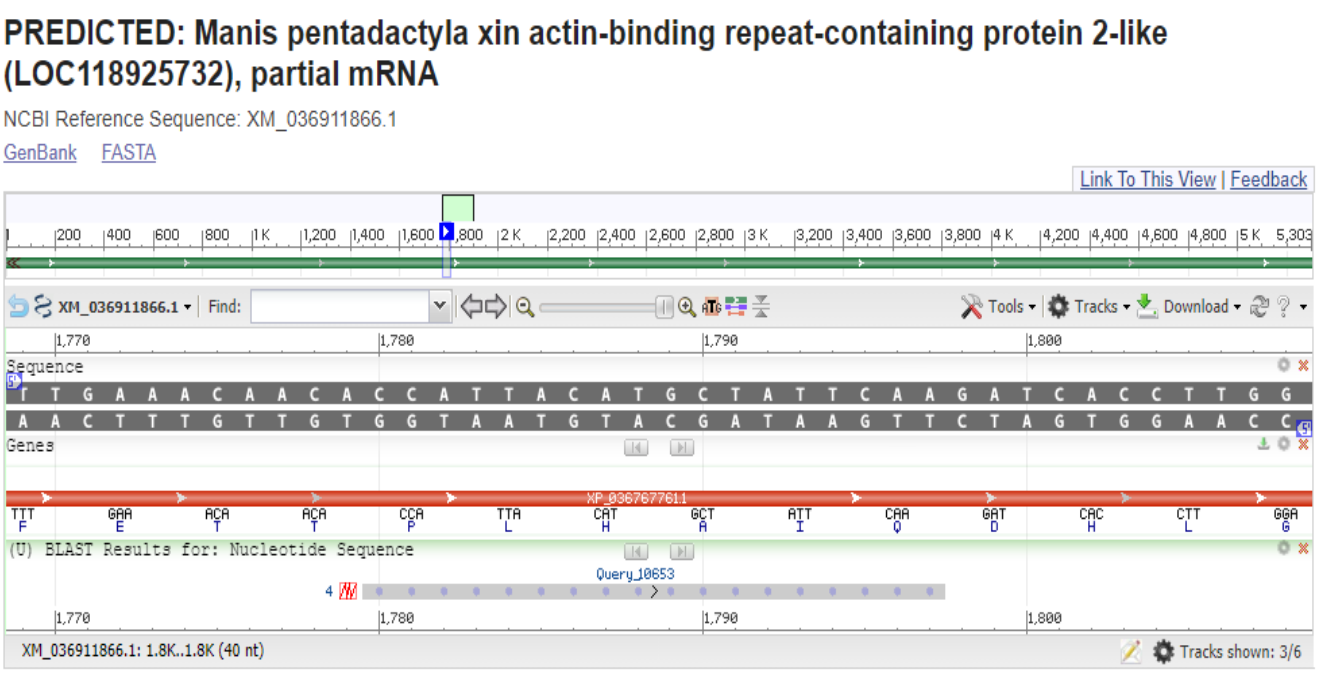
121 gccacctcct caacactggt gcaagataat accgttcctt caggagggac agaagaattt
181 cctcctcccc cacccagcag ccttcaaact ccaatagatg tgatggcatt ttcccagtcc
241 cctgaattcc ccagccctcc tagaatgcca ccagtctcca aagaattatg ttctaaacaa
301 agaaatttgt atgagctaaa ccatttatat aaacatattc atccggagtt aaggaagaac
361 ttagaaaaag attatatcag tcaggtttct gagattgttt ctagtcaagt gaatccaagc
421 agctcagttt ctgcagatgt tcagcaagcc cggtatgttt ttgaaaacac aaatatcagt
481 tgtccaaaag gcctgaactc agaaagagaa tatttggagt gtaatgagat tctgaagggg
541 gaggtgcagt ccatgaggtg gatttttgaa aatcagccat tagatgccat caacaatggc
601 tctccagagg aaggtaacac tnccaagggc attgctgatc aagaaatcat tgctgttggt
661 gatgtgaaat ataccacatg gatgtttgaa acacaaccca tagatatgct gggggatcat
721 tcttctggta ctgaagaaaa tgctgagaaa attcctgagc tagccagagg agatgtttgt
781 acagccagat ggatgtttga aacaaagcca ttagattcaa tgaataaaat gcatccaagt
841 caagaagaac caggggtatc ttccataaag gatataattg gcggggatgt caagactgtg
901 agatacatgt ttgaaactct acatctggat caacttggac agtttcatac agtggatcag
961 gttcacttac tgcagctcag gtctgagctc caagaaatta agggaaatgt taagcaaagc
1021 ataaaatgtt ttgaaactca accattctgt gttattagag atggtttagg tcaaacgcta
1081 gaaattaaaa ctgttcacag agaagatgtt gaaaaggggg atgtgaaaac agcacaaggg
1141 cttgaaacac agcccttgga cacaatcaat aaggatatca cagaaattaa agtcatccaa
1201 gaaatatcca tggaagaaaa ggtcaaaggt ggggtgagca gggcaaagtg gttgtttgaa
1261 actcagcctt tggagaaaat caaagagtca gaagaggaca tgattgagaa ggaaacaatt
1321 atgggtacag atgtctccag aaagtgttgg atgtttgaaa cacagccatc agacattcta
1381 aaggaagttc ccgacgcaga ttctgtaaga tctgaagaga taataggggg cgatgtacaa
1441 actaccaagc ctctatttca aacacttcca atagaagctt taaaagatag ccctgatgta
1501 ggtaaacttc aaaaaattac cacctctgaa gaagaaaagg gggatgttag gcaccaaaag
1561 tggattttcg aaacccaacc tctggaagag attagagaag ataaaaaaga gtacatacga
1621 acagtgaaac ttgaagaggt tgacagagga gatatcagaa attacacaaa tacctttgag
1681 tcaaacaatt taattaaatt tgatgcatca cataagatag aggtgaaagg agtcacaaga
1741 ggtgctgtag aattaaataa atcactcttt gaaacaacac cattacatgc tattcaagat
1801 caccttggaa aatgtcatca ggtaaagaca gtccagcaag aagaaatcct aagaggtgaa
1861 gtaagaagct gcaggtggct ttttgaaaca aggcccattg accagtttga tgaaagcatt
1921 cataactttc aaataatcag aggaatatct gctcaggaca tacagactgg aaatgtgaaa
1981 tctgctaagt ggttgtttga aacccaacct cttgattcaa ttaaatattt tagtaatatg
2041 gaagtagaaa gtaaaattga acaagctaca gatatcatta aagggaatgt caaaacctgt
2101 agccagccaa tggagtctct ttatgaaaaa gtctcattaa tgacagacag tgaagaaatt
2161 cctaagggag acatcaaagc ctgtacttgg ctctttgaaa ctcggcccct tgataccata
2221 aaagatgact ctgaagaaac agtcaaatca caagctgtaa aacaggagga gatcaatggt
2281 gtagatgttt gttcagcgtg ttttcttttt gagacagaaa atctggacag cataaaagga
2341 gaagaaggaa aggaaatcaa ggctgtggaa gtagatatcc aagctgggga tgtctccagc
2401 atgcggcata aatttgaaaa tcagtcctta gactctatca ggtcctcaga ggaagttttg
2461 agaaagatca catctcctgc aacactttgt cgccaaatta agatagaaac tcgtcgtagg
2521 gactctccac **ctataa**ttac aataccggta agtgtaaatc atgttgatag tagttccttc
2581 agagaatcca tggaagcgca agaggaagtt aggaaagtag ggaaaagggc gacttacatc
2641 cacaaagaca gaataagttc ctctgggtgt atagggccag atactgaatg ttttgacgca
2701 gttgaaatca tccacaaagt tgaagggccc ctcaggtcag agcacacaca gagataggaa
2761 gcagccaacc aaactgttca aatggctgaa aatttcatga atgactgtga aaatgaaata

2821 aacagatggn tcagggaatt tgaggatggc ccagttttgg aagcaaagtc aaagagaaga
2881 gtttatgtaa atggagaagc aaaccataat ataacactag aaagtcatac attttgtaag
2941 gaggaatttg gattaacatc ttcagagaac tctagtttta caggcttttc aggcaatcat
3001 cctagggaac tgcaagaaaa aattcctgtt aagcagccta gcatatgctc tgaaacaagg
3061 tacctaagtg agcatttctc aggtgtggat gcatttgaga gtcaagttgt ggggttgaag
3121 atgacagttt cttcaccaca taactcagaa gctggcaaat ctggctttga cttcaagcat
3181 gccccaccaa cctatgaaga cgtcatcgct ggccacattt tagatgtttc tgattcacct
3241 aaagaactca ggaggaattt tcaagagatg tggcaggaga gtcaaagagt ctttaaaaaa
3301 ctgggatatg caacctcaga ttcttctgca acagagatga gaactgcctt ccaagaagaa
3361 tctgcattta taagtgacaa ggaaatatgt atactttgtc aaaagacagt ttatccagag
3421 gagtgcctca tagccgacat gcagattttc catgagtcct gcttcagatg ccacaactgt
3481 gacagtaaac tgagtttggg atattatgca tcacttcatg gacaaatata ttgtaaacca
3541 cactttgagc aacttttcaa atctaaagga aattatgatg aaggttttgg acacaaacag
3601 cataaagacc gatggaattg caaaattcag agcagctcag tggactgtat tcctaatgaa
3661 gcaccaaata tgtgtaaaaa cagtatagaa aacaccccca cacttaggga tcttaacaaa
3721 catttagatg ctgagaacag caaagggcaa agggatgagc tgagaaaagt cagggagagg
3781 agaaaactaa aaatcatttg gcccccatct aggaagctgc ctgagaaaac ctttcccctt
3841 gaggaagaac tcaaaatgag caaaccgaaa tggccacctg aaatgacagc ccctgcatct
3901 gctgaattta aagatgaact aagagaacac aataacactc tggaaaataa aggacaagag
3961 caagatactt ctctccttga gacatatctg cagcacatcc atatgtgtca gaaagatgtt
4021 acaggaatcc aagaaatgga angttataaa gcaaagtaag atgagaagga agggagtaag
4081 aatgtgccag ataatctgaa tgaggctgaa gaaataaaga gtaagaggaa aagtgaaatg
4141 gatgttagca agaataatat tggggttgtt cagatagctg aggagaaaaa tgaaccaggt
4201 aatgaacttg atgatgcaga agttatacag gttactgaca ctgatgatga ggtggtacca
4261 gaaaatcata aagagaattt gaataagaag aataataatt atgtagcagt ctcggatctg
4321 aataatggca ggcagaagac atctatttta gaatttccta atctattacc actgtcaagt
4381 gaagcaaact gtgttgcaag tgactatgtg attgaaaact tagaaagtgt acccaggatc
4441 tctgaagaga cgtactcgag gaatgtacca gctatggctt gcaagaagca ggctggcaga
4501 gctactgctg gcagtcctgt gcagtctgcc cccaagtcag tcctcagtag cagcttaaga
4561 ggaaagagga acagctcaat ggcctctcct cctaacacaa ataccttaaa catcaaagaa
4621 aaccactcaa ataacaagaa tctatgcttc ttcttttcta acactgtgaa aattgctgct
4681 ttttccaaga agaatgagga tcctgattta atagagtctg tagatccaat taaaaacatg
4741 ccatgtttgt atttaagaga acttgggaag aacatcaaac attggcatgg tgaagcagga
4801 gcagcaggaa acacaagctt tgatgctctg agcagtgaat gtgcaactaa gccttcattt
4861 cccagagtgg aggaccagtt tgaacaactc actgtggaag agcaaattaa aaggaatagg
4921 tgctacagtg aaacagagta aaacatctct ggccacttgc agtggacttc aggcttggaa
4981 acacgcaggc atggacttta gggattttga tgacgtttag tagaacttgt aagcttaatg
5041 ttcagaaatc tcatgtctat cacaatggga tattccttgt gttacaccgt atgtttttcc
5101 caattggttt gaatctactt ttaatccgaa tggaatgaaa aaaaatgaaa caatatggct
5161 atgttttctg tcaagtggca cttacatgta tcgctctctt tttcaacatg ccagagaaca
5221 tcaccaggga gattctattc tgctaatgtt tatcataatt cttatacctt **ctctataatt**
5281 tctgctgaaa taaaattcaa tca //

PREDICTED: Manis pentadactyla xin actin-binding repeat-containing protein 2-like (LOC118925732), partial mRNA
Sequence ID: XM_036911866.1 Length: 5303 Number of Matches: 1
Range 1: 1780 to 1797 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
36.2 bits(18)	9.1	18/18(100%)	0/18(0%)	Plus/Plus

```
CDS: Putative 1        1      P  L  H  A  I  Q
Query                  5      CCATTACATGCTATTCAA  22
                              ||||||||||||||||||
Sbjct                  1780   CCATTACATGCTATTCAA  1797
CDS:LOW QUALITY PROT   594     P  L  H  A  I  Q
```



Pipistrellus pipistrellus genome assembly, chromosome: 11

Sequence ID: LR862367.1 **Length: 69386144 Number of Matches: 2**

Range 1: 58336268 to 58336284 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
34.2 bits(17)	35	17/17(100%)	0/17(0%)	Plus/Plus

```
Query  2         GCTTTGCCAGTGCATAT  18
                 ||||||||||||||||
Sbjct  58336268  GCTTTGCCAGTGCATAT  58336284
```

Range 2: 68321361 to 68321376 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
32.2 bits(16)	138	16/16(100%)	0/16(0%)	Plus/Plus

```
Query  3         CTTTGCCAGTGCATAT  18
                 ||||||||||||||||
Sbjct  68321361  CTTTGCCAGTGCATAT  68321376
```

ALIGNMENT VIEW FLAT QUERY-ANCHORED WITH LETTERS FOR IDENTITIES

Query range 1: 1 to 22

Query	**1**	**ACTCCCATTACATGCTATTCAA**	**22**
CP017762.1	862396	TCCCATTACATGCTATTCAA	862377
CP027633.1	19889483	CCCATTACATGCTATTCAA	19889465
LT972194.1	21776087	CTCCCATTACATGCTATTC	21776105
OC164279.1	271	CCCATTACATGCTATTCAA	289
HE601627.2	7312442	CTCCCATTACATGCTATTC	7312460
AP022721.1	18314661	CTCCCATTACATGCTATT	18314644
CP047562.1	12580305	CCATTACATGCTATTCAA	12580288
CP047562.1	33453785	ATTACATGCTATTCAA	33453770
CP046689.1	12584707	CCATTACATGCTATTCAA	12584690
CP046689.1	33461441	ATTACATGCTATTCAA	33461426
CP046700.1	12582316	CCATTACATGCTATTCAA	12582299
CP046700.1	33462030	ATTACATGCTATTCAA	33462015
CP039345.1	39954919	ACTCCCATTACATGCTAT	39954936
XR_003327131.1	1995	ACTCCCATTACATGCTAT	1978
XR_003327129.1	2368	ACTCCCATTACATGCTAT	2351
XR_003327128.1	2648	ACTCCCATTACATGCTAT	2631
CP025675.1	2093443	CTCCCATTACATGCTATT	2093460
CP025662.1	8191839	CTCCCATTACATGCTATT	8191856
CP025646.1	27002995	CTCCCATTACATGCTATT	27002978
CP025646.1	6565722	CCATTACATGCTATT	6565736

CP034499.1	60345042	ACTCCCATTACATGCTAT	60345059
CP034499.1	4598376	CCATTACATGCTATT	4598390
CP034499.1	66836077	CCATTACATGCTATT	66836091
CP023762.1	35297423	CCATTACATGCTATTCAA	35297440
XR_470386.2	511	CCCATTACATGCTATTCA	494
LK065388.1	1196630	CTCCCATTACATGCTATT	1196613
HG975518.1	31786909	CCATTACATGCTATTCAA	31786926
HG975445.1	44716352	CCATTACATGCTATTCAA	44716369
XM_036911866.1	1780	CCATTACATGCTATTCAA	1797
AP023480.1	59911569	ACTCCCATTACATGCTAT	59911586
AP023480.1	4610624	CCATTACATGCTATT	4610638
LR812537.1	6511371	CCATTACATGCTATTCAA	6511388
AC166324.1	102314	CCCATTACATGCTATTCA	102331
AL365401.20	51972	ACTCCCATTACATGCTAT	51989
CT030659.8	186345	CCCATTACATGCTATTCA	186362
LR812501.1	40119541	CCATTACATGCTATTCA	40119525
LR812497.1	50452228	CATTACATGCTATTCAA	50452212
LR812497.1	26544760	TTACATGCTATTCAA	26544746
AP023162.1	29214796	ACTCCCATTACATGCTA	29214812
AP023162.1	40097834	ACTCCCATTACATGC	40097820
CP050569.1	53692157	ACTCCCATTACATGCTA	53692173
CP050569.1	25090251	TTACATGCTATTCAA	25090237
CP050569.1	45363170	TTACATGCTATTCAA	45363184
CP050620.1	53707877	ACTCCCATTACATGCTA	53707893
CP050620.1	25093979	TTACATGCTATTCAA	25093965
CP050620.1	45380290	TTACATGCTATTCAA	45380304
LR778278.1	35760565	CATTACATGCTATTCAA	35760549
AP022706.1	17994870	TCCCATTACATGCTATT	17994854
AP022706.1	25884063	ACTCCCATTACATGC	25884077
XM_001953727.3	1753	CCCATTACATGCTATTC	1769
CP047567.1	68100976	CATTACATGCTATTCAA	68100960
CP047558.1	56727502	CATTACATGCTATTCAA	56727486
CP047558.1	60127672	ATTACATGCTATTCAA	60127687
CP046684.1	20468724	CATTACATGCTATTCAA	20468740
CP046684.1	31157508	CATTACATGCTATTCA	31157493
CP046681.1	56794043	CATTACATGCTATTCAA	56794027
CP046681.1	45680165	CATTACATGCTATTCA	45680150
CP046681.1	60194350	ATTACATGCTATTCAA	60194365
CP046702.1	68128230	CATTACATGCTATTCAA	68128214
CP046699.1	5919201	CATTACATGCTATTCAA	5919185
CP046697.1	56765144	CATTACATGCTATTCAA	56765128
CP046697.1	60165451	ATTACATGCTATTCAA	60165466
CP046691.1	20468875	CATTACATGCTATTCAA	20468891
CP046691.1	31157659	CATTACATGCTATTCA	31157644
LR738626.1	24957934	ACTCCCATTACATGCTA	24957918
LR738626.1	19101037	CATTACATGCTATTC	19101051
LR738626.1	47387176	ATTACATGCTATTCA	47387162
LR738626.1	57404148	TTACATGCTATTCAA	57404134
LR738604.1	64115669	ACTCCCATTACATGCTA	64115685
LR738604.1	32851063	TTACATGCTATTCAA	32851077
LR738604.1	42356261	ATTACATGCTATTCA	42356275
LR738604.1	65443594	ACTCCCATTACATGC	65443608
LR738604.1	69827040	CATTACATGCTATTC	69827026
CP045896.1	24123180	CTCCCATTACATGCTAT	24123164
CP045896.1	6908178	ACTCCCATTACATGC	6908164
CP045896.1	27338383	TTACATGCTATTCAA	27338369
CP045894.1	21543532	ACTCTCATTACATGCTATTCA	21543512
CP045894.1	21568810	ACTCTCATTACATGCTATTCA	21568830
CP045894.1	21573646	ACTCTCATTACATGCTATTCA	21573666
XM_031403826.1	461	CCCATTACATGCTATTC	445

XM_031412472.1	211	CCCATTACATGCTATTC	195
CP032575.1	40054738	ACTCCCATTACAAGCTATT	40054756
CP032575.1	59502103	CCCATTACATGCTAGTCAA	59502121
CP032575.1	59985713	CATTACATGCTATTCAA	59985729
LR697107.1	14547457	ACTCCCATTACATGCTA	14547473
LR697107.1	22114195	TTACATGCTATTCAA	22114181
LR697107.1	24648831	CCATTACATGCTATT	24648817
CP035805.1	46204447	CCCATTACATGCTATTC	46204431
CP035805.1	55256433	CTCCCATTACATGCT	55256447
CP035805.1	55258099	CTCCCATTACATGCT	55258113
CP035805.1	55259851	CTCCCATTACATGCT	55259865
LR633958.1	28541402	TCCCATTACATGCTATT	28541386
LR633958.1	5320960	TCCCATTACATGCTA	5320974
LR633958.1	19337100	ATTACATGCTATTCA	19337114
LR633957.1	9359176	TCCCATTACATGCTATT	9359160
LR633944.1	328293	TCCCATTACATGCTATT	328277
LR633944.1	329012	TCCCATTACATGCTATT	328996
LR633944.1	27927790	TTACATGCTATTCAA	27927776
LR606182.1	36877406	TCCCATTACATGCTATT	36877422
LR606182.1	18920084	CCATTACATGCTATTC	18920099
LR606182.1	63221325	TTACATGCTATTCAA	63221339
LR597457.1	5998417	CATTACATGCTATTCAA	5998401
LR597457.1	22387956	TTACATGCTATTCAA	22387942
LR597457.1	29432160	TTACATGCTATTCAA	29432146
CP040924.1	2843564	TCCCATTACATGCTATT	2843580
CP032557.1	20366047	ACTCCCATTACAAGCTATT	20366029
CP032557.1	490197	CATTACATGCTATTCAA	490181
CP032557.1	19549316	CATTACATGCTATTC	19549302
LR594556.1	25948892	TCCCATTGCATGCTATTCA	25948874
LR594556.1	34441629	TCCCATTACATGCTATT	34441613
LR594589.1	503296	CATTACATGCTATTCAA	503312
LR594589.1	4281444	TTACATGCTATTCAA	4281430
LR594589.1	12950896	CATTACATGCTATTC	12950910
CP033012.1	266895	CATTACATGCTATTCAA	266879
XM_028731005.1	2961	CTCCCATTACATGCTAT	2945
LR584444.1	331044	CATTACATGCTATTCAA	331028
LR584420.1	38651378	CATTACATGCTATTCAA	38651394
LR584403.1	49326681	TCCCATTACATTCTATTCAA	49326700
LR584403.1	39254652	ACTCCCATTACATGCTA	39254636
LR584403.1	26460935	CCATTACATGCTATT	26460949
LR584403.1	45730826	CCATTACATGCTATT	45730840
LR584403.1	48250383	CCATTACATGCTATT	48250397
LR584403.1	60477111	CCATTACATGCTATT	60477097
XM_028449733.1	16	TCCCATTACATGCTATT	32
XM_028449732.1	16	TCCCATTACATGCTATT	32
LR535852.1	7366302	TCCCATTACATGCTATT	7366286
LR132035.1	5782616	CATTACATGCTATTCAA	5782600
LR132035.1	424592	CATTACATGCTATTC	424606
LR132000.1	17549045	ACTCCCATTACACGCTATTC	17549026
LR132000.1	8021281	TCCCATTACATGCTATT	8021297
LR131993.1	10909371	CTCCCATTACATGCTTTTCAA	10909351
LR131993.1	33793925	CATTACATGCTATTC	33793911
AP019308.1	5689581	CCCATTACATGCTATTC	5689565
XM_026092999.1	3767	ACTCCCATTACATGCTA	3751
XM_026092992.1	4004	ACTCCCATTACATGCTA	3988
XM_026092982.1	4274	ACTCCCATTACATGCTA	4258
XM_026042941.1	3647	ACTCCCATTACATGCTA	3631
XM_026042863.1	3701	ACTCCCATTACATGCTA	3685
XM_014902763.2	1990	CCATTACATGCTATTCA	2006
CP023118.1	37884455	CATTACATGCTATTCAA	37884439

CP023115.1	27769250	CATTACATGCTATTCAA	27769234
CP020637.1	24681771	TCCCATTACATGCTATT	24681787
CP020795.1	20992119	TCCCATTACATGCTATT	20992135
CP020681.1	23867421	TCCCATTACATGCTATT	23867437
XM_021018677.1	683	ACTCCCATTACATGCTA	699
XM_002867023.2	834	ACTCCCATTACATGCTA	850
CP019645.1	1341986	CATTACATGCTATTCAA	1341970
XM_017981523.1	642	CATTACATGCTATTCAA	626
HF933223.1	24480052	TCCCATTACATGCTATT	24480068
KU359346.1	165	CCATTACATGCTATTCA	181
CP062013.1	27469064	ACTCCCATTACATGCTA	27469080
CP062013.1	49312549	CCATTACATGCTATT	49312563
CP055245.1	63604109	CATTACATGCTATTCAA	63604093
CP055245.1	66937583	ATTACATGCTATTCAA	66937598
CP055244.1	34789223	CATTACATGCTATTCAA	34789207
CP055244.1	42996857	CATTACATGCTATTCA	42996842
CP055241.1	31712695	CATTACATGCTATTCAA	31712679
CP055237.1	55508622	CATTACATGCTATTCAA	55508606
LR877218.1	24591882	ACTCCCATTACATGCTA	24591898

PRODUCTS OF SIGNIFICANT ALIGNMENTS (CDS FEATURE)

- **DNA repair exonuclease** GenBank
- **LOW QUALITY PROTEIN: xin actin-binding repeat-containing protein 2-like** GenBank
- **ADP-ribosylation factor GTPase-activating protein AGD2-like** GenBank
- **ADP-ribosylation factor GTPase-activating protein AGD4-like** GenBank
- **Anthranilate phosphoribosyltransferase** GenBank
- **Hypothetical protein** GenBank
- **Disease resistance RPP13-like protein** GenBank
- **Uncharacterized protein LOC9305153** GenBank; GenBank
- **Protein lethal(2)k10201** GenBank
- **Ammonia monooxygenase** GenBank

SEQUENCE N° 3/3 (T=2; C=0; G=3; A=1)

Query: **ACTCCCATTACATGCTATTCAA**

Search Parameters	
Program	blastn
Word size	7
Expect value	1000
Hitlist size	250
Match/Mismatch scores	1,-3
Gapcosts	5,2
Filter string	F
Genetic Code	1

247 sequences selected:

Accession Seq	Description	Comomm Name	Max Score	Total Score	Query C.	E value	Per. Id.	Acc. Len
.....	.....	.....	.....	.....	.....	.....	.....	.....
CP054816.1	Bacillus anthracis strain FDAARGOS_695 chromosome	Bacillus anthracis	32.2	32.2	72%	138	100.00%	5135792
CP054800.1	Bacillus anthracis strain FDAARGOS_702 chromosome	Bacillus anthracis	32.2	32.2	72%	138	100.00%	5272559
CP054797.1	Bacillus anthracis strain FDAARGOS_703 chromosome	Bacillus anthracis	32.2	32.2	72%	138	100.00%	5261520
CP050973.1	Bacillus anthracis strain FDAARGOS_699 chromosome	Bacillus anthracis	32.2	32.2	72%	138	100.00%	5227345
CP050972.1	Bacillus anthracis strain FDAARGOS_697 chromosome	Bacillus anthracis	32.2	32.2	72%	138	100.00%	5240765
CP050971.1	Bacillus anthracis strain FDAARGOS_696 chromosome	Bacillus anthracis	32.2	32.2	72%	138	100.00%	5255974
CP050970.1	Bacillus anthracis strain FDAARGOS_694 chromosome	Bacillus anthracis	32.2	32.2	72%	138	100.00%	5251490
CP064079.1	Bacillus paranthracis strain EFR-4 chromosome	Bacillus paranthracis	32.2	32.2	72%	138	100.00%	5315901
CP060194.1	Bacillus anthracis strain Kanchipuram chromosome	Bacillus anthracis	32.2	32.2	72%	138	100.00%	5227419
LR862374.1 1/2 2/2 3/2 3/5 6/1 7/2 8/1 9/1 10/1	*Pipistrellus pipistrellus genome assembly, chromosome: 18*	common pipistrelle	32.2	62.4	90%	138	95.00%	39450869
LR862361.1 1/2 2/2 3/2 3/5 6/1 7/2 8/1 9/1 10/1	*Pipistrellus pipistrellus genome assembly, chromosome: 5*	common pipistrelle	32.2	32.2	72%	138	100.00%	94929986

SEQUENCE N° 3/4

(T=2; C=0; G=3; A=1)

SEQUENCE N° 3/4 (T=2; C=0; G=3; A=1)

Original Sequence	**G**	**T**	**G**	**A**	**A**	**A**	**T**	**G**	**G**	**T**	**C**	**A**	**T**	**G**	**T**	**G**	**T**	**G**	**G**	**C**	**G**	**G**
Trend n° 3 (T=2; C=0; G=3; A=1)	**3**	**2**	**3**	**1**	**1**	**1**	**2**	**3**	**3**	**2**	**0**	**1**	**2**	**3**	**2**	**3**	**2**	**3**	**3**	**0**	**3**	**3**
Trend n° 3 (I i D d)	**-**	**D**	**I**	**D**	**d**	**d**	**I**	**I**	**i**	**D**	**D**	**I**	**I**	**I**	**D**	**I**	**D**	**I**	**i**	**D**	**I**	**i**
Trend n° 3/4	**1**	**0**	**1**	**0**	**0**	**0**	**1**	**2**	**2**	**1**	**0**	**1**	**2**	**3**	**0**	**3**	**0**	**1**	**1**	**0**	**1**	**1**
Sequence n° 3/4 (72,73%)	**A**	**C**	**A**	**C**	**C**	**C**	**A**	**T**	**T**	**A**	**C**	**A**	**T**	**G**	**C**	**G**	**C**	**A**	**A**	**C**	**A**	**A**

Query: **ACACCCATTACATGCGCAACAA**

Search Parameters	
Program	blastn
Word size	7
Expect value	1000
Hitlist size	100
Match/Mismatch scores	1,-3
Gapcosts	5,2
Filter string	F
Genetic Code	1

99 sequences selected:

Accession Seq	Description	Comomm Name	Max Score	Total Score	Query C.	E value	Per. Id.	Acc. Len
NC_047815.1	Erwinia phage vB_EamM_Yoloswag, complete genome	Erwinia phage vB_EamM_Yoloswag	36.2	36.2	81%	8.9	100.00%	259700
LR697116.1 1/2 2/2 3/1 3/3 8/1	Chanos chanos genome assembly, chromosome: 11	milkfish	36.2	36.2	81%	8.9	100.00%	25523583
LR132003.1 1/2 3/3 3/6	Gouania willdenowi genome assembly, chromosome: 3	blunt-snouted clingfish	36.2	36.2	81%	8.9	100.00%	46057556
XM_022376476.1	PREDICTED: Drosophila obscura protein O-mannosyltransferase 1 (LOC111080699), mRNA	Drosophila obscura	36.2	36.2	81%	8.9	100.00%	3834
LN554852.1	Moritella viscosa genome assembly MVIS1, chromosome : 1	Moritella viscosa	36.2	36.2	81%	8.9	100.00%	5086074
AB970869.1	Uncultured archaeon mcrA gene for methyl coenzyme-M reductase subunit A, partial cds, clone: m-2-10STMZ-128	uncultured archaeon	36.2	36.2	81%	8.9	100.00%	482
MK645239.1	Drosophila C virus isolate EB, complete genome	Drosophila C virus	34.2	34.2	95%	35	95.24%	9062
MK645238.1	Drosophila C virus isolate CYG, complete genome	Drosophila C virus	34.2	34.2	95%	35	95.24%	9062
LR697106.1 1/2 2/2 3/1 3/3 8/1	Chanos chanos genome assembly, chromosome: 1	milkfish	34.2	64.4	81%	35	100.00%	63121898
XM_029577608.1	PREDICTED: Rhinatrema bivittatum mitochondrial ribosomal protein S30 (MRPS30), transcript variant X2, mRNA	two-lined caecilian	34.2	34.2	77%	35	100.00%	1263
XM_029577603.1	PREDICTED: Rhinatrema bivittatum mitochondrial ribosomal protein S30 (MRPS30), transcript variant X1, mRNA	two-lined caecilian	34.2	34.2	77%	35	100.00%	5071
LR584407.1 1/2 2/2 3/3 3/6 8/1	Salmo trutta genome assembly, chromosome: 8	river trout	34.2	34.2	77%	35	100.00%	51188276
LR535816.1	Denticeps clupeoides genome assembly, chromosome: 4	denticle herring	34.2	34.2	77%	35	100.00%	35364899
XR_003294056.1	PREDICTED: Carassius auratus uncharacterized LOC113116706 (LOC113116706), ncRNA	goldfish	34.2	34.2	77%	35	100.00%	4417
CP034970.1	Vibrio chagasii strain	Vibrio chagasii	34.2	34.2	95%	35	95.24%	3563604

	ECSMB14107 chromosome 1, complete sequence							
XM_010099710.2	PREDICTED: Morus notabilis DNA-directed **RNA polymerases II, IV and V subunit 11** (LOC21387719), mRNA	Morus notabilis	34.2	34.2	77%	35	100.00%	777
XR_002760741.1	PREDICTED: Seriola lalandi dorsalis uncharacterized LOC111654400 (LOC111654400), ncRNA	Seriola lalandi dorsalis	34.2	34.2	77%	35	100.00%	1769
XM_018247752.1	PREDICTED: Xenopus laevis sterile alpha motif domain-containing protein 9-like (LOC108708729), mRNA	African clawed frog	34.2	34.2	77%	35	100.00%	5335
CP016603.1	Comamonas aquatica strain CJG genome	Comamonas aquatica	34.2	34.2	77%	35	100.00%	3764434
CP012069.2	Olsenella sp. oral taxon 807 strain F0089, complete genome	Olsenella sp. oral taxon 807	34.2	34.2	77%	35	100.00%	3213585
XM_037790127.1	PREDICTED: Sebastes umbrosus sacsin-like (LOC119500421), mRNA	honeycomb rockfish	34.2	34.2	77%	35	100.00%	9521
CP062014.1	Macrobrachium nipponense isolate FS-2020 chromosome 14	Macrobrachium nipponense	34.2	34.2	77%	35	100.00%	99588160
CP003282.1	Spirochaeta africana DSM 8902, complete genome	Spirochaeta africana DSM 8902	34.2	34.2	77%	35	100.00%	3285855
AC211727.1	Populus trichocarpa clone ACSB2229-H14, complete sequence	black cottonwood	34.2	34.2	77%	35	100.00%	38255
AF014388.1	Drosophila C virus strain EB, complete genome	Drosophila C virus	34.2	34.2	95%	35	95.24%	9264
LR790227.1	Phallusia mammillata mRNA for superkiller viralicidic activity 2-like 2 (Skiv2l2 gene)	Phallusia mammillata	34.2	34.2	77%	35	100.00%	3384
CP049246.1	Sphingobacterium lactis strain DSM 22361 chromosome, complete genome	Sphingobacterium lactis	34.2	34.2	95%	35	95.24%	4030188
CP046457.1	Candidatus Syntrophocurvum alkaliphilum isolate B(2M) chromosome, complete genome	Candidatus Syntrophocurvum alkaliphilum	34.2	34.2	77%	35	100.00%	2360781
MH138257.1	Cydia pomonella cytochrome P450 4V2 mRNA, complete cds	codling moth	34.2	34.2	77%	35	100.00%	999
CP020105.1 **1/2 2/2 3/3 8/1**	Spirosoma rigui strain KCTC 12531 chromosome, complete genome	Spirosoma rigui	34.2	34.2	77%	35	100.00%	5828404
CP011542.1	Corynebacterium mustelae strain DSM 45274, complete genome	Corynebacterium mustelae	34.2	34.2	77%	35	100.00%	3391554
XR_912820.1	PREDICTED: Ceratosolen solmsi marchali uncharacterized LOC105360997 (LOC105360997), transcript variant X2, misc_RNA	Ceratosolen solmsi marchali	34.2	34.2	77%	35	100.00%	1433
XM_011498069.1	PREDICTED: Ceratosolen solmsi marchali uncharacterized LOC105360997 (LOC105360997), transcript variant X1, mRNA	Ceratosolen solmsi marchali	34.2	34.2	77%	35	100.00%	1609
XM_007384682.1	Punctularia strigosozonata HHB-11173 SS5 ATP-dependent DNA helicase (PUNSTDRAFT_103671), partial mRNA	Punctularia strigosozonata HHB-11173 SS5	34.2	34.2	77%	35	100.00%	1572
AM454937.2	Vitis vinifera contig VV78X056749.7, whole genome shotgun sequence	wine grape	34.2	34.2	77%	35	100.00%	51851
LR812075.1	Danio rerio genome assembly, chromosome: 13	zebrafish	32.2	32.2	72%	138	100.00%	53214489
CP052041.1	Fusarium oxysporum Fo47 chromosome IV	Fusarium oxysporum Fo47	32.2	32.2	72%	138	100.00%	4731052
XM_034156366.1	Diutina rugosa uncharacterized protein (DIURU_003588), partial mRNA	Diutina rugosa	32.2	32.2	90%	138	95.00%	14265
CP050589.1 **1/2 2/2 3/3 3/5 3/6 5/1 10/1**	Canis lupus familiaris breed Labrador retriever chromosome 14a	dog	32.2	32.2	72%	138	100.00%	62855878
CP050623.1 **1/2 2/2 3/3 3/5 3/6 5/1 10/1**	Canis lupus familiaris breed Labrador retriever chromosome 14b	dog	32.2	32.2	72%	138	100.00%	62857718
AP022695.1	Epinephelus fuscoguttatus DNA, LG21, complete sequence	brown-marbled grouper	32.2	32.2	72%	138	100.00%	39642397

AP022690.1	Epinephelus fuscoguttatus DNA, LG16, complete sequence	brown-marbled grouper	32.2	32.2	72%	138	100.00%	42848118
CP032675.1	Rhodococcus rhodochrous strain ATCC BAA870 chromosome, complete genome	Rhodococcus rhodochrous	32.2	32.2	90%	138	95.00%	5370536
XM_031178426.1	Fusarium oxysporum NRRL 32931 uncharacterized protein (FOYG_04454), mRNA	Fusarium oxysporum NRRL 32931	32.2	32.2	72%	138	100.00%	2160
CP032375.1	Trichinella spiralis isolate TY2 chromosome 1	Trichinella spiralis	32.2	62.4	72%	138	100.00%	19988360
CP032372.1	Trichinella spiralis isolate Shisler1 chromosome 1	Trichinella spiralis	32.2	94.6	72%	138	100.00%	19956450
CP032369.1	Trichinella spiralis isolate g279 chromosome 1	Trichinella spiralis	32.2	94.6	72%	138	100.00%	19955631
LR699094.1	Asterias rubens genome assembly, chromosome: 3	European starfish	32.2	32.2	72%	138	100.00%	22888987
CP042389.1	Lactobacillus curvatus strain CBA3617 chromosome, complete genome	Lactobacillus curvatus	32.2	32.2	72%	138	100.00%	1915352
LR537136.1	**Sparus aurata genome assembly, chromosome: 16**	gilthead seabream	32.2	32.2	72%	138	100.00%	30078578
CP035801.1	Haemonchus contortus strain NZ_Hco_NP chromosome 2	barber pole worm	32.2	62.4	77%	138	100.00%	77295023
LR607350.1	Klebsiella oxytoca strain 4928STDY7071151 genome assembly, chromosome: 1	Klebsiella oxytoca	32.2	32.2	72%	138	100.00%	6835957
CP027858.1	Streptomyces clavuligerus strain ATCC 27064 chromosome, complete genome	Streptomyces clavuligerus	32.2	32.2	72%	138	100.00%	6748591
XM_021039339.2	PREDICTED: Exaiptasia pallida U6 snRNA-associated Sm-like protein LSm8 (LOC110233967), mRNA	Exaiptasia diaphana	32.2	32.2	72%	138	100.00%	1300
LR584401.1	Rhinatrema bivittatum genome assembly, chromosome: 15	two-lined caecilian	32.2	32.2	72%	138	100.00%	80992113
LR535831.1	Denticeps clupeoides genome assembly, chromosome: 19	denticle herring	32.2	32.2	72%	138	100.00%	19707071
LR890312.1	Klebsiella oxytoca isolate MSB1_10D-sc-2280340 genome assembly, chromosome: 1	Klebsiella oxytoca	32.2	32.2	72%	138	100.00%	5812602
CP052886.1	Fusarium oxysporum isolate 170 chromosome 2	Fusarium oxysporum	32.2	32.2	72%	138	100.00%	4546955
LR880646.1	Poecilia reticulata genome assembly, chromosome: 2	guppy	32.2	62.4	95%	138	100.00%	45373629
LR877148.1	Angomonas deanei strain Crithidia deanei Carvalho (ATCC PRA-265) genome assembly, chromosome: 04	Angomonas deanei	32.2	32.2	72%	138	100.00%	1175096
CP057330.1	Klebsiella oxytoca strain RHB30-C02 chromosome, complete genome	Klebsiella oxytoca	32.2	32.2	72%	138	100.00%	6051148
CP056567.1	Klebsiella oxytoca strain RHBSTW-00373 chromosome, complete genome	Klebsiella oxytoca	32.2	32.2	72%	138	100.00%	6091081
CP045504.1	Desulfovibrio sulfodismutans DSM 3696 chromosome, complete genome	Desulfovibrio sulfodismutans DSM 3696	32.2	32.2	72%	138	100.00%	4388462
LR812606.1	Danio rerio genome assembly, chromosome: 13	zebrafish	32.2	32.2	72%	138	100.00%	52210330
LR812581.1	Danio rerio strain Nadia (NA) genome assembly, chromosome: 13	zebrafish	32.2	32.2	72%	138	100.00%	54512446
LR812556.1	Danio rerio strain Cooch Behar (CB) genome assembly, chromosome: 13	zebrafish	32.2	32.2	72%	138	100.00%	52525257
CP053263.1	Fusarium oxysporum Fo5176 chromosome 4	Fusarium oxysporum Fo5176	32.2	32.2	72%	138	100.00%	5006977
LR792560.1	Pseudochaenichthys georgianus genome assembly, chromosome: 15	South Georgia icefish	32.2	32.2	72%	138	100.00%	39528010
LR778311.1	Raphanus sativus genome assembly, chromosome: 2	radish	32.2	32.2	72%	138	100.00%	53721154
AP022706.1	Plectropomus leopardus DNA, chromosome 7, nearly complete sequence	leopard coralgrouper	32.2	94.6	90%	138	100.00%	36063594
CP035042.1	Halomonas sp. BC-M4-5 chromosome, complete genome	Halomonas sp. BC-M4-5	32.2	32.2	90%	138	95.00%	5015118
CP045427.1	Shewanella sp. YLB-09 chromosome, complete genome	Shewanella sp. YLB-09	32.2	32.2	72%	138	100.00%	6225487
CP032580.1	Gossypium turneri isolate D10-2 chromosome D10_10	Gossypium turneri	32.2	32.2	72%	138	100.00%	63881757
	Chanos chanos genome	milkfish	32.2	92.7	81%	138	100.00%	49187548

LR697112.1 1/2 2/2 3/1 3/3 8/1	assembly, chromosome: 7								
XM_029861887.1	PREDICTED: Aedes albopictus progestin and adipoQ receptor family member 3-like (LOC109432372), transcript variant X3, mRNA	Asian tiger mosquito	32.2	32.2	72%	138	100.00%	1991
XM_029861886.1	PREDICTED: Aedes albopictus progestin and adipoQ receptor family member 3-like (LOC109432372), transcript variant X2, mRNA	Asian tiger mosquito	32.2	32.2	72%	138	100.00%	1836
XM_029861885.1	PREDICTED: Aedes albopictus progestin and adipoQ receptor family member 3-like (LOC109432372), transcript variant X1, mRNA	Asian tiger mosquito	32.2	32.2	72%	138	100.00%	1416
XM_020635225.2	PREDICTED: Labrus bergylta protein THEM6-like (LOC109985015), mRNA	ballan wrasse	32.2	32.2	72%	138	100.00%	1880
CP039451.1	Psychroserpens sp. NJDZ02 chromosome, complete genome	Psychroserpens sp. NJDZ02	32.2	32.2	90%	138	95.00%	4319537
LR584428.1 1/2 2/2 3/3 3/6 8/1	Salmo trutta genome assembly, chromosome: 13	river trout	32.2	32.2	72%	138	100.00%	91488822
CP038470.1	Dehalococcoides mccartyi strain FL2 chromosome, complete genome	Dehalococcoides mccartyi	32.2	32.2	72%	138	100.00%	1422358
XM_028542838.1	PREDICTED: Dendronephthya gigantea NACHT, LRR and PYD domains-containing protein 12-like (LOC114522204), mRNA	Dendronephthya gigantea	32.2	32.2	72%	138	100.00%	3953
LR899426.1	Corylus avellana genome assembly, chromosome: ca4	Corylus avellana	32.2	32.2	72%	138	100.00%	36845065
CP045503.1	Shewanella sp. YLB-08 chromosome, complete genome	Shewanella sp. YLB-08	32.2	32.2	72%	138	100.00%	5785459
XM_037424772.1	PREDICTED: Rhipicephalus microplus uncharacterized LOC119173993 (LOC119173993), mRNA	southern cattle tick	32.2	32.2	72%	138	100.00%	9238
XM_037323409.1	Colletotrichum aenigma Cytochrome P450 4F4 (Cyp4f4), partial mRNA	Colletotrichum aenigma	32.2	32.2	72%	138	100.00%	1659
XM_037319458.1	Colletotrichum aenigma (S)-mandelate dehydrogenase (CGCA056_v004203), partial mRNA	Colletotrichum aenigma	32.2	32.2	72%	138	100.00%	2187
XM_036640383.1	Colletotrichum siamense Cytochrome P450 4F8 (CYP4F8), partial mRNA	Colletotrichum siamense	32.2	32.2	72%	138	100.00%	1659
XM_036686071.1	Fusarium subglutinans fad binding domain protein (FSUBG_7736), partial mRNA	Fusarium subglutinans	32.2	32.2	90%	138	95.00%	1515
CP053382.1	Halomonas sp. MCCC 1A13316 chromosome, complete genome	Halomonas sp. MCCC 1A13316	32.2	32.2	90%	138	95.00%	4490931
AP024096.1	Ktedonobacteria bacterium brp13 DNA, complete genome	Ktedonobacteria bacterium brp13	32.2	32.2	72%	138	100.00%	8879929
CP052898.1	Fusarium oxysporum f. sp. koae isolate 44 chromosome 2	Fusarium oxysporum f. sp. koae	32.2	32.2	72%	138	100.00%	4660957
CP027557.1	Rhodococcus rhodochrous strain BX2 chromosome, complete genome	Rhodococcus rhodochrous	32.2	32.2	90%	138	95.00%	5459556
XM_036046082.1	PREDICTED: Anopheles stephensi aminopeptidase N (LOC118507521), mRNA	Asian malaria mosquito	32.2	32.2	72%	138	100.00%	4309
XM_035959857.1	PREDICTED: Zea mays B3 domain-containing protein IDEF1 (LOC103628923), transcript variant X4, mRNA	Zea mays	32.2	32.2	90%	138	95.00%	1726
XM_020539753.3	PREDICTED: Zea mays B3 domain-containing protein IDEF1 (LOC103628923), transcript variant X3, mRNA	Zea mays	32.2	32.2	90%	138	95.00%	1802
XM_020539752.3	PREDICTED: Zea mays B3 domain-containing protein IDEF1 (LOC103628923), transcript variant X2, mRNA	Zea mays	32.2	32.2	90%	138	95.00%	1784
XM_020539751.2	PREDICTED: Zea mays B3 domain-containing protein IDEF1 (LOC103628923), transcript variant X1, mRNA	Zea mays	32.2	32.2	90%	138	95.00%	1862

CP049139.1	Herbaspirillum frisingense strain IAC152 chromosome, complete genome	Herbaspirillum frisingense	32.2	32.2	72%	138	100.00%	5548491

ALIGNMENT VIEW FLAT QUERY-ANCHORED WITH LETTERS FOR IDENTITIES

Query range 1: 1 to 22

Query	**1**	**ACACCCATTACATGCGCAACAA**	**22**
NC_047815.1	21660	CCCATTACATGCGCAACA	21643
LR697116.1	225932	CCATTACATGCGCAACAA	225949
LR132003.1	14505064	CACCCATTACATGCGCAA	14505081
XM_022376476.1	1593	CCCATTACATGCGCAACA	1610
LN554852.1	844798	CCATTACATGCGCAACAA	844815
AB970869.1	275	CCATTACATGCGCAACAA	292
MK645239.1	8120	CACCCATTACTTGCGCAACAA	8100
MK645238.1	8120	CACCCATTACTTGCGCAACAA	8100
LR697106.1	33711251	CCATTACATGCGCAACA	33711235
LR697106.1	17761206	TTACATGCGCAACAA	17761192
XM_029577608.1	35	CCATTACATGCGCAACA	19
XM_029577603.1	32	CCATTACATGCGCAACA	16
LR584407.1	3836456	CACCCATTACATGCGCA	3836440
LR535816.1	24999106	CCATTACATGCGCAACA	24999090
XR_003294056.1	712	CATTACATGCGCAACAA	696
CP034970.1	1384417	ACACCAATTACATGCGCAACA	1384397
XM_010099710.2	633	CCATTACATGCGCAACA	617
XR_002760741.1	799	CATTACATGCGCAACAA	783
XM_018247752.1	2875	CATTACATGCGCAACAA	2859
CP016603.1	997773	ACACCCATTACATGCGC	997757
CP012069.2	144002	CATTACATGCGCAACAA	143986
XM_037790127.1	3422	CATTACATGCGCAACAA	3406
CP062014.1	14890328	CCATTACATGCGCAACA	14890312
CP003282.1	93081	CCCATTACATGCGCAAC	93065
AC211727.1	8264	CATTACATGCGCAACAA	8248
AF014388.1	8184	CACCCATTACTTGCGCAACAA	8164
LR790227.1	887	CATTACATGCGCAACAA	903
CP049246.1	3145168	ACACCTATTACATGCGCAACA	3145188
CP046457.1	495369	CATTACATGCGCAACAA	495385
MH138257.1	608	CATTACATGCGCAACAA	624
CP020105.1	2283394	CATTACATGCGCAACAA	2283410
CP011542.1	3342868	CCCATTACATGCGCAAC	3342884
XR_912820.1	974	CATTACATGCGCAACAA	990
XM_011498069.1	974	CATTACATGCGCAACAA	990
XM_007384682.1	365	ACACCCATTACATGCGC	381
AM454937.2	41714	CCATTACATGCGCAACA	41730
LR812075.1	12296124	ATTACATGCGCAACAA	12296109
CP052041.1	3582335	CCCATTACATGCGCAA	3582320
XM_034156366.1	3984	ACCCATTACTTGCGCAACAA	3965
CP050589.1	10488151	CCATTACATGCGCAAC	10488136

CP050623.1	10487775	CCATTACATGCGCAAC	10487760
AP022695.1	23257140	CCCATTACATGCGCAA	23257125
AP022690.1	15029198	ACCCATTACATGCGCA	15029183
CP032675.1	1712573	CACCCATTACATGCTCAACA	1712554
XM_031178426.1	2061	CCCATTACATGCGCAA	2046
CP032375.1	5326937	ATTACATGCGCAACAA	5326922
CP032375.1	17112705	TTACATGCGCAACAA	17112691
CP032372.1	5318831	ATTACATGCGCAACAA	5318816
CP032372.1	5328027	ATTACATGCGCAACAA	5328012
CP032372.1	17083568	TTACATGCGCAACAA	17083554
CP032369.1	5317970	ATTACATGCGCAACAA	5317955
CP032369.1	5327176	ATTACATGCGCAACAA	5327161
CP032369.1	17083083	TTACATGCGCAACAA	17083069
LR699094.1	11771640	CACCCATTACATGCGC	11771625
CP042389.1	772295	ATTACATGCGCAACAA	772280
LR537136.1	25217121	CATTACATGCGCAACA	25217106
CP035801.1	56629303	CACCCATTACATGCGC	56629288
CP035801.1	62158347	ACACCCATTACATGC	62158333
LR607350.1	737908	ATTACATGCGCAACAA	737893
CP027858.1	6473139	CCATTACATGCGCAAC	6473124
XM_021039339.2	621	ATTACATGCGCAACAA	606
LR584401.1	61207559	ATTACATGCGCAACAA	61207544
LR535831.1	12835665	ATTACATGCGCAACAA	12835650
LR890312.1	2041547	ATTACATGCGCAACAA	2041532
CP052886.1	833445	CCCATTACATGCGCAA	833430
LR880646.1	34989247	CATTACATGCGCAACA	34989232
LR880646.1	29182986	ACACCCATTACATGC	29183000
LR877148.1	262317	ATTACATGCGCAACAA	262302
CP057330.1	2169365	ATTACATGCGCAACAA	2169350
CP056567.1	2180977	ATTACATGCGCAACAA	2180962
CP045504.1	978787	CACCCATTACATGCGC	978772
LR812606.1	13576925	ATTACATGCGCAACAA	13576910
LR812581.1	11898295	ATTACATGCGCAACAA	11898280
LR812556.1	14423538	ATTACATGCGCAACAA	14423523
CP053263.1	1029999	CCCATTACATGCGCAA	1030014
LR792560.1	29256488	CATTACATGCGCAACA	29256503
LR778311.1	48051809	ATTACATGCGCAACAA	48051824
AP022706.1	33783454	ACCCATTACATGCGCA	33783469
AP022706.1	33804113	ACCCATTACATGCGCA	33804128
AP022706.1	5021490	TTACATGCGCAACAA	5021476
CP035042.1	3998223	CACCCAGTACATGCGCAACA	3998242
CP045427.1	3331462	ATTACATGCGCAACAA	3331477
CP032580.1	40222466	CATTACATGCGCAACA	40222481
LR697112.1	7022426	CCATTACATGCGCAAC	7022441
LR697112.1	38998851	TTACATGCGCAACAA	38998837
LR697112.1	39042207	TTACATGCGCAACAA	39042193
XM_029861887.1	935	CCCATTACATGCGCAA	950
XM_029861886.1	360	CCCATTACATGCGCAA	375
XM_029861885.1	360	CCCATTACATGCGCAA	375
XM_020635225.2	338	CCATTACATGCGCAAC	353
CP039451.1	1153313	ACCCATTTCATGCGCAACAA	1153332
LR584428.1	69879931	ACCCATTACATGCGCA	69879946
CP038470.1	1202561	ACCCATTACATGCGCA	1202576
XM_028542838.1	617	ACACCCATTACATGCG	632
LR899426.1	3011913	ACCCATTACATGCGCA	3011928
CP045503.1	2956236	ATTACATGCGCAACAA	2956251
XM_037424772.1	7105	ATTACATGCGCAACAA	7120
XM_037323409.1	461	ATTACATGCGCAACAA	476
XM_037319458.1	748	ATTACATGCGCAACAA	763
XM_036640383.1	461	ATTACATGCGCAACAA	476

XM_036686071.1	1106	ACACCCATTTCATGCGCAAC	1125
CP053382.1	3523668	ACCCAGTACATGCGCAACAA	3523687
AP024096.1	6635369	ACCCATTACATGCGCA	6635384
CP052898.1	1085581	CCCATTACATGCGCAA	1085596
CP027557.1	4946093	CACCCATTACATGCTCAACA	4946112
XM_036046082.1	3912	ATTACATGCGCAACAA	3927
XM_035959857.1	734	ACCCATTACATGCTCAACAA	753
XM_020539753.3	729	ACCCATTACATGCTCAACAA	748
XM_020539752.3	722	ACCCATTACATGCTCAACAA	741
XM_020539751.2	741	ACCCATTACATGCTCAACAA	760
CP049139.1	5187934	CACCCATTACATGCGC	5187949

PRODUCTS OF SIGNIFICANT ALIGNMENTS (CDS FEATURE)

- **Hypothetical protein** GenBank; GenBank; GenBank; GenBank
- **Protein O-mannosyltransferase 1** GenBank
- **Molybdopterin biosynthesis protein MoeA** GenBank
- **Methyl coenzyme-M reductase subunit A** GenBank
- **Capsid polyprotein** GenBank; GenBank; GenBank
- **Sterile alpha motif domain-containing protein 9-like** GenBank
- **Sacsin-like** GenBank
- **NADH dehydrogenase, FAD-containing subunit** GenBank
- **Superkiller viralicidic activity 2-like 2** GenBank
- **Cytochrome P450 4V2** GenBank
- **AAA ATPase domain** GenBank
- **Uncharacterized protein LOC105360997** GenBank
- **ATP-dependent DNA helicase** GenBank
- **Uncharacterized protein** GenBank
- **Formate dehydrogenase** GenBank; GenBank
- **Glycosyltransferase family 8 protein** GenBank
- **Uncharacterized protein family UPF0054,putative** GenBank
- **Efflux RND transporter periplasmic adaptor subunit** GenBank; GenBank
- **Peptidoglycan DD-metalloendopeptidase family protein** GenBank
- **Choline-sulfatase** GenBank; GenBank
- **HAD hydrolase-like protein** GenBank; GenBank
- **Progestin and adipoQ receptor family member 3-like** GenBank; GenBank; GenBank
- **Protein THEM6-like** GenBank
- **T9SS type A sorting domain-containing protein** GenBank
- **Glycerol-3-phosphate acyltransferase** GenBank
- **NACHT, LRR and PYD domains-containing protein 12-like** GenBank
- **Cytochrome P450 4F4** GenBank
- **(S)-mandelate dehydrogenase** GenBank
- **Cytochrome P450 4F8** GenBank
- **Fad binding domain protein** GenBank
- **HAMP domain-containing histidine kinase** GenBank
- **Aminopeptidase N** GenBank
- **B3 domain-containing protein IDEF1** GenBank; GenBank; GenBank; GenBank
- **Thiolase family protein** GenBank

SEQUENCE N° 3/5

(T=2; C=0; G=3; A=1)

SEQUENCE N° 3/5 (T=2; C=0; G=3; A=1)

Original Sequence	**G**	**T**	**G**	**A**	**A**	**A**	**T**	**G**	**G**	**T**	**C**	**A**	**T**	**G**	**T**	**G**	**T**	**G**	**G**	**C**	**G**	**G**
Trend n° 3 (T=2; C=0; G=3; A=1)	**3**	**2**	**3**	**1**	**1**	**1**	**2**	**3**	**3**	**2**	**0**	**1**	**2**	**3**	**2**	**3**	**2**	**3**	**3**	**0**	**3**	**3**
Trend n° 3 (I i D d)	**-**	**D**	**I**	**D**	**d**	**d**	**I**	**I**	**i**	**D**	**D**	**I**	**I**	**I**	**D**	**I**	**D**	**I**	**i**	**D**	**I**	**i**
Trend n° 3/5	**2**	**0**	**2**	**0**	**0**	**0**	**1**	**2**	**2**	**1**	**0**	**1**	**2**	**3**	**1**	**3**	**1**	**2**	**2**	**0**	**1**	**1**
Sequence n° 3/5 (72,73%)	**T**	**C**	**T**	**C**	**C**	**C**	**A**	**T**	**T**	**A**	**C**	**A**	**T**	**G**	**A**	**G**	**A**	**T**	**T**	**C**	**A**	**A**

Query: **TCTCCCATTACATGAGATTCAA**

Search Parameters	
Program	blastn
Word size	7
Expect value	1000
Hitlist size	100
Match/Mismatch scores	1,-3
Gapcosts	5,2
Filter string	F
Genetic Code	1

100 sequences selected:

Accession Seq	Description	Comomm Name	Max Score	Total Score	Query C.	E value	Per. Id.	Acc. Len
XM_031253612.1	**PREDICTED: Ipomoea triloba auxin transport protein BIG (LOC116013716), mRNA**	trilobed morning glory	40.1	40.1	90%	0.57	100.00%	15549
CP025662.1	**Ipomoea triloba cultivar NCNSP0323 chromosome 3**	trilobed morning glory	40.1	40.1	90%	0.57	100.00%	33448119
CP025646.1	**Ipomoea trifida cultivar NCNSP0306 chromosome 3**	Ipomoea trifida	40.1	40.1	90%	0.57	100.00%	28619729
XM_019329813.1	PREDICTED: Ipomoea nil auxin transport protein BIG (LOC109180310), mRNA	Japanese morning glory	40.1	40.1	90%	0.57	100.00%	15622
CP032554.1	Gossypium raimondii isolate D5-4 chromosome D5_02	Gossypium raimondii	38.2	38.2	86%	2.2	100.00%	46322098
CP025781.1	Iodobacter fluviatilis strain PCH194 chromosome, complete genome	Iodobacter fluviatilis	38.2	38.2	86%	2.2	100.00%	4441511
LT669788.1	Arabis alpina genome assembly, chromosome: 1	gray rockcress	38.2	38.2	86%	2.2	100.00%	29724344
AL356461.15 **1/2 3/3**	Human DNA sequence from clone RP11-180D15 on chromosome Xq13.2-21.2, complete sequence	human	38.2	38.2	86%	2.2	100.00%	127615
XM_030979643.1	PREDICTED: Strongylocentrotus purpuratus A disintegrin and metalloproteinase with thrombospondin motifs 3 (LOC579423), transcript variant X3, mRNA	purple sea urchin	36.2	36.2	81%	8.9	100.00%	6189
XM_011679278.2	PREDICTED: Strongylocentrotus purpuratus A disintegrin and metalloproteinase with thrombospondin motifs 3 (LOC579423), transcript variant X2, mRNA	purple sea urchin	36.2	36.2	81%	8.9	100.00%	6199
XM_011679285.2	PREDICTED: Strongylocentrotus purpuratus A disintegrin and metalloproteinase with thrombospondin motifs 3 (LOC579423), transcript variant X1, mRNA	purple sea urchin	36.2	36.2	81%	8.9	100.00%	6198
CP035802.1	Haemonchus contortus strain NZ_Hco_NP chromosome 3	barber pole worm	36.2	102	90%	8.9	100.00%	83970805
LR606180.1 **1/2 3/2 3/3 9/1**	Aquila chrysaetos chrysaetos genome assembly, chromosome: Z	Aquila chrysaetos chrysaetos	36.2	219	100%	8.9	100.00%	88216475
LR597476.1 **1/2 2/2 3/1 3/6**	Sphaeramia orbicularis genome assembly, chromosome: 19	orbiculate cardinalfish	36.2	96.6	95%	8.9	100.00%	53342843

CP035095.1	Bacillus megaterium NBRC 15308 = ATCC 14581 plasmid unnamed1, complete sequence	Bacillus megaterium NBRC 15308 = ATCC 14581	36.2	36.2	81%	8.9	100.00%	301273
CP032250.1	Gossypioides kirkii chromosome KI_08	Gossypioides kirkii	36.2	96.6	81%	8.9	100.00%	39602670
CP032245.1	Gossypioides kirkii chromosome KI_2_4	Gossypioides kirkii	36.2	128	81%	8.9	100.00%	41238926
CP026261.1	Scophthalmus maximus chromosome 19	turbot	36.2	36.2	81%	8.9	100.00%	20258455
CP009921.1	Bacillus megaterium NBRC 15308 = ATCC 14581 plasmid pBMV_2, complete sequence	Bacillus megaterium NBRC 15308 = ATCC 14581	36.2	36.2	81%	8.9	100.00%	301264
MT261175.1	Uvularia grandiflora plastid, complete genome	Uvularia grandiflora	34.2	34.2	77%	35	100.00%	157025
MT261174.1	Tripladenia cunninghamii plastid, complete genome	Tripladenia cunninghamii	34.2	34.2	77%	35	100.00%	155652
MT261148.1	Androcymbium greuterocymbium plastid, complete genome	Androcymbium greuterocymbium	34.2	34.2	77%	35	100.00%	154804
CP050574.1 **1/2 2/2 3/3 3/4 3/6 5/1 10/1**	Canis lupus familiaris breed Labrador retriever chromosome 02a	dog	34.2	66.4	77%	35	100.00%	83829384
CP050570.1 **1/2 2/2 3/3 3/4 3/6 5/1 10/1**	Canis lupus familiaris breed Labrador retriever chromosome 26a	dog	34.2	34.2	77%	35	100.00%	38035513
CP050613.1 **1/2 2/2 3/3 3/4 3/6 5/1 10/1**	Canis lupus familiaris breed Labrador retriever chromosome 02b	dog	34.2	66.4	77%	35	100.00%	83851004
CP050610.1 **1/2 2/2 3/3 3/4 3/6 5/1 10/1**	Canis lupus familiaris breed Labrador retriever chromosome 26b	dog	34.2	34.2	77%	35	100.00%	38033444
AP022632.1	Lotus japonicus B-129 DNA, chromosome 4, complete sequence	Lotus japonicus	34.2	64.4	77%	35	100.00%	83711341
AP022704.1	Plectropomus leopardus DNA, chromosome 5, nearly complete sequence	leopard coralgrouper	34.2	64.4	77%	35	100.00%	37020500
XM_002936994.4	PREDICTED: Xenopus tropicalis recoverin, gene 1 (rcvrn.1), mRNA	tropical clawed frog	34.2	34.2	77%	35	100.00%	1736
LR722981.1	**Thalassophryne amazonica genome assembly, chromosome: 16**	Thalassophryne amazonica	34.2	98.6	77%	35	100.00%	92122960
CP042455.1	Kluyveromyces lactis strain CBS 2105 chromosome 1	Kluyveromyces lactis	34.2	34.2	77%	35	100.00%	1602874
XM_030198566.1	PREDICTED: Microcaecilia unicolor RNA binding motif protein 33 (RBM33), transcript variant X7, mRNA	Microcaecilia unicolor	34.2	34.2	77%	35	100.00%	7121
XM_030198557.1	PREDICTED: Microcaecilia unicolor RNA binding motif protein 33 (RBM33), transcript variant X6, mRNA	Microcaecilia unicolor	34.2	34.2	77%	35	100.00%	7130
XM_030198551.1	PREDICTED: Microcaecilia unicolor RNA binding motif protein 33 (RBM33), transcript variant X5, mRNA	Microcaecilia unicolor	34.2	34.2	77%	35	100.00%	7133
XM_030198543.1	PREDICTED: Microcaecilia unicolor RNA binding motif protein 33 (RBM33), transcript variant X4, mRNA	Microcaecilia unicolor	34.2	34.2	77%	35	100.00%	7250
XM_030198536.1	PREDICTED: Microcaecilia unicolor RNA binding motif protein 33 (RBM33), transcript variant X3, mRNA	Microcaecilia unicolor	34.2	34.2	77%	35	100.00%	7253
XM_030198529.1	PREDICTED: Microcaecilia unicolor RNA binding motif protein 33 (RBM33), transcript variant X2, mRNA	Microcaecilia unicolor	34.2	34.2	77%	35	100.00%	7253
XM_030198521.1	PREDICTED: Microcaecilia unicolor RNA binding motif protein 33 (RBM33), transcript variant X1, mRNA	Microcaecilia unicolor	34.2	34.2	77%	35	100.00%	7256
LR537130.1	**Sparus aurata genome assembly, chromosome: 10**	gilthead seabream	34.2	66.4	77%	35	100.00%	35512760
CP035801.1	Haemonchus contortus strain NZ_Hco_NP chromosome 2	barber pole worm	34.2	64.4	95%	35	95.24%	77295023
CP035800.1	Haemonchus contortus strain NZ_Hco_NP chromosome X	barber pole worm	34.2	34.2	77%	35	100.00%	41849371

LR606198.1 1/2 3/2 3/3 9/1	Aquila chrysaetos chrysaetos genome assembly, chromosome: 18	Aquila chrysaetos chrysaetos	34.2	66.4	100%	35	100.00%	28555659
LR597552.1	Myripristis murdjan genome assembly, chromosome: 3	pinecone soldierfish	34.2	64.4	100%	35	95.24%	47119472
XM_029315943.1	PREDICTED: Nylanderia fulva cytochrome c1-2, heme protein, mitochondrial-like (LOC114941085), transcript variant X2, mRNA	Nylanderia fulva	34.2	34.2	77%	35	100.00%	1393
XM_029315942.1	PREDICTED: Nylanderia fulva cytochrome c1-2, heme protein, mitochondrial-like (LOC114941085), transcript variant X1, mRNA	Nylanderia fulva	34.2	34.2	77%	35	100.00%	1979
CP032252.1	Gossypioides kirkii chromosome KI_10	Gossypioides kirkii	34.2	223	81%	35	100.00%	47257441
CP032251.1	Gossypioides kirkii chromosome KI_09	Gossypioides kirkii	34.2	64.4	77%	35	100.00%	38743931
CP032244.1	Gossypioides kirkii chromosome KI_01	Gossypioides kirkii	34.2	94.6	86%	35	100.00%	38174113
LR535841.1	Mastacembelus armatus genome assembly, chromosome: 9	zig-zag eel	34.2	64.4	100%	35	100.00%	25822229
LR132043.1	Anabas testudineus genome assembly, chromosome: 4	climbing perch	34.2	66.4	86%	35	100.00%	26090302
MK072386.1	Hyperionvirus sp. clone Hyperionvirus_4 genomic sequence	Hyperionvirus sp.	34.2	34.2	95%	35	95.24%	145918
XM_026573109.1	PREDICTED: Papaver somniferum 1-aminocyclopropane-1-carboxylate oxidase homolog 1-like (LOC113324819), mRNA	opium poppy	34.2	34.2	77%	35	100.00%	1439
CP032418.1	Paenisporosarcina sp. K2R23-3 chromosome, complete genome	Paenisporosarcina sp. K2R23-3	34.2	34.2	77%	35	100.00%	2537131
CP025672.1	**Ipomoea triloba cultivar NCNSP0323 chromosome 13**	trilobed morning glory	34.2	64.4	77%	35	100.00%	32066333
CP025656.1	**Ipomoea trifida cultivar NCNSP0306 chromosome 13**	Ipomoea trifida	34.2	34.2	77%	35	100.00%	22914374
LS997567.1	Haemonchus contortus, ISE/inbred ISE, WGS project CAVP01000000 data, chromosome: X	barber pole worm	34.2	34.2	77%	35	100.00%	46012677
CP032604.1	Lateolabrax maculatus linkage group 7 sequence	spotted sea bass	34.2	64.4	77%	35	100.00%	27713553
CP027281.1	Lateolabrax maculatus chromosome Lm20	spotted sea bass	34.2	64.4	77%	35	100.00%	21152183
XM_024658498.1	Cryptococcus neoformans var. neoformans JEC21 hypothetical protein (CND04685), mRNA	Cryptococcus neoformans var. neoformans JEC21	34.2	34.2	77%	35	100.00%	2087
XM_024168143.1	PREDICTED: Morus notabilis serine/threonine-protein kinase BLUS1 (LOC21410377), mRNA	Morus notabilis	34.2	34.2	77%	35	100.00%	2094
XR_002864067.1	PREDICTED: Cyanistes caeruleus uncharacterized LOC111928373 (LOC111928373), ncRNA	blue tit	34.2	34.2	77%	35	100.00%	1664
LT972181.1	Larimichthys crocea genome assembly, chromosome: XV	large yellow croaker	34.2	34.2	77%	35	100.00%	16065804
CP023113.2	Lupinus angustifolius cultivar Tanjil chromosome LG-01	narrow-leaved blue lupine	34.2	64.4	95%	35	95.24%	42828440
XM_022430592.1	PREDICTED: Crassostrea virginica phosphoenolpyruvate phosphomutase (LOC111099182), mRNA	eastern oyster	34.2	34.2	77%	35	100.00%	1753
CP021239.1	Kluyveromyces lactis strain GG799 chromosome A, complete sequence	Kluyveromyces lactis	34.2	68.4	90%	35	100.00%	1050866
CP018160.1	**Oryza sativa Indica Group cultivar Shuhui498 chromosome 4 sequence**	long-grained rice	34.2	34.2	77%	35	100.00%	35849732
XM_012406738.2	PREDICTED: Athalia rosae uncharacterized LOC105689598 (LOC105689598), mRNA	coleseed sawfly	34.2	34.2	77%	35	100.00%	1824
XM_018676841.1	PREDICTED: Lates calcarifer microtubule-associated serine/threonine-protein kinase 2-like (LOC108883545), mRNA	barramundi perch	34.2	34.2	77%	35	100.00%	10871
CP011299.1	Buchnera aphidicola (Schlechtendalia chinensis)	Buchnera aphidicola (Schlechtendalia	34.2	34.2	77%	35	100.00%	607835

	strain SC, complete genome	chinensis)							
KP125337.1	Colchicum autumnale chloroplast, complete genome	Colchicum autumnale	34.2	34.2	77%	35	100.00%	156462
AP015034.1	Vigna angularis var. angularis DNA, chromosome 1, almost complete sequence, cultivar: Shumari	Vigna angularis var. angularis	34.2	66.4	77%	35	100.00%	67115795
AP015039.1	Vigna angularis var. angularis DNA, chromosome 6, almost complete sequence, cultivar: Shumari	Vigna angularis var. angularis	34.2	34.2	77%	35	100.00%	38860970
KJ888251.1	Sardinella longiceps isolate CMFRIMBTDS 419 control region, partial sequence; mitochondrial	Indian oil sardine	34.2	34.2	77%	35	100.00%	1110
KJ888227.1	Sardinella longiceps isolate CMFRIMBTDS 325 control region, partial sequence; mitochondrial	Indian oil sardine	34.2	34.2	77%	35	100.00%	1076
AP014968.1	**Oryza sativa Japonica Group DNA, chromosome 12, cultivar: Nipponbare, complete sequence**	Japanese rice	34.2	64.4	100%	35	100.00%	27531856
AP014960.1	**Oryza sativa Japonica Group DNA, chromosome 4, cultivar: Nipponbare, complete sequence**	Japanese rice	34.2	34.2	77%	35	100.00%	35502694
AP014958.1	**Oryza sativa Japonica Group DNA, chromosome 2, cultivar: Nipponbare, complete sequence**	Japanese rice	34.2	34.2	77%	35	100.00%	35937250
CP012610.1	**Oryza sativa Indica Group cultivar RP Bio-226 chromosome 2 sequence**	long-grained rice	34.2	34.2	77%	35	100.00%	36385228
LK075972.1	Apteryx australis mantelli genome assembly AptMant0, scaffold scaffold11160	Apteryx australis mantelli	34.2	34.2	77%	35	100.00%	203429
LN713258.1	Cucumis melo genomic chromosome, chr_4	muskmelon	34.2	64.4	90%	35	100.00%	33123230
LN713255.1	Cucumis melo genomic chromosome, chr_1	muskmelon	34.2	34.2	77%	35	100.00%	35383099
LN681841.1	Cucumis melo genomic scaffold, anchoredscaffold00011	muskmelon	34.2	34.2	77%	35	100.00%	6909323
LN681802.1	Cucumis melo genomic scaffold, anchoredscaffold00046	muskmelon	34.2	34.2	77%	35	100.00%	3261579
LN590681.1	Cyprinus carpio genome assembly common carp genome, scaffold: LG47, chromosome: 47	common carp	34.2	64.4	77%	35	100.00%	16759857
HG975451.1 **1/2 2/2 3/2 3/3 6/1 8/1 9/1 10/1**	Solanum pennellii chromosome ch12, complete genome	Solanum pennellii	34.2	124	77%	35	100.00%	83305730
OA566334.1	**1_Tdi_b3v08**	Timema douglasi	34.2	34.2	77%	35	100.00%	125251
CP062021.1	Macrobrachium nipponense isolate FS-2020 chromosome 25	Macrobrachium nipponense	34.2	34.2	77%	35	100.00%	78920981
CP062051.1	Macrobrachium nipponense isolate FS-2020 chromosome 42	Macrobrachium nipponense	34.2	64.4	77%	35	100.00%	54564452
NC_049026.1	Uvularia grandiflora plastid, complete genome	Uvularia grandiflora	34.2	34.2	77%	35	100.00%	157025
NC_049025.1	Tripladenia cunninghamii plastid, complete genome	Tripladenia cunninghamii	34.2	34.2	77%	35	100.00%	155652
NC_049014.1	Androcymbium greuterocymbium plastid, complete genome	Androcymbium greuterocymbium	34.2	34.2	77%	35	100.00%	154804
LS423614.1	Anas platyrhynchos genome assembly, chromosome: 4	mallard	34.2	185	90%	35	100.00%	74522197
LR812533.1	Danio kyathit genome assembly, chromosome: 15	Danio kyathit	34.2	96.6	77%	35	100.00%	59247637
XM_771023.1	Cryptococcus neoformans var. neoformans B-3501A hypothetical protein (CNBD1640) partial mRNA	Cryptococcus neoformans var. neoformans B-3501A	34.2	34.2	77%	35	100.00%	2103
CU075459.1	Xenopus tropicalis finished cDNA, clone THdA011m01	tropical clawed frog	34.2	34.2	77%	35	100.00%	1620
AC102166.13	Mus musculus chromosome 1, clone RP23-447B15, complete sequence	house mouse	34.2	34.2	77%	35	100.00%	182043
AE017344.1	Cryptococcus neoformans var. neoformans JEC21 chromosome 4 sequence	Cryptococcus neoformans var. neoformans JEC21	34.2	34.2	77%	35	100.00%	1783081
AC125233.4	Mus musculus BAC clone RP23-9D24 from chromosome 5,	house mouse	34.2	34.2	77%	35	100.00%	212792

	complete sequence							
CR382121.1	Kluyveromyces lactis strain NRRL Y-1140 chromosome A complete sequence	Kluyveromyces lactis	34.2	68.4	90%	35	100.00%	1062590
AC158950.3	Mus musculus chromosome 1, clone RP24-215K13, complete sequence	house mouse	34.2	34.2	77%	35	100.00%	208060

ALIGNMENT VIEW FLAT QUERY-ANCHORED WITH LETTERS FOR IDENTITIES

Query range 1: 1 to 22

Query	1	TCTCCCATTACATGAGATTCAA	22
XM_031253612.1	5222	TCCCATTACATGAGATTCAA	5203
CP025662.1	1281008	TCCCATTACATGAGATTCAA	1280989
CP025646.1	1139018	TCCCATTACATGAGATTCAA	1138999
XM_019329813.1	5281	TCCCATTACATGAGATTCAA	5262
CP032554.1	6737245	TCCCATTACATGAGATTCA	6737227
CP025781.1	1900961	TCTCCCATTACATGAGATT	1900943
LT669788.1	22674659	TCCCATTACATGAGATTCA	22674677
AL356461.15	29562	TCTCCCATTACATGAGATT	29544
XM_030979643.1	5556	CCATTACATGAGATTCAA	5573
XM_011679278.2	5566	CCATTACATGAGATTCAA	5583
XM_011679285.2	5565	CCATTACATGAGATTCAA	5582
CP035802.1	45968877	TCCCATTACATGAGATTC	45968894
CP035802.1	45969896	TCCCATTACATGAGATTC	45969913
CP035802.1	41738010	TTACATGAGATTCAA	41737996
LR606180.1	43712628	TCTCCCATTACATGAGAT	43712645
LR606180.1	47654345	ATTACATGAGATTCAA	47654360
LR606180.1	3332135	TCCCATTACATGAGA	3332121
LR606180.1	7667991	TCCCATTACATGAGA	7667977
LR606180.1	32312744	TCTCCCATTACATGA	32312758
LR606180.1	56948018	ATTACATGAGATTCA	56948032
LR606180.1	85170661	CCATTACATGAGATT	85170675
LR597476.1	5944255	CTCCCATTACATGAGATT	5944272
LR597476.1	35246834	TTACATGAGATTCAA	35246848
LR597476.1	49557654	CATTACATGAGATTC	49557640
CP035095.1	56545	CCATTACATGAGATTCAA	56562
CP032250.1	30345129	CCATTACATGAGATTCAA	30345146
CP032250.1	5216992	TTACATGAGATTCAA	5217006
CP032250.1	25125222	TTACATGAGATTCAA	25125236
CP032245.1	38678537	CCATTACATGAGATTCAA	38678554
CP032245.1	851779	CATTACATGAGATTCA	851794
CP032245.1	4199046	ATTACATGAGATTCA	4199060
CP032245.1	34114295	CCATTACATGAGATT	34114309
CP026261.1	8768179	CCATTACATGAGATTCAA	8768162
CP009921.1	119160	CCATTACATGAGATTCAA	119143
MT261175.1	82718	TCCCATTACATGAGATT	82702
MT261174.1	81902	TCCCATTACATGAGATT	81886
MT261148.1	80875	TCCCATTACATGAGATT	80859
CP050574.1	52511578	CATTACATGAGATTCAA	52511594
CP050574.1	45396145	CATTACATGAGATTCA	45396130
CP050570.1	15520519	TCTCCCATTACATGAGA	15520503
CP050613.1	52531299	CATTACATGAGATTCAA	52531315

CP050613.1	45418493	CATTACATGAGATTCA	45418478
CP050610.1	15510022	TCTCCCATTACATGAGA	15510006
AP022632.1	81609444	CTCCCATTACATGAGAT	81609460
AP022632.1	1185563	CTCCCATTACATGAG	1185549
AP022704.1	31638031	CCATTACATGAGATTCA	31638047
AP022704.1	13523044	ATTACATGAGATTCA	13523058
XM_002936994.4	1647	TCTCCCATTACATGAGA	1631
LR722981.1	16834644	CATTACATGAGATTCAA	16834660
LR722981.1	89069024	CATTACATGAGATTCA	89069009
LR722981.1	89112696	CATTACATGAGATTCA	89112681
CP042455.1	1241841	CCATTACATGAGATTCA	1241825
XM_030198566.1	996	TCTCCCATTACATGAGA	980
XM_030198557.1	996	TCTCCCATTACATGAGA	980
XM_030198551.1	1122	TCTCCCATTACATGAGA	1106
XM_030198543.1	1122	TCTCCCATTACATGAGA	1106
XM_030198536.1	1122	TCTCCCATTACATGAGA	1106
XM_030198529.1	1122	TCTCCCATTACATGAGA	1106
XM_030198521.1	1122	TCTCCCATTACATGAGA	1106
LR537130.1	21695109	CATTACATGAGATTCAA	21695093
LR537130.1	29966081	CATTACATGAGATTCA	29966066
CP035801.1	33249112	TCTCCCATTATATGAGATTCA	33249132
CP035801.1	6199412	TCCCATTACATGAGA	6199398
CP035800.1	186911	CATTACATGAGATTCAA	186927
LR606198.1	25126242	TCTCCCATTACATGAGA	25126258
LR606198.1	27821319	ATTACATGAGATTCAA	27821304
LR597552.1	23845435	TCTCCCAGTACATGAGATTCA	23845455
LR597552.1	16379503	TTACATGAGATTCAA	16379489
XM_029315943.1	1108	CATTACATGAGATTCAA	1124
XM_029315942.1	1107	CATTACATGAGATTCAA	1123
CP032252.1	6434823	CCATTACATGAGATTCA	6434807
CP032252.1	44215530	CCATTACATGAGATTCA	44215514
CP032252.1	26950127	CATTACATGAGATTCA	26950142
CP032252.1	33839702	CCATTACATGAGATTC	33839717
CP032252.1	34237892	CATTACATGAGATTC	34237906
CP032252.1	37887277	CATTACATGAGATTC	37887263
CP032252.1	44549642	TTACATGAGATTCAA	44549656
CP032251.1	35471183	CCATTACATGAGATTCA	35471199
CP032251.1	37264375	CATTACATGAGATTC	37264361
CP032244.1	18935716	TCCCATTACATAAGATTCA	18935734
CP032244.1	19967470	CCATTACATGAGATTCA	19967486
CP032244.1	4230653	CATTACATGAGATTC	4230639
LR535841.1	1112990	CATTACATGAGATTCAA	1112974
LR535841.1	25095307	TCTCCCATTACATGA	25095293
LR132043.1	24477558	CCATTACATGAGATTCA	24477542
LR132043.1	18950349	TCCCATTACATGAGAT	18950364
MK072386.1	18967	TCTCCCAATACATGAGATTCA	18947
XM_026573109.1	1343	TCTCCCATTACATGAGA	1327
CP032418.1	1289823	TCTCCCATTACATGAGA	1289839
CP025672.1	2204754	CATTACATGAGATTCAA	2204770
CP025672.1	18960837	TTACATGAGATTCAA	18960851
CP025656.1	1330796	CATTACATGAGATTCAA	1330812
LS997567.1	259351	CATTACATGAGATTCAA	259367
CP032604.1	21005603	CATTACATGAGATTCAA	21005619
CP032604.1	21347900	TTACATGAGATTCAA	21347886
CP027281.1	8637131	CATTACATGAGATTCAA	8637115
CP027281.1	8191845	ATTACATGAGATTCA	8191831
XM_024658498.1	633	CTCCCATTACATGAGAT	617
XM_024168143.1	1717	CATTACATGAGATTCAA	1701
XR_002864067.1	1518	CATTACATGAGATTCAA	1534
LT972181.1	7109283	TCCCATTACATGAGATT	7109299

CP023113.2	20058368	CTCCCATTACATGAAATTCAA	20058348
CP023113.2	19208163	TTACATGAGATTCAA	19208177
XM_022430592.1	1510	CATTACATGAGATTCAA	1494
CP021239.1	700004	CCATTACATGAGATTCA	699988
CP021239.1	965864	CTCCCATTACATGAGAT	965880
CP018160.1	2265361	TCTCCCATTACATGAGA	2265377
XM_012406738.2	1057	TCTCCCATTACATGAGA	1041
XM_018676841.1	10516	CCATTACATGAGATTCA	10532
CP011299.1	555657	CATTACATGAGATTCAA	555641
KP125337.1	81891	TCCCATTACATGAGATT	81875
AP015034.1	54524717	CATTACATGAGATTCAA	54524733
AP015034.1	66723276	CATTACATGAGATTCA	66723291
AP015039.1	37843921	TCCCATTACATGAGATT	37843905
KJ888251.1	748	CCATTACATGAGATTCA	732
KJ888227.1	742	CCATTACATGAGATTCA	726
AP014968.1	21053515	TCTCCCATTACATGAGA	21053531
AP014968.1	8709153	TTACATGAGATTCAA	8709139
AP014960.1	2251499	TCTCCCATTACATGAGA	2251483
AP014958.1	9627530	TCTCCCATTACATGAGA	9627546
CP012610.1	9941563	TCTCCCATTACATGAGA	9941579
LK075972.1	10878	TCTCCCATTACATGAGA	10862
LN713258.1	29807408	TCCCATTACATGAGATT	29807424
LN713258.1	1409995	TTACATGAGATTCAA	1410009
LN713255.1	26026003	CATTACATGAGATTCAA	26025987
LN681841.1	3315823	TCCCATTACATGAGATT	3315807
LN681802.1	2219811	CATTACATGAGATTCAA	2219795
LN590681.1	3767544	CATTACATGAGATTCAA	3767528
LN590681.1	16263666	TTACATGAGATTCAA	16263652
HG975451.1	27599434	CATTACATGAGATTCAA	27599450
HG975451.1	63830996	TTACATGAGATTCAA	63830982
HG975451.1	66010204	CATTACATGAGATTC	66010218
HG975451.1	68655790	CATTACATGAGATTC	68655804
OA566334.1	71577	TCCCATTACATGAGATT	71593
CP062021.1	15212325	CTCCCATTACATGAGAT	15212309
CP062051.1	9417621	CATTACATGAGATTCAA	9417605
CP062051.1	40451573	TTACATGAGATTCAA	40451559
NC_049026.1	82718	TCCCATTACATGAGATT	82702
NC_049025.1	81902	TCCCATTACATGAGATT	81886
NC_049014.1	80875	TCCCATTACATGAGATT	80859
LS423614.1	65154382	TCCCATTACATGAGATT	65154398
LS423614.1	10082855	TCCCATTACATGAGA	10082841
LS423614.1	12504604	CATTACATGAGATTC	12504590
LS423614.1	41585139	TTACATGAGATTCAA	41585153
LS423614.1	45137039	TTACATGAGATTCAA	45137053
LS423614.1	58730326	TTACATGAGATTCAA	58730312
LR812533.1	34706224	CATTACATGAGATTCAA	34706208
LR812533.1	45428608	ATTACATGAGATTCAA	45428623
LR812533.1	17485149	ATTACATGAGATTCA	17485135
XM_771023.1	460	CTCCCATTACATGAGAT	444
CU075459.1	1524	TCTCCCATTACATGAGA	1508
AC102166.13	98747	CATTACATGAGATTCAA	98763
AE017344.1	1290155	CTCCCATTACATGAGAT	1290171
AC125233.4	153179	CCATTACATGAGATTCA	153195
CR382121.1	700120	CCATTACATGAGATTCA	700104
CR382121.1	965985	CTCCCATTACATGAGAT	966001
AC158950.3	54888	CATTACATGAGATTCAA	54872

PRODUCTS OF SIGNIFICANT ALIGNMENTS (CDS FEATURE)

- **Auxin transport protein BIG** GenBank; GenBank
- **Hypothetical protein** GenBank
- **Sgo1** GenBank
- **RNA-binding protein 33** GenBank; GenBank; GenBank; GenBank; GenBank; GenBank; GenBank
- **Cytochrome c1-2, heme protein,mitochondrial-like** GenBank; GenBank
- **Transcription antitermination factor NusB** GenBank
- **Hypothetical protein** GenBank; GenBank; GenBank; GenBank
- **Serine/threonine-protein kinase BLUS1** GenBank
- **KLLA0A07865p** GenBank

SEQUENCE N° 3/5 (T=2; C=0; G=3; A=1)

Query: **TCTCCCATTACATGAGATTCAA**

Search Parameters	
Program	blastn
Word size	7
Expect value	1000
Hitlist size	250
Match/Mismatch scores	1,-3
Gapcosts	5,2
Filter string	F
Genetic Code	1

250 sequences selected:

Accession Seq	Description	Comomm Name	Max Score	Total Score	Query C.	E value	Per. Id.	Acc. Len
LR862377.1 1/2 2/2 3/2 3/3 6/1 7/2 8/1 9/1 10/1	***Pipistrellus pipistrellus genome assembly, chromosome: 21***	common pipistrelle	32.2	32.2	72%	138	100.00%	7781141
LR862372.1 1/2 2/2 3/2 3/3 6/1 7/2 8/1 9/1 10/1	***Pipistrellus pipistrellus genome assembly, chromosome: 16***	common pipistrelle	32.2	32.2	72%	138	100.00%	46231460
LR862368.1 1/2 2/2 3/2 3/3 6/1 7/2 8/1 9/1 10/1	***Pipistrellus pipistrellus genome assembly, chromosome: 12***	common pipistrelle	32.2	122	90%	138	100.00%	58291996
LR862367.1 1/2 2/2 3/2 3/3 6/1 7/2 8/1 9/1 10/1	***Pipistrellus pipistrellus genome assembly, chromosome: 11***	common pipistrelle	32.2	92.7	100%	138	100.00%	69386144
LR862363.1 1/2 2/2 3/2 3/3 6/1 7/2 8/1 9/1 10/1	***Pipistrellus pipistrellus genome assembly, chromosome: 7***	common pipistrelle	32.2	32.2	72%	138	100.00%	84548104

SEQUENCE N° 3/6

(T=2; C=0; G=3; A=1)

SEQUENCE N° 3/6 (T=2; C=0; G=3; A=1)

Original Sequence	G	T	G	A	A	A	T	G	G	T	C	A	T	G	T	G	T	G	G	C	G	G
Trend n° 3 (T=2; C=0; G=3; A=1)	3	2	3	1	1	1	2	3	3	2	0	1	2	3	2	3	2	3	3	0	3	3
Trend n° 3 (I i D d)	-	D	I	D	d	d	I	I	i	D	D	I	I	I	D	I	D	I	i	D	I	i
Trend n° 3/6	3	2	3	2	2	2	1	2	2	1	0	1	2	3	0	2	0	3	3	0	1	1
Sequence n° 3/6 (54,54%)	G	T	G	T	T	T	A	T	T	A	C	A	T	G	C	T	C	G	G	C	A	A

Query: **GTGTTTATTACATGCTCGGCAA**

Search Parameters	
Program	blastn
Word size	7
Expect value	1000
Hitlist size	100
Match/Mismatch scores	1,-3
Gapcosts	5,2
Filter string	F
Genetic Code	1

100 sequences selected:

Accession Seq	Description	Comomm Name	Max Score	Total Score	Query C.	E value	Per. Id.	Acc. Len
LR584428.1 1/2 2/2 3/3 3/4 8/1	Salmo trutta genome assembly, chromosome: 13	river trout	40.1	40.1	90%	0.57	100.00%	91488822
CP017747.1	Bacillus velezensis strain SYBC H47, complete genome	Bacillus velezensis	38.2	38.2	86%	2.2	100.00%	3884433
CP014700.1	Bacillus amyloliquefaciens strain S499, complete genome	Bacillus amyloliquefaciens	38.2	38.2	86%	2.2	100.00%	3927922
CP004065.1	Bacillus amyloliquefaciens IT-45, complete genome	Bacillus amyloliquefaciens IT-45	38.2	38.2	86%	2.2	100.00%	3928857
CP035533.1	Bacillus velezensis strain DTU001 chromosome, complete genome	Bacillus velezensis	38.2	38.2	86%	2.2	100.00%	3927210
CP063768.1	Bacillus velezensis strain KMU01 chromosome, complete genome	Bacillus velezensis	38.2	38.2	86%	2.2	100.00%	3932437
CP053764.1	Bacillus velezensis strain B268 chromosome, complete genome	Bacillus velezensis	38.2	38.2	86%	2.2	100.00%	3908940
CP042271.1	Bacillus velezensis strain LPL061 chromosome, complete genome	Bacillus velezensis	38.2	38.2	86%	2.2	100.00%	3907268
CP014990.2	Bacillus velezensis strain KD1 chromosome, complete genome	Bacillus velezensis	38.2	38.2	86%	2.2	100.00%	3921966
CP021011.1	Bacillus velezensis strain GFP-2 chromosome, complete genome	Bacillus velezensis	38.2	38.2	86%	2.2	100.00%	3975220
CP019040.1	Bacillus velezensis strain GH1-13, complete genome	Bacillus velezensis	38.2	38.2	86%	2.2	100.00%	4071980
CP017953.1	Bacillus amyloliquefaciens strain Y14, complete genome	Bacillus amyloliquefaciens	38.2	38.2	86%	2.2	100.00%	3957164
LK064797.1	Apteryx australis mantelli genome assembly AptMant0, scaffold scaffold152	Apteryx australis mantelli	38.2	38.2	86%	2.2	100.00%	6182939
CP011278.1	Bacillus velezensis strain L-S60 chromosome, complete genome	Bacillus velezensis	38.2	38.2	86%	2.2	100.00%	3903017
CP010556.1	Bacillus velezensis strain L-H15 chromosome, complete genome	Bacillus velezensis	38.2	38.2	86%	2.2	100.00%	3905973
CP061087.1	Bacillus velezensis strain CLA178 chromosome, complete genome	Bacillus velezensis	38.2	38.2	86%	2.2	100.00%	4068479
LR812074.1	Danio rerio genome assembly, chromosome: 12	zebrafish	36.2	66.4	81%	8.9	100.00%	36192843
CP030886.1	Cytophagaceae bacterium isolate ECT2AJA-110-C chromosome	Cytophagaceae bacterium	36.2	36.2	81%	8.9	100.00%	4167162
LR812049.1	Danio rerio genome assembly, chromosome: 12	zebrafish	36.2	36.2	81%	8.9	100.00%	49740644
LR812580.1	Danio rerio strain Nadia (NA) genome assembly, chromosome: 12	zebrafish	36.2	66.4	81%	8.9	100.00%	46630904
AC277094.1	Coccomyxa subellipsoidea C-169 clone FAHX53-E22, complete sequence	Coccomyxa subellipsoidea C-169	36.2	36.2	81%	8.9	100.00%	39353

XM_012768883.2	PREDICTED: Microcebus murinus B-box and SPRY domain containing (BSPRY), transcript variant X2, mRNA	gray mouse lemur	36.2	36.2	81%	8.9	100.00%	2116
XM_012768882.2	PREDICTED: Microcebus murinus B-box and SPRY domain containing (BSPRY), transcript variant X1, mRNA	gray mouse lemur	36.2	36.2	81%	8.9	100.00%	2119
XM_009865080.1	PREDICTED: Apaloderma vittatum angiopoietin 2 (ANGPT2), partial mRNA	bar-tailed trogon	36.2	36.2	100%	8.9	95.45%	1601
LR812605.1	Danio rerio genome assembly, chromosome: 12	zebrafish	36.2	36.2	81%	8.9	100.00%	46867306
LR812078.1	Danio rerio genome assembly, chromosome: 16	zebrafish	34.2	94.6	77%	35	100.00%	55350661
CP051157.1	Wolbachia endosymbiont of Ctenocephalides felis wCfeJ chromosome, complete genome	Wolbachia endosymbiont of Ctenocephalides felis wCfeJ	34.2	34.2	77%	35	100.00%	1201647
AP022630.1	Lotus japonicus B-129 DNA, chromosome 2, complete sequence	Lotus japonicus	34.2	34.2	77%	35	100.00%	95628465
LR778318.1	Raphanus sativus genome assembly, chromosome: 9	radish	34.2	34.2	77%	35	100.00%	28894300
LR778315.1	Raphanus sativus genome assembly, chromosome: 6	radish	34.2	34.2	77%	35	100.00%	35522300
LR778310.1	Raphanus sativus genome assembly, chromosome: 1	radish	34.2	64.4	81%	35	100.00%	58898396
LR738417.1	Lutra lutra genome assembly, chromosome: 15	Eurasian river otter	34.2	126	81%	35	100.00%	69992071
LR738605.1	Sciurus carolinensis genome assembly, chromosome: 15	gray squirrel	34.2	64.4	77%	35	100.00%	83138899
CP045898.1	Caligus rogercresseyi isolate FCH chromosome 9	Caligus rogercresseyi	34.2	102	95%	35	95.24%	25126535
LR597472.1 1/2 2/2 3/1 3/5	Sphaeramia orbicularis genome assembly, chromosome: 15	orbiculate cardinalfish	34.2	34.2	77%	35	100.00%	54816899
LR597464.1 1/2 2/2 3/1 3/5	Sphaeramia orbicularis genome assembly, chromosome: 7	orbiculate cardinalfish	34.2	283	77%	35	100.00%	57165184
CP027523.1	Pseudoalteromonas carrageenovora strain KCTC 22325 chromosome I, complete sequence	Pseudoalteromonas carrageenovora	34.2	34.2	95%	35	95.24%	3633318
LR131964.1	Parambassis ranga genome assembly, chromosome: 14	Indian glassy fish	34.2	34.2	77%	35	100.00%	21894654
XM_026419049.1	PREDICTED: Frankliniella occidentalis CDGSH iron-sulfur domain-containing protein 2 homolog (LOC113204050), mRNA	western flower thrips	34.2	34.2	77%	35	100.00%	989
CP022985.1	Mariniflexile sp. TRM1-10 chromosome, complete genome	Mariniflexile sp. TRM1-10	34.2	68.4	95%	35	95.24%	4858325
CP035704.1	Pseudolysobacter antarcticus strain AQ6-296 chromosome, complete genome	Xanthomonadaceae bacterium AQ6-296	34.2	34.2	77%	35	100.00%	4711779
LT965928.1	Pseudoalteromonas carrageenovora IAM 12662 strain ATCC43555T genome assembly, chromosome: PCAR9a	Pseudoalteromonas carrageenovora IAM 12662	34.2	34.2	95%	35	95.24%	3620649
CP022522.1	Pseudoalteromonas sp. NC201 chromosome, complete genome	Pseudoalteromonas sp. NC201	34.2	34.2	77%	35	100.00%	4134317
XM_022836198.1	PREDICTED: Varroa jacobsoni adenylyl cyclase-associated protein 1-like (LOC111262154), transcript variant X2, mRNA	Varroa jacobsoni	34.2	34.2	95%	35	95.24%	2800
XM_022836197.1	PREDICTED: Varroa jacobsoni adenylyl cyclase-associated protein 1-like (LOC111262154), transcript variant X1, mRNA	Varroa jacobsoni	34.2	34.2	95%	35	95.24%	2960
XM_022792557.1	PREDICTED: Varroa destructor adenylyl cyclase-associated protein 1-like (LOC111244941), transcript variant X2, mRNA	honeybee mite	34.2	34.2	95%	35	95.24%	2445
XM_022792556.1	PREDICTED: Varroa destructor adenylyl cyclase-associated protein 1-like (LOC111244941), transcript variant X1, mRNA	honeybee mite	34.2	34.2	95%	35	95.24%	2644
XM_015846147.1	Paracoccidioides lutzii Pb01 hypothetical protein (PAAG_07160), partial mRNA	Paracoccidioides lutzii Pb01	34.2	34.2	77%	35	100.00%	5970
CP011895.1	Ovis canadensis canadensis isolate 43U chromosome 10 sequence	Ovis canadensis canadensis	34.2	126	95%	35	95.24%	86551329
XM_010757741.1	Paracoccidioides brasiliensis Pb18 hypothetical protein partial mRNA	Paracoccidioides brasiliensis Pb18	34.2	34.2	77%	35	100.00%	5970
LK995546.1	Rhabditophanes sp. KR3021 genome assembly	Rhabditophanes sp. KR3021	34.2	34.2	77%	35	100.00%	7271431

	Rhabditophanes_sp_KR3021, scaffold RSKR_scaffold0000001							
XM_008876883.1	Aphanomyces invadans hypothetical protein, variant mRNA	Aphanomyces invadans	34.2	34.2	77%	35	100.00%	2034
XM_008876882.1	Aphanomyces invadans hypothetical protein mRNA	Aphanomyces invadans	34.2	34.2	77%	35	100.00%	2099
XM_007408047.1	Melampsora larici-populina 98AG31 uncharacterized protein (MELLADRAFT_84704), partial mRNA	Melampsora larici-populina 98AG31	34.2	34.2	77%	35	100.00%	2841
FO081946.3	Pig DNA sequence from clone WTSI_1061-44B3 on chromosome Y, complete sequence	pig	34.2	34.2	77%	35	100.00%	38230
LR812053.1	Danio rerio genome assembly, chromosome: 16	zebrafish	34.2	124	81%	35	100.00%	55934183
LR812584.1	Danio rerio strain Nadia (NA) genome assembly, chromosome: 16	zebrafish	34.2	94.6	77%	35	100.00%	56178339
LR812520.1	Danio kyathit genome assembly, chromosome: 2	Danio kyathit	34.2	124	95%	35	100.00%	73064863
CP000612.1	Desulfotomaculum reducens MI-1, complete genome	Desulfotomaculum reducens MI-1	34.2	34.2	77%	35	100.00%	3608104
CR854942.16	Zebrafish DNA sequence from clone DKEY-246I14 in linkage group 16, complete sequence	zebrafish	34.2	34.2	77%	35	100.00%	240480
CP050595.1 **1/2 2/2 3/3 3/4 3/5 5/1 10/1**	Canis lupus familiaris breed Labrador retriever chromosome 05a	dog	34.2	64.4	77%	35	100.00%	86926571
CP050642.1 **1/2 2/2 3/3 3/4 3/5 5/1 10/1**	Canis lupus familiaris breed Labrador retriever chromosome 05b	dog	34.2	64.4	77%	35	100.00%	86915062
XM_030328380.2	PREDICTED: Lynx canadensis intersectin 1 (ITSN1), transcript variant X5, mRNA	Canada lynx	34.2	34.2	77%	35	100.00%	5276
XM_030328379.2	PREDICTED: Lynx canadensis intersectin 1 (ITSN1), transcript variant X4, mRNA	Canada lynx	34.2	34.2	77%	35	100.00%	5291
XM_030328378.2	PREDICTED: Lynx canadensis intersectin 1 (ITSN1), transcript variant X3, mRNA	Canada lynx	34.2	34.2	77%	35	100.00%	5489
XM_030328376.2	PREDICTED: Lynx canadensis intersectin 1 (ITSN1), transcript variant X1, mRNA	Canada lynx	34.2	34.2	77%	35	100.00%	5504
CP040558.1	Pseudoalteromonas sp. 16-SW-7 chromosome L1, complete sequence	Pseudoalteromonas sp. 16-SW-7	34.2	34.2	95%	35	95.24%	3735685
XM_015263782.2	PREDICTED: Diachasma alloeum filamin-C (LOC107042647), mRNA	Diachasma alloeum	34.2	34.2	77%	35	100.00%	9274
LR132000.1 **1/2 3/3 3/4**	Gouania willdenowi genome assembly, chromosome: 6	blunt-snouted clingfish	34.2	128	86%	35	100.00%	72451976
XM_027041670.1	PREDICTED: Acinonyx jubatus intersectin 1 (ITSN1), transcript variant X6, mRNA	cheetah	34.2	34.2	77%	35	100.00%	5435
CP033065.1	Pseudoalteromonas agarivorans strain Hao 2018 chromosome I, complete sequence	Pseudoalteromonas agarivorans	34.2	34.2	95%	35	95.24%	3611742
XM_026134041.1	PREDICTED: Solenopsis invicta WD repeat-containing protein on Y chromosome (LOC105207815), mRNA	red fire ant	34.2	34.2	77%	35	100.00%	2976
CP030990.1	Arachis hypogaea cultivar Shitouqi chromosome A08	peanut	34.2	64.4	81%	35	100.00%	51955169
CP027087.1 **1/2 7/2 10/1**	Bos mutus isolate yakQH1 chromosome 19	wild yak	34.2	34.2	77%	35	100.00%	60756558
XM_023238782.1	PREDICTED: Felis catus intersectin 1 (ITSN1), transcript variant X6, mRNA	domestic cat	34.2	34.2	77%	35	100.00%	5421
CP023464.1	Pseudoalteromonas atlantica strain ECSMB14104 chromosome, complete genome	Pseudoalteromonas atlantica	34.2	34.2	95%	35	95.24%	3669595
CP011011.1	Pseudoalteromonas agarivorans DSM 14585 chromosome I, complete sequence	Pseudoalteromonas agarivorans DSM 14585	34.2	34.2	95%	35	95.24%	3701940
CP011028.1	Pseudoalteromonas espejiana strain ATCC 29659 chromosome I, complete sequence	Pseudoalteromonas espejiana DSM 9414	34.2	34.2	95%	35	95.24%	3720756
CP015399.2	Lachnoclostridium sp. YL32, complete genome	Lachnoclostridium sp. YL32	34.2	34.2	77%	35	100.00%	7225339
LT634572.1	Sus scrofa genome assembly, chromosome: Y	pig	34.2	34.2	77%	35	100.00%	43547828

XM_015541209.1	PREDICTED: Panthera tigris altaica intersectin 1 (ITSN1), transcript variant X1, mRNA	Amur tiger	34.2	34.2	77%	35	100.00%	5200
XM_012986555.1	PREDICTED: Erythranthe guttatus tricyclene synthase 0e23, chloroplastic-like (LOC105962255), transcript variant X2, mRNA	spotted monkey flower	34.2	34.2	77%	35	100.00%	1907
XM_012986554.1	PREDICTED: Erythranthe guttatus tricyclene synthase 0e23, chloroplastic-like (LOC105962255), transcript variant X1, mRNA	spotted monkey flower	34.2	34.2	77%	35	100.00%	2025
XM_012699782.1	PREDICTED: Hydra vulgaris alpha-catulin-like (LOC101236340), mRNA	Hydra vulgaris	34.2	34.2	77%	35	100.00%	468
LR812609.1	Danio rerio genome assembly, chromosome: 16	zebrafish	34.2	124	81%	35	100.00%	53959708
FP074886.7	Pig DNA sequence from clone CH242-283J17 on chromosome X, complete sequence	pig	34.2	34.2	77%	35	100.00%	157466
FP929039.1	Coprococcus sp. ART55/1 draft genome	Coprococcus sp. ART55/1	34.2	34.2	77%	35	100.00%	3123007
AC149883.2	Xenopus tropicalis clone CH216-80L11, complete sequence	tropical clawed frog	34.2	34.2	77%	35	100.00%	146021
AY016023.1	Sphoeroides nephelus alpha globin gene cluster, complete sequence	southern puffer	34.2	34.2	77%	35	100.00%	167215
AB006531.1	Plautia stali intestine virus RNA for nonstructural polyprotein, capsid protein precursor, complete cds	Plautia stali intestine virus	34.2	34.2	77%	35	100.00%	8797
OA566189.1	**1_Tdi_b3v08**	Timema douglasi	32.2	32.2	90%	138	95.00%	130219
OA564290.1	**1_Tdi_b3v08**	Timema douglasi	32.2	32.2	72%	138	100.00%	624938
CP064845.1	Bacillus amyloliquefaciens strain BZR 277 chromosome, complete genome	Bacillus amyloliquefaciens	32.2	32.2	72%	138	100.00%	3913640
CP064846.1	Bacillus siamensis strain BZR 86 chromosome, complete genome	Bacillus siamensis	32.2	32.2	72%	138	100.00%	3913613
OA565377.1	**1_Tdi_b3v08**	Timema douglasi	32.2	32.2	72%	138	100.00%	180607
CP064904.1	Chrysoporthe sp. MAF-2017a isolate CMW54409 chromosome 5	Chrysoporthe sp. MAF-2017a	32.2	32.2	72%	138	100.00%	4781437
CP064851.1	Klebsiella pneumoniae strain 130002 chromosome, complete genome	Klebsiella pneumoniae	32.2	32.2	72%	138	100.00%	5297461
CP043357.1	Klebsiella pneumoniae subsp. pneumoniae strain WCHKP020120 chromosome, complete genome	Klebsiella pneumoniae subsp. pneumoniae	32.2	32.2	72%	138	100.00%	5462067
CP043353.1	Klebsiella pneumoniae strain WCHKP020115 chromosome, complete genome	Klebsiella pneumoniae	32.2	32.2	72%	138	100.00%	5445745
CP064820.1	Klebsiella pneumoniae subsp. pneumoniae strain K219 chromosome, complete genome	Klebsiella pneumoniae subsp. pneumoniae	32.2	32.2	72%	138	100.00%	5038803

ALIGNMENT VIEW FLAT QUERY-ANCHORED WITH LETTERS FOR IDENTITIES

Query range 1: 1 to 22

Query	1	GTGTTTATTACATGCTCGGCAA	22
LR584428.1	14898421	GTGTTTATTACATGCTCGGC	14898402
CP017747.1	798002	GTTTATTACATGCTCGGCA	798020
CP014700.1	832522	GTTTATTACATGCTCGGCA	832540
CP004065.1	1541556	GTTTATTACATGCTCGGCA	1541574
CP035533.1	2371031	GTTTATTACATGCTCGGCA	2371049
CP063768.1	2442032	GTTTATTACATGCTCGGCA	2442050
CP053764.1	2390680	GTTTATTACATGCTCGGCA	2390662
CP042271.1	1370468	GTTTATTACATGCTCGGCA	1370450
CP014990.2	2393605	GTTTATTACATGCTCGGCA	2393587
CP021011.1	365467	GTTTATTACATGCTCGGCA	365449
CP019040.1	2532062	GTTTATTACATGCTCGGCA	2532044
CP017953.1	2442226	GTTTATTACATGCTCGGCA	2442208
LK064797.1	242219	TTTATTACATGCTCGGCAA	242201
CP011278.1	2397601	GTTTATTACATGCTCGGCA	2397583
CP010556.1	2400776	GTTTATTACATGCTCGGCA	2400758
CP061087.1	2488573	GTTTATTACATGCTCGGCA	2488555
LR812074.1	30132754	GTGTTTATTACATGCTCG	30132771
LR812074.1	6704537	TGTTTATTACATGCT	6704551
CP030886.1	1283744	GTGTTTATTACATGCTCG	1283761
LR812049.1	43764999	GTGTTTATTACATGCTCG	43765016
LR812580.1	43528452	GTGTTTATTACATGCTCG	43528469
LR812580.1	6408889	TGTTTATTACATGCT	6408903
AC277094.1	19898	TTATTACATGCTCGGCAA	19881
XM_012768883.2	1574	GTGTTTATTACATGCTCG	1557
XM_012768882.2	1577	GTGTTTATTACATGCTCG	1560
XM_009865080.1	1114	GTGTTTATTATATGCTCGGCAA	1093
LR812605.1	41867345	GTGTTTATTACATGCTCG	41867328
LR812078.1	22763827	TGTTTATTACATGCTCG	22763843
LR812078.1	13684367	TGTTTATTACATGCT	13684381
LR812078.1	13333107	TGTTTATTACATGCT	13333093
CP051157.1	593713	TTTATTACATGCTCGGC	593729
AP022630.1	45479831	TTATTACATGCTCGGCA	45479847
LR778318.1	4233446	GTTTATTACATGCTCGG	4233462
LR778315.1	29657508	GTTTATTACATGCTCGG	29657524
LR778310.1	19339571	GTTTATTACATGCTCGG	19339587
LR778310.1	17970188	TATTACATGCTCGGC	17970174
LR738417.1	30385450	TGTTTATTACATGCTCG	30385466
LR738417.1	27964933	GTGTTTATTACATGCT	27964918
LR738417.1	3163553	TGTTTATTACATGCT	3163567
LR738417.1	21615284	GTGTTTATTACATGC	21615270
LR738605.1	79522714	GTGTTTATTACATGCTC	79522730
LR738605.1	47118396	GTTTATTACATGCTC	47118410
CP045898.1	4157504	GTGTTTATGACATGCTCGGCA	4157524
CP045898.1	4164309	GTGTTTATGACATGCTCGGCA	4164329
CP045898.1	3836317	GTGTTTATGACATGCTCGGCA	3836297
LR597472.1	18939790	TTATTACATGCTCGGCA	18939806
LR597464.1	43973606	GTGTTTATTACATGCTC	43973622
LR597464.1	3464358	TGTTTATTACATGCTC	3464343
LR597464.1	3465908	TGTTTATTACATGCTC	3465893
LR597464.1	3468491	TGTTTATTACATGCTC	3468476
LR597464.1	3477966	TGTTTATTACATGCTC	3477951
LR597464.1	3410873	TGTTTATTACATGCT	3410887
LR597464.1	56438488	TGTTTATTACATGCT	56438502

LR597464.1	3473140	GTTTATTACATGCTC	3473126
LR597464.1	9261435	GTGTTTATTACATGC	9261421
CP027523.1	773271	GTGTTTAATACATGCTCGGCA	773291
LR131964.1	2784092	GTGTTTATTACATGCTC	2784108
XM_026419049.1	749	TGTTTATTACATGCTCG	765
CP022985.1	263866	TGTTTATTACATTCTCGGCAA	263886
CP022985.1	1767496	TGTTTATTACATTCTCGGCAA	1767476
CP035704.1	2627575	TTTATTACATGCTCGGC	2627591
LT965928.1	767567	GTGTTTAATACATGCTCGGCA	767587
CP022522.1	1439540	GTTTATTACATGCTCGG	1439556
XM_022836198.1	1412	TGTTGATTACATGCTCGGCAA	1432
XM_022836197.1	1572	TGTTGATTACATGCTCGGCAA	1592
XM_022792557.1	1612	TGTTGATTACATGCTCGGCAA	1632
XM_022792556.1	1811	TGTTGATTACATGCTCGGCAA	1831
XM_015846147.1	4378	TATTACATGCTCGGCAA	4394
CP011895.1	47859170	GTGTTTATTACATGCTAGGCA	47859190
CP011895.1	32793092	GTGTTTATTACATGCT	32793107
CP011895.1	8531593	TGTTTATTACATGCT	8531607
CP011895.1	70986903	TGTTTATTACATGCT	70986917
XM_010757741.1	4378	TATTACATGCTCGGCAA	4394
LK995546.1	4646482	GTGTTTATTACATGCTC	4646498
XM_008876883.1	1892	TTATTACATGCTCGGCA	1908
XM_008876882.1	1957	TTATTACATGCTCGGCA	1973
XM_007408047.1	923	GTGTTTATTACATGCTC	939
FO081946.3	2338	GTGTTTATTACATGCTC	2354
LR812053.1	23405961	TGTTTATTACATGCTCG	23405977
LR812053.1	13835351	TGTTTATTACATGCT	13835365
LR812053.1	13475574	TGTTTATTACATGCT	13475560
LR812053.1	14202541	GTGTTTATTACATGC	14202527
LR812584.1	23158624	TGTTTATTACATGCTCG	23158640
LR812584.1	13560226	TGTTTATTACATGCT	13560240
LR812584.1	13220015	TGTTTATTACATGCT	13220001
LR812520.1	33859889	TTATTACATGCTCGGCA	33859905
LR812520.1	48573693	TGTTTATTACATGCT	48573707
LR812520.1	11621984	TGTTTATTACATGCT	11621970
LR812520.1	35856217	GTGTTTATTACATGC	35856203
CP000612.1	864290	GTGTTTATTACATGCTC	864306
CR854942.16	198069	TGTTTATTACATGCTCG	198085
CP050595.1	51054656	GTGTTTATTACATGCTC	51054640
CP050595.1	43698638	GTGTTTATTACATGC	43698652
CP050642.1	51047486	GTGTTTATTACATGCTC	51047470
CP050642.1	43691978	GTGTTTATTACATGC	43691992
XM_030328380.2	4435	TTATTACATGCTCGGCA	4419
XM_030328379.2	4450	TTATTACATGCTCGGCA	4434
XM_030328378.2	4648	TTATTACATGCTCGGCA	4632
XM_030328376.2	4663	TTATTACATGCTCGGCA	4647
CP040558.1	2730007	GTGTTTAATACATGCTCGGCA	2729987
XM_015263782.2	4941	TGTTTATTACATGCTCG	4925
LR132000.1	13471894	GTGTTTATTACATGCTC	13471878
LR132000.1	15682035	GTGTTTATTACATGCT	15682050
LR132000.1	17723161	TTTATTACATGCTCGG	17723146
LR132000.1	39203337	TGTTTATTACATGCT	39203323
XM_027041670.1	4594	TTATTACATGCTCGGCA	4578
CP033065.1	2858632	GTGTTTAATACATGCTCGGCA	2858612
XM_026134041.1	1180	TGTTTATTACATGCTCG	1164
CP030990.1	47809991	GTTTATTACATGCTCGG	47809975
CP030990.1	50122539	TGTTTATTACATGCT	50122553
CP027087.1	16137289	TTTATTACATGCTCGGC	16137273
XM_023238782.1	4580	TTATTACATGCTCGGCA	4564
CP023464.1	2993631	GTGTTTAATACATGCTCGGCA	2993611

CP011011.1	3015947	GTGTTTAATACATGCTCGGCA	3015927
CP011028.1	3054898	GTGTTTAATACATGCTCGGCA	3054878
CP015399.2	4853392	GTGTTTATTACATGCTC	4853376
LT634572.1	1541167	GTGTTTATTACATGCTC	1541151
XM_015541209.1	4358	TTATTACATGCTCGGCA	4342
XM_012986555.1	1456	TGTTTATTACATGCTCG	1440
XM_012986554.1	1574	TGTTTATTACATGCTCG	1558
XM_012699782.1	69	TTTATTACATGCTCGGC	53
LR812609.1	22130831	TGTTTATTACATGCTCG	22130815
LR812609.1	14729804	TGTTTATTACATGCT	14729818
LR812609.1	14396777	TGTTTATTACATGCT	14396763
LR812609.1	15096455	GTGTTTATTACATGC	15096441
FP074886.7	124177	GTGTTTATTACATGCTC	124161
FP929039.1	1812939	TTTATTACATGCTCGGC	1812923
AC149883.2	21346	GTGTTTATTACATGCTC	21330
AY016023.1	163893	GTGTTTATTACATGCTC	163877
AB006531.1	5925	GTGTTTATTACATGCTC	5909
OA566189.1	69235	TGTTTATTACTTGCTCGGCA	69254
OA564290.1	68076	TTATTACATGCTCGGC	68091
CP064845.1	1321270	TATTACATGCTCGGCA	1321285
CP064846.1	182505	TATTACATGCTCGGCA	182520
OA565377.1	104438	GTGTTTATTACATGCT	104423
CP064904.1	1014034	TTTATTACATGCTCGG	1014019
CP064851.1	27375	TATTACATGCTCGGCA	27360
CP043357.1	27582	TATTACATGCTCGGCA	27567
CP043353.1	28507	TATTACATGCTCGGCA	28492
CP064820.1	2571746	TATTACATGCTCGGCA	2571731

PRODUCTS OF SIGNIFICANT ALIGNMENTS (CDS FEATURE)

- **Citrate synthase 3** GenBank GenBank GenBank GenBank GenBank GenBank GenBank GenBank GenBank GenBank GenBank GenBank GenBank
- **Beta-galactosidase** GenBank GenBank GenBank GenBank GenBank GenBank
- **K(+)/H(+) antiporter NhaP** GenBank GenBank
- **DNA cytosine methyltransferase** GenBank
- **Putative glycosyltransferase EpsJ** GenBank
- **Adenylyl cyclase-associated protein 1-like** GenBank GenBank GenBank GenBank
- **Hypothetical protein** GenBank GenBank GenBank
- **Uncharacterized protein** GenBank
- **Type II secretion system protein E** GenBank
- **Filamin-C** GenBank
- **WD repeat-containing protein on Y chromosome** GenBank
- **Tricyclene synthase 0e23, chloroplastic-like** GenBank GenBank
- **Alpha-catulin-like** GenBank
- **Nonstructural polyprotein** GenBank
- **D-serine transporter DsdX** GenBank GenBank GenBank GenBank

SEQUENCE N° 4/1

(T=1; C=3; G=0; A=2)

SEQUENCE N° 4/1 (T=1; C=3; G=0; A=2)

Original Sequence	**G**	**T**	**G**	**A**	**A**	**A**	**T**	**G**	**G**	**T**	**C**	**A**	**T**	**G**	**T**	**G**	**T**	**G**	**G**	**C**	**G**	**G**
Trend n° 4 (T=1; C=3; G=0; A=2)	**0**	**1**	**0**	**2**	**2**	**2**	**1**	**0**	**0**	**1**	**3**	**2**	**1**	**0**	**1**	**0**	**1**	**0**	**0**	**3**	**0**	**0**
Trend n° 4 (I i D d)	**-**	**I**	**D**	**I**	**i**	**i**	**D**	**D**	**d**	**I**	**I**	**D**	**D**	**D**	**I**	**D**	**I**	**D**	**d**	**I**	**D**	**d**
Trend n° 4/1	**1**	**2**	**1**	**3**	**3**	**3**	**2**	**1**	**1**	**2**	**3**	**2**	**1**	**0**	**2**	**1**	**2**	**1**	**1**	**2**	**1**	**1**
Sequence n° 4/1 (81,82%)	**T**	**A**	**T**	**C**	**C**	**C**	**A**	**T**	**T**	**A**	**C**	**A**	**T**	**G**	**A**	**T**	**A**	**T**	**T**	**A**	**T**	**T**

Query: **TATCCCATTACATGATATTATT**

The bases of **Sequence n ° 4/1** (T = 1; C = 3; G = 0; A = 2) are identical to the bases of **Sequence n ° 3/1** (T = 2; C = 0; G = 3; A = 1). For details, see what is written about **Sequence n ° 3/1**.

SEQUENCE N° 5/1

(T=3; C=1; G=2; A=0)

SEQUENCE N° 5/1 (T=3; C=1; G=2; A=0)

Original Sequence	G	T	G	A	A	A	T	G	G	T	C	A	T	G	T	G	T	G	G	C	G	G
Trend n° 5 (T=3; C=1; G=2; A=0)	2	3	2	0	0	0	3	2	2	3	1	0	3	2	3	2	3	2	2	1	2	2
Trend n° 5 (I i D d)	-	I	D	D	d	d	I	D	d	I	D	D	I	D	I	D	I	D	d	D	I	i
Trend n° 5/1	1	3	2	1	1	1	2	1	1	2	1	0	1	0	1	0	2	1	1	0	1	1
Sequence n° 5/1 (81,82%)	C	T	G	C	C	C	G	C	C	G	C	A	C	A	C	A	G	C	C	A	C	C

Query: **CTGCCCGCCGCACACAGCCACC**

Search Parameters	
Program	blastn
Word size	7
Expect value	1000
Hitlist size	100
Match/Mismatch scores	1,-3
Gapcosts	5,2
Filter string	F
Genetic Code	1

100 sequences selected:

Accession Seq	Description	Comomm Name	Max Score	Total Score	Query C.	E value	Per. Id.	Acc. Len
CP003872.1	Acidovorax sp. KKS102, complete genome	Acidovorax sp. KKS102	44.1	108	100%	0.036	100.00%	5196935
CP050591.1 1/2 2/2 3/3 3/4 3/5 3/6 10/1	Canis lupus familiaris breed Labrador retriever chromosome 10a	dog	38.2	68.4	100%	2.2	100.00%	69938001
CP050611.1 1/2 2/2 3/3 3/4 3/5 3/6 10/1	Canis lupus familiaris breed Labrador retriever chromosome 10b	dog	38.2	68.4	100%	2.2	100.00%	69942321
CP040244.1	Streptomyces exfoliatus strain A1013Y chromosome, complete genome	Streptomyces exfoliatus	38.2	38.2	86%	2.2	100.00%	7646296
CP030277.1	Rhodobiaceae bacterium strain SMS8 chromosome, complete genome	Rhodobiaceae bacterium	38.2	38.2	86%	2.2	100.00%	3862255
CP026304.1	Streptomyces lunaelactis strain MM109 chromosome, complete genome	Streptomyces lunaelactis	38.2	68.4	86%	2.2	100.00%	8396100
XM_020930819.1	PREDICTED: Boleophthalmus pectinirostris peroxisomal biogenesis factor 10 (pex10), mRNA	great blue-spotted mudskipper	38.2	38.2	86%	2.2	100.00%	1190
XM_019534010.1	PREDICTED: Crocodylus porosus transient receptor potential cation channel subfamily C member 3 (TRPC3), transcript variant X3, mRNA	Australian saltwater crocodile	38.2	38.2	86%	2.2	100.00%	2878
XM_019534008.1	PREDICTED: Crocodylus porosus transient receptor potential cation channel subfamily C member 3 (TRPC3), transcript variant X2, mRNA	Australian saltwater crocodile	38.2	38.2	86%	2.2	100.00%	3623
XM_019534007.1	PREDICTED: Crocodylus porosus transient receptor potential cation channel subfamily C member 3 (TRPC3), transcript variant X1, mRNA	Australian saltwater crocodile	38.2	38.2	86%	2.2	100.00%	3704
XM_019494543.1	PREDICTED: Alligator mississippiensis transient receptor potential cation channel subfamily C member 3 (TRPC3), transcript variant X2, mRNA	American alligator	38.2	38.2	86%	2.2	100.00%	3649
XM_019494538.1	PREDICTED: Alligator mississippiensis transient receptor potential cation channel subfamily C member 3 (TRPC3), transcript	American alligator	38.2	38.2	86%	2.2	100.00%	3725

	variant X1, mRNA							
LT629973.1	Akkermansia glycaniphila isolate APytT genome assembly, chromosome: I	Akkermansia glycaniphila	38.2	70.4	95%	2.2	100.00%	3074078
LR812118.1 **1/2 3/2**	Erithacus rubecula genome assembly, chromosome: 15	European robin	36.2	36.2	81%	8.7	100.00%	19110993
AP023165.1	Felis catus Senzu DNA, chromosome: E1, American Shorthair breed	domestic cat	36.2	66.4	86%	8.7	100.00%	65441144
XM_002189001.3	PREDICTED: Taeniopygia guttata protein Wnt-11b-2 (LOC100226560), transcript variant X2, mRNA	zebra finch	36.2	36.2	81%	8.7	100.00%	1909
XM_030272468.2	PREDICTED: Taeniopygia guttata protein Wnt-11b-2 (LOC100226560), transcript variant X1, mRNA	zebra finch	36.2	36.2	81%	8.7	100.00%	2886
LR698958.1	Lachnospiraceae bacterium isolate MGYG-HGUT-01373 genome assembly, chromosome: 1	Lachnospiraceae bacterium	36.2	36.2	81%	8.7	100.00%	2219029
LR597567.1	**Myripristis murdjan genome assembly, chromosome: 18**	pinecone soldierfish	36.2	36.2	81%	8.7	100.00%	29228082
LR597438.1	Salarias fasciatus genome assembly, chromosome: 3	jewelled blenny	36.2	66.4	81%	8.7	100.00%	30411982
LR584060.1	Echeneis naucrates genome assembly, chromosome: 19	live sharksucker	36.2	36.2	100%	8.7	95.45%	17971980
CP033069.1	Acidovorax sp. 1608163 chromosome, complete genome	Acidovorax sp. 1608163	36.2	36.2	100%	8.7	95.45%	5045878
CP029042.1	Streptomyces lydicus strain WYEC 108 chromosome, complete genome	Streptomyces lydicus	36.2	96.6	100%	8.7	100.00%	9125666
XM_026121011.1	PREDICTED: Dromaius novaehollandiae pygopus family PHD finger 2 (PYGO2), mRNA	emu	36.2	36.2	100%	8.7	95.45%	1454
AP017899.1	Sphingopyxis sp. FD7 plasmid pSFD01 DNA, complete sequence	Sphingopyxis sp. FD7	36.2	36.2	81%	8.7	100.00%	242612
CP022129.1	Methylovulum psychrotolerans strain HV10_M2, complete genome	Methylovulum psychrotolerans	36.2	66.4	81%	8.7	100.00%	4923400
LN854557.1	Sodalis glossinidius str. 'morsitans' isolate B4 genome assembly, chromosome: SgGMMB4_chromosome	Sodalis glossinidius str. 'morsitans'	36.2	66.4	81%	8.7	100.00%	4171874
XM_014943042.1	PREDICTED: Calidris pugnax dachsous cadherin-related 2 (DCHS2), mRNA	ruff	36.2	36.2	100%	8.7	95.45%	10705
XM_012646526.1	PREDICTED: Propithecus coquereli ABL proto-oncogene 2, non-receptor tyrosine kinase (ABL2), mRNA	Coquerel's sifaka	36.2	36.2	100%	8.7	95.45%	3381
XM_017700999.2	PREDICTED: Pygocentrus nattereri Cbp/p300-interacting transactivator, with Glu/Asp-rich carboxy-terminal domain, 2 (cited2), mRNA	red-bellied piranha	36.2	36.2	81%	8.7	100.00%	1662
CP062147.1	Komagataeibacter hansenii strain C110 chromosome, complete genome	Komagataeibacter hansenii	36.2	36.2	81%	8.7	100.00%	3530618
CP060748.1	Escherichia coli strain EC96 chromosome, complete genome	Escherichia coli	36.2	36.2	81%	8.7	100.00%	4461910
CP059165.1	Mycobacterium gordonae strain 24T chromosome, complete genome	Mycobacterium gordonae	36.2	36.2	81%	8.7	100.00%	6266765
GU134344.1	Chlamydomonas reinhardtii strain CC425 Mn superoxide dismutase 4 (MSD4) mRNA, complete cds	Chlamydomonas reinhardtii	36.2	36.2	81%	8.7	100.00%	1342
AC236542.1	Oryza minuta clone OM__Ba0185B12, complete sequence	Oryza minuta	36.2	36.2	81%	8.7	100.00%	188666
AC236527.1	Oryza minuta clone OM__Ba0059E04, complete sequence	Oryza minuta	36.2	36.2	81%	8.7	100.00%	131828
XM_001690884.1	Chlamydomonas reinhardtii superoxide dismutase [Mn] (MSD4), partial mRNA	Chlamydomonas reinhardtii	36.2	36.2	81%	8.7	100.00%	657
AP008232.1	Sodalis glossinidius str. 'morsitans' DNA, complete genome	Sodalis glossinidius str. 'morsitans'	36.2	66.4	81%	8.7	100.00%	4171146
LR812078.1	Danio rerio genome assembly, chromosome: 16	zebrafish	34.2	96.6	95%	35	100.00%	55350661
XM_035129734.1	PREDICTED: Zootoca vivipara glycyl-tRNA synthetase 1 (GARS1), mRNA	common lizard	34.2	34.2	77%	35	100.00%	2327
AP023162.1	Felis catus Senzu DNA, chromosome: D2, American Shorthair breed	domestic cat	34.2	34.2	77%	35	100.00%	90643714

XM_034315246.1	PREDICTED: Osmia lignaria uncharacterized LOC117600181 (LOC117600181), transcript variant X2, mRNA	orchard mason bee	34.2	34.2	77%	35	100.00%	1225
XM_034315245.1	PREDICTED: Osmia lignaria uncharacterized LOC117600181 (LOC117600181), transcript variant X1, mRNA	orchard mason bee	34.2	34.2	77%	35	100.00%	1231
CP050951.1 1/2 7/1 7/2 9/1	Pseudomonas putida strain JQ581 chromosome, complete genome	Pseudomonas putida	34.2	34.2	77%	35	100.00%	6656695
CP052884.1	Gordonia ajococcus strain A2 chromosome, complete genome	Gordonia ajococcus	34.2	34.2	77%	35	100.00%	5090254
CP050803.1	Setaria viridis cultivar ME034v chromosome 9	Setaria viridis	34.2	98.6	77%	35	100.00%	56635340
CP050797.1	Setaria viridis cultivar ME034v chromosome 3	Setaria viridis	34.2	64.4	77%	35	100.00%	49814079
XM_033896783.1	PREDICTED: Pecten maximus excitatory amino acid transporter-like (LOC117336294), mRNA	Pecten maximus	34.2	34.2	77%	35	100.00%	2340
XM_033819947.1	Trematosphaeria pertusa protoheme IX farnesyltransferase mitochondrial precursor (BU26DRAFT_149234), mRNA	Trematosphaeria pertusa	34.2	34.2	77%	35	100.00%	1694
LR792829.1	Digitaria exilis annotation	Digitaria exilis	34.2	94.6	77%	35	100.00%	38778756
LR792828.1	Digitaria exilis annotation	Digitaria exilis	34.2	64.4	77%	35	100.00%	40393928
LR792821.1	Digitaria exilis annotation	Digitaria exilis	34.2	64.4	77%	35	100.00%	35142973
LR792820.1	Digitaria exilis annotation	Digitaria exilis	34.2	94.6	77%	35	100.00%	37200191
XM_033591836.1	Didymella exigua CBS 183.55 membrane protein Gsa14p (M421DRAFT_418578), mRNA	Didymella exigua CBS 183.55	34.2	34.2	77%	35	100.00%	3050
LR761614.1	Digitaria exilis genome assembly, chromosome: 5B	Digitaria exilis	34.2	94.6	77%	35	100.00%	38778756
LR761613.1	Digitaria exilis genome assembly, chromosome: 5A	Digitaria exilis	34.2	64.4	77%	35	100.00%	40393928
LR761606.1	Digitaria exilis genome assembly, chromosome: 1B	Digitaria exilis	34.2	64.4	77%	35	100.00%	35142973
LR761605.1	Digitaria exilis genome assembly, chromosome: 1A	Digitaria exilis	34.2	94.6	77%	35	100.00%	37200191
XM_017289033.2	PREDICTED: Drosophila miranda E3 ubiquitin-protein ligase MARCH5 (LOC108157133), transcript variant X2, mRNA	Drosophila miranda	34.2	34.2	77%	35	100.00%	1747
XM_017289032.2	PREDICTED: Drosophila miranda E3 ubiquitin-protein ligase MARCH5 (LOC108157133), transcript variant X1, mRNA	Drosophila miranda	34.2	34.2	77%	35	100.00%	1726
CP050976.1	Streptomyces sp. RPA4-5 chromosome	Streptomyces sp. RPA4-5	34.2	64.4	95%	35	95.24%	9051652
CP050571.1 1/2 2/2 3/3 3/4 3/5 3/6 10/1	Canis lupus familiaris breed Labrador retriever chromosome 20a	dog	34.2	64.4	77%	35	100.00%	58324418
CP050631.1 1/2 2/2 3/3 3/4 3/5 3/6 10/1	Canis lupus familiaris breed Labrador retriever chromosome 20b	dog	34.2	64.4	77%	35	100.00%	58321546
XM_033060848.1	PREDICTED: Catharus ustulatus dachsous cadherin-related 2 (DCHS2), mRNA	Swainson's thrush	34.2	34.2	95%	35	95.24%	1839
AP022720.1	Plectropomus leopardus DNA, chromosome 21, nearly complete sequence	leopard coralgrouper	34.2	64.4	81%	35	100.00%	25625633
XM_015886902.2	PREDICTED: Coturnix japonica SWI/SNF related, matrix associated, actin dependent regulator of chromatin subfamily c member 2 (SMARCC2), transcript variant X6, mRNA	Japanese quail	34.2	34.2	77%	35	100.00%	3347
XM_015886901.2	PREDICTED: Coturnix japonica SWI/SNF related, matrix associated, actin dependent regulator of chromatin subfamily c member 2 (SMARCC2), transcript variant X5, mRNA	Japanese quail	34.2	34.2	77%	35	100.00%	3539
XM_015886900.2	PREDICTED: Coturnix japonica SWI/SNF related, matrix associated, actin dependent regulator of chromatin subfamily c member 2 (SMARCC2), transcript variant X4,	Japanese quail	34.2	34.2	77%	35	100.00%	3542

	mRNA							
XM_015886899.2	PREDICTED: Coturnix japonica SWI/SNF related, matrix associated, actin dependent regulator of chromatin subfamily c member 2 (SMARCC2), transcript variant X3, mRNA	Japanese quail	34.2	34.2	77%	35	100.00%	3629
XM_015886898.2	PREDICTED: Coturnix japonica SWI/SNF related, matrix associated, actin dependent regulator of chromatin subfamily c member 2 (SMARCC2), transcript variant X2, mRNA	Japanese quail	34.2	34.2	77%	35	100.00%	3632
XM_015886897.2	PREDICTED: Coturnix japonica SWI/SNF related, matrix associated, actin dependent regulator of chromatin subfamily c member 2 (SMARCC2), transcript variant X1, mRNA	Japanese quail	34.2	34.2	77%	35	100.00%	3635
CP048641.1	Streptomyces aureoverticillatus strain HN6 chromosome, complete genome	Streptomyces aureoverticillatus	34.2	64.4	100%	35	100.00%	9617533
XM_032198087.1	PREDICTED: Aythya fuligula chromodomain helicase DNA binding protein 6 (CHD6), transcript variant X8, mRNA	tufted duck	34.2	34.2	77%	35	100.00%	10822
XM_037414197.1	PREDICTED: Rhipicephalus microplus probable cation-transporting ATPase 13A3 (LOC119161628), mRNA	southern cattle tick	34.2	34.2	77%	35	100.00%	12141
XM_028886172.2	PREDICTED: Peromyscus leucopus inorganic pyrophosphatase 1 (Ppa1), transcript variant X2, mRNA	white-footed mouse	34.2	34.2	77%	35	100.00%	1301
CP063212.1	Trueperella sp. 19M2397 strain 19OD0592 chromosome, complete genome	Trueperella sp. 19M2397	34.2	64.4	86%	35	100.00%	2373092
CP063213.1	Trueperella sp. 19M2397 strain 15A0121 chromosome, complete genome	Trueperella sp. 19M2397	34.2	34.2	77%	35	100.00%	2327149
CP053291.1	Trueperella sp. 19M2397 chromosome, complete genome	Trueperella sp. 19M2397	34.2	34.2	77%	35	100.00%	2282684
CP062789.1	Janibacter indicus strain TT2 chromosome, complete genome	Janibacter indicus	34.2	34.2	77%	35	100.00%	3663756
AP024094.1	Jeongeupia sp. HS-3 DNA, complete genome	Jeongeupia sp. HS-3	34.2	34.2	77%	35	100.00%	3396565
CP061723.1 **1/2 7/1 7/2 9/1**	Pseudomonas **putida** strain ZXPA-20 chromosome, complete genome	Pseudomonas putida	34.2	64.4	86%	35	100.00%	6511858
AP023477.1								
1/2 2/2 3/2 3/3 6/1 10/1	Homo sapiens DNA chromosome 17 nearly complete genome	human	34.2	64.4	100%	35	100.00%	80688777
XM_036166397.1	PREDICTED: Onychomys torridus coordinator of PRMT5 and differentiation stimulator (Coprs), transcript variant X2, mRNA	southern grasshopper mouse	34.2	34.2	77%	35	100.00%	809
XM_036166396.1	PREDICTED: Onychomys torridus coordinator of PRMT5 and differentiation stimulator (Coprs), transcript variant X1, mRNA	southern grasshopper mouse	34.2	34.2	77%	35	100.00%	866
CP045886.1	Beauveria bassiana strain HN6 chromosome 1	Beauveria bassiana	34.2	94.6	100%	35	100.00%	6424025
AP023408.1	Streptomyces libani subsp. rufus NBRC 15424 DNA, complete genome	Streptomyces libani subsp. rufus	34.2	94.6	95%	35	95.24%	8522827
LR877204.1	Onychomys torridus genome assembly, chromosome: 17	southern grasshopper mouse	34.2	64.4	86%	35	100.00%	64256681
CP046907.1	Streptomyces sp. SCUT-3 chromosome, complete genome	Streptomyces sp. SCUT-3	34.2	64.4	86%	35	100.00%	6075149
XM_035424987.1	PREDICTED: Anguilla anguilla N-acyl-aromatic-L-amino acid amidohydrolase (carboxylate-forming) B-like (LOC118231297), transcript variant X4, mRNA	European eel	34.2	34.2	77%	35	100.00%	1355
XM_035424986.1	PREDICTED: Anguilla anguilla N-acyl-aromatic-L-amino acid amidohydrolase (carboxylate-forming) B-like (LOC118231297), transcript variant X3, mRNA	European eel	34.2	34.2	77%	35	100.00%	1452
XM_035424985.1	PREDICTED: Anguilla anguilla N-acyl-aromatic-L-amino acid	European eel	34.2	34.2	77%	35	100.00%	1581

	amidohydrolase (carboxylate-forming) B-like (LOC118231297), transcript variant X2, mRNA							
XM_035424984.1	PREDICTED: Anguilla anguilla N-acyl-aromatic-L-amino acid amidohydrolase (carboxylate-forming) B-like (LOC118231297), transcript variant X1, mRNA	European eel	34.2	34.2	77%	35	100.00%	1621
CP054635.1	Exserohilum turcica Et28A chromosome 9	Exserohilum turcica Et28A	34.2	34.2	77%	35	100.00%	2134525
CP056779.1	Enterobacter cloacae strain D41-sc-1712200 chromosome, complete genome	Enterobacter cloacae	34.2	34.2	77%	35	100.00%	4826830
CP056117.1	Enterobacter cloacae strain DSM 26481 chromosome	Enterobacter cloacae	34.2	34.2	77%	35	100.00%	4896691
LR812053.1	Danio rerio genome assembly, chromosome: 16	zebrafish	34.2	96.6	95%	35	100.00%	55934183
LS423635.1	Anas platyrhynchos genome assembly, chromosome: 25	mallard	34.2	34.2	77%	35	100.00%	7329573
LR812609.1	Danio rerio genome assembly, chromosome: 16	zebrafish	34.2	96.6	95%	35	100.00%	53959708
LR812584.1	Danio rerio strain Nadia (NA) genome assembly, chromosome: 16	zebrafish	34.2	96.6	95%	35	100.00%	56178339
LR812534.1	Danio kyathit genome assembly, chromosome: 16	Danio kyathit	34.2	96.6	95%	35	100.00%	72833698

ALIGNMENT VIEW FLAT QUERY-ANCHORED WITH LETTERS FOR IDENTITIES

Query range 1: 1 to 22

Query	1	CTGCCCGCCGCACACAGCCACC	22
CP003872.1	787459	CTGCCCGCCGCACACAGCCACC	787480
CP003872.1	1514441	CCGCCGCACACAGCCAC	1514457
CP003872.1	1959438	CTGCCCGCCGCACAC	1959424
CP050591.1	67864953	CTGCCCGCCGCACACAGCC	67864971
CP050591.1	34135575	CCGCACACAGCCACC	34135589
CP050611.1	67867264	CTGCCCGCCGCACACAGCC	67867282
CP050611.1	34132726	CCGCACACAGCCACC	34132740
CP040244.1	3249512	CTGCCCGCCGCACACAGCC	3249530
CP030277.1	2066811	TGCCCGCCGCACACAGCCA	2066829
CP026304.1	6760890	CTGCCCGCCGCACACAGCC	6760872
CP026304.1	7972870	TGCCCGCCGCACACA	7972856
XM_020930819.1	603	CTGCCCGCCGCACACAGCC	585
XM_019534010.1	29	CTGCCCGCCGCACACAGCC	47
XM_019534008.1	29	CTGCCCGCCGCACACAGCC	47
XM_019534007.1	26	CTGCCCGCCGCACACAGCC	44
XM_019494543.1	81	CTGCCCGCCGCACACAGCC	99
XM_019494538.1	73	CTGCCCGCCGCACACAGCC	91
LT629973.1	2024842	GCCCGCCGCACACAGCCAC	2024824
LT629973.1	2207850	CTGCCCGCCGCACACA	2207835
LR812118.1	12113376	TGCCCGCCGCACACAGCC	12113393

AP023165.1	7483224	GCCCGCCGCACACAGCCA	7483241
AP023165.1	42916581	TGCCCGCCGCACACA	42916595
XM_002189001.3	188	CTGCCCGCCGCACACAGC	205
XM_030272468.2	1165	CTGCCCGCCGCACACAGC	1182
LR698958.1	2091545	CCGCCGCACACAGCCACC	2091562
LR597567.1	1890318	CTGCCCGCCGCACACAGC	1890335
LR597438.1	14354613	GCCCGCCGCACACAGCCA	14354630
LR597438.1	2628531	GCCCGCCGCACACAG	2628517
LR584060.1	7230904	CTGCCCGCTGCACACAGCCACC	7230883
CP033069.1	4128501	CTGCCCGCCGCACACGGCCACC	4128480
CP029042.1	9021441	CCGCCGCACACAGCCACC	9021424
CP029042.1	4426031	CTGCCCGCCGCACAC	4426045
CP029042.1	5109719	CCGCCGCACACAGCC	5109705
XM_026121011.1	986	CTGCCCGCCGGACACAGCCACC	1007
AP017899.1	199026	TGCCCGCCGCACACAGCC	199009
CP022129.1	3124410	CCGCCGCACACAGCCACC	3124427
CP022129.1	3119491	CCGCACACAGCCACC	3119505
LN854557.1	573116	GCCCGCCGCACACAGCCA	573133
LN854557.1	2801961	CGCCGCACACAGCCA	2801975
XM_014943042.1	113	CTGCCCGCCGCCCACAGCCACC	92
XM_012646526.1	2611	CTGCCCGCCGCCCACAGCCACC	2632
XM_017700999.2	1261	CTGCCCGCCGCACACAGC	1278
CP062147.1	534142	TGCCCGCCGCACACAGCC	534125
CP060748.1	2150440	CCCGCCGCACACAGCCAC	2150423
CP059165.1	640750	CCGCCGCACACAGCCACC	640767
GU134344.1	824	GCCCGCCGCACACAGCCA	807
AC236542.1	171617	CCCGCCGCACACAGCCAC	171600
AC236527.1	100118	CCCGCCGCACACAGCCAC	100101
XM_001690884.1	423	GCCCGCCGCACACAGCCA	406
AP008232.1	572106	GCCCGCCGCACACAGCCA	572123
AP008232.1	2814214	CGCCGCACACAGCCA	2814228
LR812078.1	10800211	CCGCCGCACACAGCCAC	10800195
LR812078.1	31573378	GCCGCACACAGCCACC	31573393
LR812078.1	16243231	TGCCCGCCGCACACA	16243245
XM_035129734.1	1953	CGCCGCACACAGCCACC	1969
AP023162.1	1969624	CGCCGCACACAGCCACC	1969608
XM_034315246.1	232	CGCCGCACACAGCCACC	248
XM_034315245.1	232	CGCCGCACACAGCCACC	248
CP050951.1	3341770	TGCCCGCCGCACACAGC	3341754
CP052884.1	533557	CGCCGCACACAGCCACC	533541
CP050803.1	34152504	CCGCCGCACACAGCCAC	34152488
CP050803.1	55602526	CCGCCGCACACAGCCAC	55602542
CP050803.1	2350286	CCGCCGCACACAGCC	2350300
CP050797.1	18434526	CTGCCCGCCGCACACAG	18434542
CP050797.1	3528640	TGCCCGCCGCACACA	3528654
XM_033896783.1	247	CCCGCCGCACACAGCCA	231
XM_033819947.1	1129	CGCCGCACACAGCCACC	1113
LR792829.1	6876829	CCGCCGCACACAGCCAC	6876845
LR792829.1	5539866	CGCCGCACACAGCCA	5539852
LR792829.1	37962811	CGCCGCACACAGCCA	37962825
LR792828.1	7121733	CCGCCGCACACAGCCAC	7121749
LR792828.1	39921398	CGCCGCACACAGCCA	39921412
LR792821.1	5378840	CCGCCGCACACAGCCAC	5378824
LR792821.1	14804528	GCCGCACACAGCCAC	14804542
LR792820.1	5334243	CCGCCGCACACAGCCAC	5334227
LR792820.1	19482845	CGCCGCACACAGCCA	19482831
LR792820.1	29045300	CGCCGCACACAGCCA	29045286
XM_033591836.1	2605	CGCCGCACACAGCCACC	2621
LR761614.1	6876829	CCGCCGCACACAGCCAC	6876845
LR761614.1	5539866	CGCCGCACACAGCCA	5539852

LR761614.1	37962811	CGCCGCACACAGCCA	37962825
LR761613.1	7121733	CCGCCGCACACAGCCAC	7121749
LR761613.1	39921398	CGCCGCACACAGCCA	39921412
LR761606.1	5378840	CCGCCGCACACAGCCAC	5378824
LR761606.1	14804528	GCCGCACACAGCCAC	14804542
LR761605.1	5334243	CCGCCGCACACAGCCAC	5334227
LR761605.1	19482845	CGCCGCACACAGCCA	19482831
LR761605.1	29045300	CGCCGCACACAGCCA	29045286
XM_017289033.2	248	GCCCGCCGCACACAGCC	264
XM_017289032.2	227	GCCCGCCGCACACAGCC	243
CP050976.1	8937379	TGCCCGCCGCCCACAGCCACC	8937399
CP050976.1	5011301	CCGCCGCACACAGCC	5011287
CP050571.1	55652137	CTGCCCGCCGCACACAG	55652121
CP050571.1	39983513	CTGCCCGCCGCACAC	39983527
CP050631.1	55647723	CTGCCCGCCGCACACAG	55647707
CP050631.1	39973297	CTGCCCGCCGCACAC	39973311
XM_033060848.1	110	CTGCCCGCCCCACACAGCCAC	90
AP022720.1	23859507	GCCCGCCGCACACAGCC	23859491
AP022720.1	9168476	CGCCGCACACAGCCA	9168490
XM_015886902.2	3236	GCCCGCCGCACACAGCC	3252
XM_015886901.2	3428	GCCCGCCGCACACAGCC	3444
XM_015886900.2	3431	GCCCGCCGCACACAGCC	3447
XM_015886899.2	3518	GCCCGCCGCACACAGCC	3534
XM_015886898.2	3521	GCCCGCCGCACACAGCC	3537
XM_015886897.2	3524	GCCCGCCGCACACAGCC	3540
CP048641.1	2537708	CGCCGCACACAGCCACC	2537692
CP048641.1	9124121	CTGCCCGCCGCACAC	9124135
XM_032198087.1	51	CCCGCCGCACACAGCCA	35
XM_032198086.1	51	CCCGCCGCACACAGCCA	35
XM_032198085.1	51	CCCGCCGCACACAGCCA	35
XM_032198084.1	51	CCCGCCGCACACAGCCA	35
XM_037414197.1	3337	CGCCGCACACAGCCACC	3321
XM_028886172.2	94	CCGCCGCACACAGCCAC	78
CP062789.1	509128	GCCCGCCGCACACAGCC	509144
AP024094.1	384518	CGCCGCACACAGCCACC	384502
CP061723.1	1554612	TGCCCCCCGCACACAGCCA	1554594
CP061723.1	3017879	TGCCCGCCGCACACAGC	3017863
AP023477.1	76694453	CTGCCCGCCGCACACAG	76694437
AP023477.1	74580794	CCGCACACAGCCACC	74580808
XM_036166397.1	56	CCCGCCGCACACAGCCA	40
XM_036166396.1	56	CCCGCCGCACACAGCCA	40
CP045886.1	2323905	CTGCCCGCCGCACACAG	2323889
CP045886.1	5608051	CCGCACACAGCCACC	5608037
CP045886.1	6346327	TGCCCGCCGCACACA	6346313
AP023408.1	27661	TGCCCGCCGCCCACAGCCACC	27641
AP023408.1	613141	CCGCCGCACACAGCC	613155
AP023408.1	7134375	CCCGCCGCACACAGC	7134361
LR877204.1	29872798	CCCGCCGCACACAGTCACC	29872816
LR877204.1	10928480	CCCGCCGCACACAGCCA	10928496
CP046907.1	2339206	CTGCCCGCCGCGCACAGCC	2339188
CP046907.1	2869276	GCCCGCCGCACACAGCC	2869260
XM_035424987.1	217	CCGCCGCACACAGCCAC	201
XM_035424986.1	314	CCGCCGCACACAGCCAC	298
XM_035424985.1	443	CCGCCGCACACAGCCAC	427
XM_035424984.1	483	CCGCCGCACACAGCCAC	467
CP054635.1	839409	GCCCGCCGCACACAGCC	839425
CP056779.1	3830421	GCCCGCCGCACACAGCC	3830437
CP056117.1	2102096	GCCCGCCGCACACAGCC	2102112
LR812053.1	11096994	CCGCCGCACACAGCCAC	11096978
LR812053.1	32205844	GCCGCACACAGCCACC	32205859

LR812053.1	16778702	TGCCCGCCGCACACA	16778716
LS423635.1	2666403	CTGCCCGCCGCACACAG	2666387
LR812609.1	10255845	CCGCCGCACACAGCCAC	10255829
LR812609.1	30652221	GCCGCACACAGCCACC	30652236
LR812609.1	15868371	TGCCCGCCGCACACA	15868357
LR812584.1	10880970	CCGCCGCACACAGCCAC	10880954
LR812584.1	32007184	GCCGCACACAGCCACC	32007199
LR812584.1	16376697	TGCCCGCCGCACACA	16376711
LR812534.1	29479406	CCGCCGCACACAGCCAC	29479422
LR812534.1	40709593	GCCGCACACAGCCACC	40709608
LR812534.1	23502108	TGCCCGCCGCACACA	23502094

PRODUCTS OF SIGNIFICANT ALIGNMENTS (CDS FEATURE)

- **RNA 3'-phosphate cyclase** GenBank; GenBank
- **Zinc/iron permease** GenBank
- **Hypothetical protein** GenBank; GenBank; GenBank; GenBank; GenBank; GenBank; GenBank; GenBank; GenBank
- **Putative esterase** GenBank
- **CsbD family protein** GenBank
- **Peroxisome biogenesis factor 10** GenBank
- **E3 ubiquitin ligase** GenBank
- **C-type lectin fold** GenBank
- **Protein Wnt-11b isoform X2** GenBank
- **Protein Wnt-11b isoform X1** GenBank
- **FxsB family radical SAM/SPASM domain protein** GenBank
- **Septum formation initiator** GenBank; GenBank
- **Lipase** GenBank
- **Glutamate--tRNA ligase** GenBank
- **DNA primase TraC** GenBank
- **Protocadherin-23** GenBank
- **Abelson tyrosine-protein kinase 2** GenBank
- **Cbp/p300-interacting transactivator 2** GenBank
- **Excinuclease ABC subunit UvrA** GenBank
- **Succinylglutamate desuccinylase** GenBank
- **Mn superoxide dismutase 4** GenBank; GenBank
- **Superoxide dismutase [Mn]** GenBank
- **Glycine--tRNA ligase** GenBank
- **Uncharacterized protein LOC117600181** GenBank; GenBank
- **CoA transferase subunit A** GenBank; GenBank
- **Excitatory amino acid transporter-like** GenBank
- **Protoheme IX farnesyltransferase mitochondrial precursor** GenBank
- **Unnamed protein product; ID:Dexi5B01G0009810.1:cds** GenBank
- **Unnamed protein product; ID:Dexi5A01G0009500.1:cds** GenBank
- **Unnamed protein product; ID:Dexi1A01G0006930.1:cds** GenBank
- **Membrane protein Gsa14p** GenBank
- **E3 ubiquitin-protein ligase MARCH5** GenBank; GenBank
- **FAD-dependent oxidoreductase** GenBank
- **LOW QUALITY PROTEIN: protocadherin-23** GenBank
- **SWI/SNF complex subunit SMARCC2** GenBank; GenBank; GenBank; GenBank; GenBank; GenBank
- **DegT/DnrJ/EryC1/StrS aminotransferase** GenBank
- **NADP-dependent succinic semialdehyde dehydrogenase** GenBank
- **LOW QUALITY PROTEIN: probable cation-transporting ATPase 13A3** GenBank
- **NADH-quinone oxidoreductase subunit I** GenBank
- **ZIP family metal transporter** GenBank
- **Monoamine oxidase** GenBank
- **N-acyl-aromatic-L-amino acid amidohydrolase (carboxylate-forming) B-like** GenBank; GenBank; GenBank; GenBank
- **DNA mismatch repair protein MutS** GenBank; GenBank

SEQUENCE N° 5/2

(T=3; C=1; G=2; A=0)

SEQUENCE N° 5/2 (T=3; C=1; G=2; A=0)

Original Sequence	**G**	**T**	**G**	**A**	**A**	**A**	**T**	**G**	**G**	**T**	**C**	**A**	**T**	**G**	**T**	**G**	**T**	**G**	**G**	**C**	**G**	**G**
Trend n° 5 (T=3; C=1; G=2; A=0)	**2**	**3**	**2**	**0**	**0**	**0**	**3**	**2**	**2**	**3**	**1**	**0**	**3**	**2**	**3**	**2**	**3**	**2**	**2**	**1**	**2**	**2**
Trend n° 5 (I i D d)	**-**	**I**	**D**	**D**	**d**	**d**	**I**	**D**	**d**	**I**	**D**	**D**	**I**	**D**	**I**	**D**	**I**	**D**	**d**	**D**	**I**	**i**
Trend n° 5/2	**1**	**2**	**1**	**0**	**0**	**0**	**2**	**1**	**1**	**3**	**2**	**1**	**2**	**1**	**2**	**1**	**2**	**1**	**1**	**0**	**1**	**1**
Sequence n° 5/2 (81,82%)	**C**	**G**	**C**	**A**	**A**	**A**	**G**	**C**	**C**	**T**	**G**	**C**	**G**	**C**	**G**	**C**	**G**	**C**	**C**	**A**	**C**	**C**

Query: **CGCAAAGCCTGCGCGCGCCACC**

Search Parameters	
Program	blastn
Word size	7
Expect value	1000
Hitlist size	100
Match/Mismatch scores	1,-3
Gapcosts	5,2
Filter string	F
Genetic Code	1

98 sequences selected:

Accession Seq	Description	Comomm Name	Max Score	Total Score	Query C.	E value	Per. Id.	Acc. Len
CP003872.1	Acidovorax sp. KKS102, complete genome	Acidovorax sp. KKS102	44.1	108	100%	0.036	100.00%	5196935
CP007444.1	Dyella jiangningensis strain SBZ 3-12, complete genome	Dyella jiangningensis	40.1	100	90%	0.58	100.00%	5396991
CP060388.1	Sphingobium sp. KCTC 72723 chromosome, complete genome	Sphingobium sp. KCTC 72723	40.1	70.4	95%	0.58	100.00%	4372398
CP049736.1	Bordetella hinzii strain FY01 chromosome, complete genome	Bordetella hinzii	38.2	128	95%	2.3	100.00%	5045412
CP044059.1	Bordetella hinzii strain FDAARGOS_621 chromosome, complete genome	Bordetella hinzii	38.2	128	95%	2.3	100.00%	4860452
CP021400.1	Bordetella hinzii strain 243-2 chromosome, complete genome	Bordetella hinzii	38.2	128	95%	2.3	100.00%	4811419
CP021399.1	Bordetella hinzii strain TR-1212 chromosome, complete genome	Bordetella hinzii	38.2	128	95%	2.3	100.00%	4930681
CP021398.1	Bordetella hinzii strain 14-3425 chromosome, complete genome	Bordetella hinzii	38.2	98.6	95%	2.3	100.00%	4885078
CP021397.1	Bordetella hinzii strain 4449 chromosome, complete genome	Bordetella hinzii	38.2	128	95%	2.3	100.00%	4898820
CP021396.1	Bordetella hinzii strain 4134 chromosome, complete genome	Bordetella hinzii	38.2	128	95%	2.3	100.00%	5042960
CP021395.1	Bordetella hinzii strain SV2 chromosome, complete genome	Bordetella hinzii	38.2	128	95%	2.3	100.00%	4809490
CP034452.1	Mesorhizobium sp. M6A.T.Cr.TU.016.01.1.1 chromosome	Mesorhizobium sp. M6A.T.Cr.TU.016.01.1.1	38.2	98.6	86%	2.3	100.00%	6784580
CP024172.1	Bordetella hinzii strain H720 chromosome, complete genome	Bordetella hinzii	38.2	128	95%	2.3	100.00%	4859886
CP018911.1	Glycocaulis alkaliphilus strain 6B-8, complete genome	Glycocaulis alkaliphilus	38.2	68.4	95%	2.3	100.00%	3019659
LR134382.1	Bordetella hinzii strain NCTC13199 genome assembly, chromosome: 1	Bordetella hinzii	38.2	128	95%	2.3	100.00%	5033537
XM_025517752.1	Ceraceosorus guamensis DUF1740-domain-containing protein (IE81DRAFT_92921), mRNA	Ceraceosorus guamensis	38.2	38.2	86%	2.3	100.00%	3308
LT906461.1	Bordetella hinzii strain NCTC13200 genome assembly, chromosome: 1	Bordetella hinzii	38.2	128	95%	2.3	100.00%	4913806
CP012077.1	Bordetella hinzii strain H568, complete genome	Bordetella hinzii	38.2	128	95%	2.3	100.00%	4859884
CP012076.1	Bordetella hinzii strain F582, complete genome	Bordetella hinzii	38.2	128	95%	2.3	100.00%	4912977
CP002156.1	Parvularcula bermudensis HTCC2503 strain HTCC2503, complete genome	Parvularcula bermudensis HTCC2503	38.2	38.2	86%	2.3	100.00%	2902643
CP054599.1	Sulfitobacter pseudonitzschiae strain H46 chromosome, complete genome	Sulfitobacter pseudonitzschiae	36.2	66.4	86%	9.1	100.00%	3891549
CP019930.1	Bordetella avium strain Ba002, complete genome	Bordetella avium	36.2	66.4	86%	9.1	100.00%	3691350

CP019929.1	Bordetella avium strain 14-4061, complete genome	Bordetella avium	36.2	36.2	81%	9.1	100.00%	3836621
CP019928.1	Bordetella avium strain SV1, complete genome	Bordetella avium	36.2	36.2	81%	9.1	100.00%	3682222
CP052769.1	Mesorhizobium japonicum strain R7AstarV2 chromosome, complete genome	Mesorhizobium japonicum	36.2	36.2	81%	9.1	100.00%	6530396
CP052770.1	Mesorhizobium japonicum strain R7ANSstar chromosome, complete genome	Mesorhizobium japonicum	36.2	36.2	81%	9.1	100.00%	6028604
CP051773.1	Mesorhizobium japonicum strain R7Astar chromosome, complete genome	Mesorhizobium japonicum	36.2	36.2	81%	9.1	100.00%	6530403
CP051772.1	Mesorhizobium japonicum R7A chromosome, complete genome	Mesorhizobium japonicum R7A	36.2	36.2	81%	9.1	100.00%	6530403
CP051684.1	Duganella sp. AF9R3 chromosome, complete genome	Duganella sp. AF9R3	36.2	128	90%	9.1	100.00%	6532782
CP041041.1	Paracoccus sp. AK26 chromosome, complete genome	Paracoccus sp. AK26	36.2	66.4	81%	9.1	100.00%	2787975
CP044550.1	Hydrogenophaga sp. BPS33 plasmid pBPS33-1, complete sequence	Hydrogenophaga sp. BPS33	36.2	72.4	81%	9.1	100.00%	339634
CP044549.1	Hydrogenophaga sp. BPS33 chromosome, complete genome	Hydrogenophaga sp. BPS33	36.2	229	86%	9.1	100.00%	6325781
LR738417.1	Lutra lutra genome assembly, chromosome: 15	Eurasian river otter	36.2	36.2	81%	9.1	100.00%	69992071
CP046823.1	Vibrio mimicus strain 06-2455 chromosome 1, complete sequence	Vibrio mimicus	36.2	36.2	81%	9.1	100.00%	3051319
CP033366.1	Mesorhizobium japonicum R7A chromosome, complete genome	Mesorhizobium japonicum R7A	36.2	36.2	81%	9.1	100.00%	6529994
AP019684.1	Xanthomonas campestris pv. campestris MAFF302021 DNA, complete genome	Xanthomonas campestris pv. campestris	36.2	191	100%	9.1	100.00%	5048651
AP019682.1	Xanthomonas campestris pv. campestris MAFF106712 DNA, complete genome	Xanthomonas campestris pv. campestris	36.2	191	100%	9.1	100.00%	5002720
CP029484.1	Xanthomonas campestris pv. campestris strain Xcc8004_Xcc2 chromosome	Xanthomonas campestris pv. campestris	36.2	443	100%	9.1	100.00%	5161275
CP029483.1	Xanthomonas campestris pv. campestris strain Xcc8004_Xcc1 chromosome	Xanthomonas campestris pv. campestris	36.2	443	100%	9.1	100.00%	5169528
CP039247.1	Corynebacterium endometrii strain LMM-1653 chromosome, complete genome	Corynebacterium endometrii	36.2	36.2	81%	9.1	100.00%	2477061
CP031395.1	Hylemonella gracilis strain NS1 chromosome	Hylemonella gracilis	36.2	66.4	90%	9.1	100.00%	3822345
CP034450.1	Mesorhizobium sp. M4B.F.Ca.ET.058.02.1.1 chromosome	Mesorhizobium sp. M4B.F.Ca.ET.058.02.1.1	36.2	96.6	81%	9.1	100.00%	6461656
CP032509.1	Georhizobium profundi strain WS11 chromosome, complete genome	Georhizobium profundi	36.2	96.6	81%	9.1	100.00%	4356607
CP024676.1	Citrobacter pasteurii strain UMH17 chromosome, complete genome	Citrobacter pasteurii	36.2	36.2	81%	9.1	100.00%	4654874
CP031606.1	Pseudomonas sp. phDV1 chromosome, complete genome	Pseudomonas sp. phDV1	36.2	66.4	81%	9.1	100.00%	4727682
CP034811.1	Paracoccus sp. Arc7-R13 plasmid unnamed1, complete sequence	Paracoccus sp. Arc7-R13	36.2	36.2	81%	9.1	100.00%	199685
CP025522.1	Bordetella avium strain JBBA chromosome, complete genome	Bordetella avium	36.2	36.2	81%	9.1	100.00%	3837904
CP025521.1	Bordetella avium strain Owl19 chromosome, complete genome	Bordetella avium	36.2	36.2	81%	9.1	100.00%	3785260
LR134485.1	Citrobacter youngae strain NCTC13709 genome assembly, chromosome: 1	Citrobacter youngae	36.2	36.2	81%	9.1	100.00%	4867355
MH449685.1	Aeromonas hydrophila O35 antigen biosynthesis gene cluster, complete sequence	Aeromonas hydrophila	36.2	36.2	81%	9.1	100.00%	48465
CP025997.1	Stenotrophomonas sp. YAU14D1_LEIMI4_1 chromosome	Stenotrophomonas sp. YAU14D1_LEIMI4_1	36.2	66.4	90%	9.1	100.00%	4319112
CP028843.1	Methylobacterium currus strain PR1016A chromosome 1	Methylobacterium currus	36.2	36.2	81%	9.1	100.00%	6553813
CP023467.1	Dickeya dadantii strain DSM 18020 chromosome, complete genome	Dickeya dadantii	36.2	96.6	95%	9.1	100.00%	4997541
CP014030.2	Citrobacter sp. FDAARGOS_156 strain FDAARGOS_155 chromosome, complete genome	Citrobacter sp. FDAARGOS_156	36.2	36.2	81%	9.1	100.00%	4727498
CP026124.1	Achromobacter sp. AONIH1 chromosome, complete genome	Achromobacter sp. AONIH1	36.2	66.4	81%	9.1	100.00%	6648943
CP025750.1	Xanthomonas campestris pv. campestris strain 3811 chromosome, complete genome	Xanthomonas campestris pv. campestris	36.2	161	100%	9.1	100.00%	5072566
CP021913.1	Sagittula sp. P11 chromosome, complete genome	Sagittula sp. P11	36.2	66.4	100%	9.1	95.45%	4624102
CP017389.1	Xanthomonas campestris pv. campestris str. CN16 chromosome, complete genome	Xanthomonas campestris pv. campestris str. CN16	36.2	161	100%	9.1	100.00%	5023147
CP017323.1	Xanthomonas campestris pv. campestris str. CN15 chromosome, complete genome	Xanthomonas campestris pv. campestris str. CN15	36.2	161	100%	9.1	100.00%	5030187

CP017319.1	Xanthomonas campestris pv. campestris str. CN18 chromosome, complete genome	Xanthomonas campestris pv. campestris str. CN18	36.2	191	100%	9.1	100.00%	5133109
CP017317.1	Xanthomonas campestris pv. campestris str. CN14 chromosome, complete genome	Xanthomonas campestris pv. campestris str. CN14	36.2	161	100%	9.1	100.00%	5033560
CP017310.1	Xanthomonas campestris pv. campestris str. CN12 chromosome, complete genome	Xanthomonas campestris pv. campestris str. CN12	36.2	191	100%	9.1	100.00%	5115137
CP017308.1	Xanthomonas campestris pv. campestris str. CN03 chromosome, complete genome	Xanthomonas campestris pv. campestris str. CN03	36.2	191	100%	9.1	100.00%	5035984
CP017307.1	Xanthomonas campestris pv. campestris str. CN17 chromosome, complete genome	Xanthomonas campestris pv. campestris str. CN17	36.2	191	100%	9.1	100.00%	5122268
CP001655.1	Dickeya chrysanthemi Ech1591 chromosome, complete genome	Dickeya chrysanthemi Ech1591	36.2	36.2	81%	9.1	100.00%	4813854
CP023344.1	Nibricoccus aquaticus strain HZ-65 chromosome, complete genome	Nibricoccus aquaticus	36.2	98.6	100%	9.1	100.00%	4730447
CP014136.1	Gibbsiella quercinecans strain FRB97, complete genome	Gibbsiella quercinecans	36.2	100	90%	9.1	100.00%	5548506
CP021424.1	Armatimonadetes bacterium Uphvl-Ar2 genome	Armatimonadetes bacterium Uphvl-Ar2	36.2	66.4	81%	9.1	100.00%	2633541
CP018812.1	Paraburkholderia sp. SOS3 chromosome 2, complete sequence	Paraburkholderia sp. SOS3	36.2	130	86%	9.1	100.00%	2991945
CP017433.1	Achromobacter ruhlandii strain SCCH3:ACH 33-1365 genome	Achromobacter ruhlandii	36.2	130	95%	9.1	100.00%	6366684
CP012661.1	Defluviimonas alba strain cai42 chromosome, complete genome	Defluviimonas alba	36.2	36.2	81%	9.1	100.00%	4738298
CP014646.1	Thauera humireducens strain SgZ-1, complete genome	Thauera humireducens	36.2	96.6	90%	9.1	100.00%	4171754
CP014505.1	Burkholderia sp. PAMC 28687 strain PAMC28687 chromosome 1, complete sequence	Burkholderia sp. PAMC 28687	36.2	36.2	81%	9.1	100.00%	3260450
CP014306.1	Burkholderia sp. PAMC 26561 chromosome 1, complete sequence	Burkholderia sp. PAMC 26561	36.2	36.2	81%	9.1	100.00%	3397021
CP013730.1	Burkholderia cepacia JBK9 chromosome 1, complete sequence	Burkholderia cepacia JBK9	36.2	66.4	100%	9.1	100.00%	3790084
CP011946.1	Xanthomonas campestris pv. campestris strain 17 genome	Xanthomonas campestris pv. campestris	36.2	159	100%	9.1	100.00%	4994426
CP012146.1	Xanthomonas campestris pv. campestris strain ICMP 4013, complete genome	Xanthomonas campestris pv. campestris	36.2	159	100%	9.1	100.00%	4908383
CP012145.1	Xanthomonas campestris pv. campestris strain ICMP 21080, complete genome	Xanthomonas campestris pv. campestris	36.2	191	100%	9.1	100.00%	4911121
CP012046.1	Achromobacter xylosoxidans strain MN001, complete genome	Achromobacter xylosoxidans	36.2	130	95%	9.1	100.00%	5876049
AP014568.1	Serpentinomonas raichei DNA, complete genome	Serpentinomonas raichei	36.2	36.2	81%	9.1	100.00%	2574013
XM_005849537.1	Chlorella variabilis expressed protein (CHLNCDRAFT_143052) mRNA, complete cds	Chlorella variabilis	36.2	36.2	100%	9.1	95.45%	894
CP004886.1	Mycobacterium tuberculosis str. Haarlem/NITR202, complete genome	Mycobacterium tuberculosis str. Haarlem/NITR202	36.2	36.2	100%	9.1	95.45%	4404786
CP003740.1	Octadecabacter antarcticus 307, complete genome	Octadecabacter antarcticus 307	36.2	36.2	81%	9.1	100.00%	4812600
AP024109.1	Mesorhizobium sp. J8 DNA, complete genome	Mesorhizobium sp. J8	36.2	96.6	81%	9.1	100.00%	6690860
XM_037898474.1	PREDICTED: Chelonia mydas microtubule interacting and trafficking domain containing 1 (MITD1), transcript variant X3, mRNA	Green sea turtle	36.2	36.2	81%	9.1	100.00%	915
XM_037898464.1	PREDICTED: Chelonia mydas microtubule interacting and trafficking domain containing 1 (MITD1), transcript variant X2, mRNA	Green sea turtle	36.2	36.2	81%	9.1	100.00%	879
XM_007068977.3	PREDICTED: Chelonia mydas microtubule interacting and trafficking domain containing 1 (MITD1), transcript variant X1, mRNA	Green sea turtle	36.2	36.2	81%	9.1	100.00%	980
LR884483.1	Acanthopagrus latus isolate v.2019 genome assembly, chromosome: 24	yellowfin seabream	36.2	36.2	81%	9.1	100.00%	16042954
AP024096.1	Ktedonobacteria bacterium brp13 DNA, complete genome	Ktedonobacteria bacterium brp13	36.2	36.2	81%	9.1	100.00%	8879929
CP062066.1	Xanthomonas campestris strain M28 chromosome, complete genome	Xanthomonas campestris	36.2	221	100%	9.1	100.00%	5102828
CP060008.1	Pseudomonas sp. B11D7D chromosome, complete genome	Pseudomonas sp. B11D7D	36.2	36.2	81%	9.1	100.00%	5387171
CP058243.1	Xanthomonas campestris pv. raphani strain MAFF106181 chromosome, complete genome	Xanthomonas campestris pv. raphani	36.2	161	100%	9.1	100.00%	4942039
CP002789.1	Xanthomonas campestris pv. raphani	Xanthomonas campestris	36.2	128	81%	9.1	100.00%	4941214

	756C, complete genome	pv. raphani 756C						
CP002038.1	Dickeya dadantii 3937, complete genome	Dickeya dadantii 3937	36.2	68.4	81%	9.1	100.00%	4922802
FN563999.1	Muriella zofingiensis lcyB gene for chloroplast lycopene beta cyclase precursor, strain SAG 211-14	Chromochloris zofingiensis	36.2	36.2	81%	9.1	100.00%	3683
FN563998.1	Muriella zofingiensis mRNA for chloroplast lycopene beta cyclase precursor (lcyB gene), strain SAG 211-14	Chromochloris zofingiensis	36.2	36.2	81%	9.1	100.00%	2131
AM920689.1	Xanthomonas campestris pv. campestris complete genome, strain B100	Xanthomonas campestris pv. campestris	36.2	161	100%	9.1	100.00%	5079002
AM167904.1	Bordetella avium 197N complete genome	Bordetella avium 197N	36.2	66.4	86%	9.1	100.00%	3732255
CP000050.1	Xanthomonas campestris pv. campestris str. 8004, complete genome	Xanthomonas campestris pv. campestris str. 8004	36.2	221	100%	9.1	100.00%	5148708

ALIGNMENT VIEW FLAT QUERY-ANCHORED WITH LETTERS FOR IDENTITIES

Query range 1: 1 to 22

Query	1	CGCAAAGCCTGCGCGCGCCACC	22
CP007444.1	3529116	CGCAAAGCCTGCGCGCGCCA	3529135
CP007444.1	2514930	AAAGCCTGCGCGCGC	2514916
CP007444.1	2948117	CGCAAAGCCTGCGCG	2948131
CP060388.1	51295	CGCAAAGCCTGCGCGCGCCA	51314
CP060388.1	45720	GCCTGCGCGCGCCAC	45734
CP049736.1	450092	CGCAAAGCCTGCGCGCGCC	450074
CP049736.1	429383	GCCTGCGCGCGCCAC	429397
CP049736.1	2423660	AAAGCCTGCGCGCGC	2423646
CP049736.1	4680990	AGCCTGCGCGCGCCA	4681004
CP044059.1	3501842	CGCAAAGCCTGCGCGCGCC	3501824
CP044059.1	538227	AAAGCCTGCGCGCGC	538213
CP044059.1	2724592	AGCCTGCGCGCGCCA	2724606
CP044059.1	3478283	GCCTGCGCGCGCCAC	3478297
CP021400.1	1251408	CGCAAAGCCTGCGCGCGCC	1251426
CP021400.1	1272124	GCCTGCGCGCGCCAC	1272110
CP021400.1	2029798	AGCCTGCGCGCGCCA	2029784
CP021400.1	4200893	AAAGCCTGCGCGCGC	4200907
CP021399.1	1281326	CGCAAAGCCTGCGCGCGCC	1281344
CP021399.1	1302039	GCCTGCGCGCGCCAC	1302025
CP021399.1	2029406	AGCCTGCGCGCGCCA	2029392
CP021399.1	4247514	AAAGCCTGCGCGCGC	4247528
CP021398.1	1325213	CGCAAAGCCTGCGCGCGCC	1325231
CP021398.1	1348788	GCCTGCGCGCGCCAC	1348774
CP021398.1	4311205	AAAGCCTGCGCGCGC	4311219
CP021397.1	1305690	CGCAAAGCCTGCGCGCGCC	1305708
CP021397.1	1326420	GCCTGCGCGCGCCAC	1326406
CP021397.1	2085053	AGCCTGCGCGCGCCA	2085039
CP021397.1	4287345	AAAGCCTGCGCGCGC	4287359
CP021396.1	1406756	CGCAAAGCCTGCGCGCGCC	1406774
CP021396.1	1430294	GCCTGCGCGCGCCAC	1430280
CP021396.1	2177356	AGCCTGCGCGCGCCA	2177342
CP021396.1	4402985	AAAGCCTGCGCGCGC	4402999
CP021395.1	1212088	CGCAAAGCCTGCGCGCGCC	1212106
CP021395.1	1232773	GCCTGCGCGCGCCAC	1232759
CP021395.1	1957055	AGCCTGCGCGCGCCA	1957041
CP021395.1	4187215	AAAGCCTGCGCGCGC	4187229
CP034452.1	4342267	CGCAAAGCCTGCGCGCGCC	4342285
CP034452.1	1509088	GCAAAGCCTGCGCGC	1509074
CP034452.1	3514502	GCAAAGCCTGCGCGC	3514516
CP024172.1	1257457	CGCAAAGCCTGCGCGCGCC	1257475
CP024172.1	1281009	GCCTGCGCGCGCCAC	1280995
CP024172.1	2007304	AGCCTGCGCGCGCCA	2007290
CP024172.1	4193644	AAAGCCTGCGCGCGC	4193658
CP018911.1	996246	GCAAAGCCTGCGCGCGCCA	996264
CP018911.1	1018589	CCTGCGCGCGCCACC	1018603
LR134382.1	3910504	CGCAAAGCCTGCGCGCGCC	3910486
LR134382.1	878833	AAAGCCTGCGCGCGC	878819
LR134382.1	3139793	AGCCTGCGCGCGCCA	3139807
LR134382.1	3886973	GCCTGCGCGCGCCAC	3886987
XM_025517752.1	1990	AAAGCCTGCGCGCGCCACC	2008
LT906461.1	4459305	CGCAAAGCCTGCGCGCGCC	4459323

LT906461.1	315255	AGCCTGCGCGCGCCA	315241
LT906461.1	2529602	AAAGCCTGCGCGCGC	2529616
LT906461.1	4482859	GCCTGCGCGCGCCAC	4482845
CP012077.1	393657	CGCAAAGCCTGCGCGCGCC	393639
CP012077.1	370105	GCCTGCGCGCGCCAC	370119
CP012077.1	2317355	AAAGCCTGCGCGCGC	2317341
CP012077.1	4503693	AGCCTGCGCGCGCCA	4503707
CP012076.1	414739	CGCAAAGCCTGCGCGCGCC	414721
CP012076.1	391185	GCCTGCGCGCGCCAC	391199
CP012076.1	2344442	AAAGCCTGCGCGCGC	2344428
CP012076.1	4557958	AGCCTGCGCGCGCCA	4557972
CP002156.1	832596	GCAAAGCCTGCGCGCGCCA	832614
CP054599.1	1153759	GCAAAGCCCGCGCGCGCCA	1153777
CP054599.1	276727	GCAAAGCCTGCGCGCGCC	276744
CP019930.1	2692072	CGCAAAGCCTGCGCGCGC	2692089
CP019930.1	1412336	AAGCCTGCGCGCGCC	1412350
CP019929.1	2727522	CGCAAAGCCTGCGCGCGC	2727539
CP019928.1	2647316	CGCAAAGCCTGCGCGCGC	2647333
CP052769.1	471372	AAGCCTGCGCGCGCCACC	471355
CP052770.1	471372	AAGCCTGCGCGCGCCACC	471355
CP051773.1	471372	AAGCCTGCGCGCGCCACC	471355
CP051772.1	471372	AAGCCTGCGCGCGCCACC	471355
CP051684.1	1599592	CGCAAAGCCTGCGCGCGC	1599609
CP051684.1	11757	GCAAAGCCTGCGCGCG	11742
CP051684.1	3718406	AGCCTGCGCGCGCCA	3718420
CP051684.1	4991950	AAAGCCTGCGCGCGC	4991964
CP041041.1	279582	GCAAAGCCTGCGCGCGCC	279599
CP041041.1	1755355	GCAAAGCCTGCGCGC	1755341
CP044550.1	82872	AAGCCTGCGCGCGCCACC	82855
CP044550.1	87100	AAGCCTGCGCGCGCCACC	87117
CP044549.1	1545431	AAGCCTGCGCGCGCCACC	1545448
CP044549.1	4565613	AAGCCTGCGCGCGCCACC	4565630
CP044549.1	5740624	AAGCCTGCGCGCGCCACC	5740607
CP044549.1	275463	AAAGCCTGCGCGCGC	275449
CP044549.1	2192626	AAGCCTGCGCGCGCC	2192612
CP044549.1	3836392	GCCTGCGCGCGCCAC	3836406
CP044549.1	5311789	GCCTGCGCGCGCCAC	5311803
LR738417.1	13084461	CGCAAAGCCTGCGCGCGC	13084478
CP046823.1	272185	CAAAGCCTGCGCGCGCCA	272168
CP033366.1	471372	AAGCCTGCGCGCGCCACC	471355
AP019684.1	3762612	CGCAAAGCCTGCGCGCGC	3762595
AP019684.1	134075	GCCTGCGCGCGCCACC	134090
AP019684.1	4185175	CGCAAAGCCTGCGCGC	4185190
AP019684.1	3011064	GCCTGCGCGCGCCAC	3011078
AP019684.1	4396954	GCAAAGCCTGCGCGC	4396940
AP019684.1	5026308	CGCAAAGCCTGCGCG	5026322
AP019682.1	3697093	CGCAAAGCCTGCGCGCGC	3697076
AP019682.1	135710	GCCTGCGCGCGCCACC	135725
AP019682.1	4145217	CGCAAAGCCTGCGCGC	4145232
AP019682.1	3021719	GCCTGCGCGCGCCAC	3021733
AP019682.1	4356147	GCAAAGCCTGCGCGC	4356133
AP019682.1	4980234	CGCAAAGCCTGCGCG	4980248
CP029484.1	3792248	CGCAAAGCCTGCGCGCGC	3792231
CP029484.1	73263	GCCTGCGCGCGCCACC	73278
CP029484.1	4208420	CGCAAAGCCTGCGCGC	4208435
CP029484.1	2553501	GCAAAGCCTGCGCGC	2553487
CP029484.1	3057760	GCCTGCGCGCGCCAC	3057774
CP029484.1	4424193	GCAAAGCCTGCGCGC	4424179
CP029484.1	5064311	CGCAAAGCCTGCGCG	5064325
CP029483.1	3042580	CGCAAAGCCTGCGCGCGC	3042597

CP029483.1	1610516	GCCTGCGCGCGCCACC	1610501
CP029483.1	2626401	CGCAAAGCCTGCGCGC	2626386
CP029483.1	1770510	CGCAAAGCCTGCGCG	1770496
CP029483.1	2410628	GCAAAGCCTGCGCGC	2410642
CP029483.1	3777068	GCCTGCGCGCGCCAC	3777054
CP029483.1	4281327	GCAAAGCCTGCGCGC	4281341
CP039247.1	2192226	CAAAGCCTGCGCGCGCCA	2192243
CP031395.1	3525202	AAAGCCTGCGCGCGCCAC	3525219
CP031395.1	2337432	GCAAAGCCTGCGCGC	2337418
CP034450.1	4086861	CGCAAAGCCTGCGCGCGC	4086878
CP034450.1	764197	AAAGCCTGCGCGCGC	764211
CP034450.1	4906734	GCAAAGCCTGCGCGC	4906748
CP032509.1	4089974	CGCAAAGCCTGCGCGCGC	4089991
CP032509.1	658493	AAAGCCTGCGCGCGC	658479
CP032509.1	1460723	CGCAAAGCCTGCGCG	1460737
CP024676.1	1844362	CGCAAAGCCTGCGCGCGC	1844379
CP031606.1	2902444	CAAAGCCTGCGCGCGCCA	2902427
CP031606.1	4679086	AAGCCTGCGCGCGCC	4679072
CP034811.1	170428	CGCAAAGCCTGCGCGCGC	170445
CP025522.1	2728803	CGCAAAGCCTGCGCGCGC	2728820
CP025521.1	2746473	CGCAAAGCCTGCGCGCGC	2746490
LR134485.1	2222998	CGCAAAGCCTGCGCGCGC	2222981
MH449685.1	21601	CAAAGCCTGCGCGCGCCA	21618
CP025997.1	284550	AAGCCTGCGCGCGCCACC	284567
CP025997.1	1037954	CAAAGCCTGCGCGCG	1037940
CP028843.1	5110156	CGCAAAGCCTGCGCGCGC	5110139
CP023467.1	2972650	CGCAAAGCCTGCGCGCGC	2972633
CP023467.1	3990021	GCCTGCGCGCGCCAC	3990007
CP023467.1	4291883	CGCAAAGCCTGCGCG	4291897
CP014030.2	2843726	CGCAAAGCCTGCGCGCGC	2843743
CP026124.1	2935136	AAGCCTGCGCGCGCCACC	2935119
CP026124.1	1189807	GCCTGCGCGCGCCAC	1189821
CP025750.1	3786086	CGCAAAGCCTGCGCGCGC	3786069
CP025750.1	137235	GCCTGCGCGCGCCACC	137250
CP025750.1	4208327	CGCAAAGCCTGCGCGC	4208342
CP025750.1	4420324	GCAAAGCCTGCGCGC	4420310
CP025750.1	5050122	CGCAAAGCCTGCGCG	5050136
CP021913.1	1359938	CGCAAAGCCTGCGGGCGCCACC	1359917
CP021913.1	460680	GCCTGCGCGCGCCAC	460666
CP017389.1	3706375	CGCAAAGCCTGCGCGCGC	3706358
CP017389.1	133681	GCCTGCGCGCGCCACC	133696
CP017389.1	4144639	CGCAAAGCCTGCGCGC	4144654
CP017389.1	4355298	GCAAAGCCTGCGCGC	4355284
CP017389.1	4990731	CGCAAAGCCTGCGCG	4990745
CP017323.1	3723299	CGCAAAGCCTGCGCGCGC	3723282
CP017323.1	133677	GCCTGCGCGCGCCACC	133692
CP017323.1	4161583	CGCAAAGCCTGCGCGC	4161598
CP017323.1	4372242	GCAAAGCCTGCGCGC	4372228
CP017323.1	5007672	CGCAAAGCCTGCGCG	5007686
CP017319.1	3751445	CGCAAAGCCTGCGCGCGC	3751428
CP017319.1	167322	GCCTGCGCGCGCCACC	167337
CP017319.1	4196367	CGCAAAGCCTGCGCGC	4196382
CP017319.1	3039234	GCCTGCGCGCGCCAC	3039248
CP017319.1	4407303	GCAAAGCCTGCGCGC	4407289
CP017319.1	5110592	CGCAAAGCCTGCGCG	5110606
CP017317.1	3706634	CGCAAAGCCTGCGCGCGC	3706617
CP017317.1	133679	GCCTGCGCGCGCCACC	133694
CP017317.1	4144914	CGCAAAGCCTGCGCGC	4144929
CP017317.1	4375615	GCAAAGCCTGCGCGC	4375601
CP017317.1	5011045	CGCAAAGCCTGCGCG	5011059

CP017310.1	1296408	CGCAAAGCCTGCGCGCGC	1296425
CP017310.1	135689	GCCTGCGCGCGCCACC	135704
CP017310.1	849581	CGCAAAGCCTGCGCGC	849566
CP017310.1	638652	GCAAAGCCTGCGCGC	638666
CP017310.1	1989563	GCCTGCGCGCGCCAC	1989549
CP017310.1	5092620	CGCAAAGCCTGCGCG	5092634
CP017308.1	3753299	CGCAAAGCCTGCGCGCGC	3753282
CP017308.1	136990	GCCTGCGCGCGCCACC	137005
CP017308.1	4169082	CGCAAAGCCTGCGCGC	4169097
CP017308.1	3047944	GCCTGCGCGCGCCAC	3047958
CP017308.1	4378878	GCAAAGCCTGCGCGC	4378864
CP017308.1	5013462	CGCAAAGCCTGCGCG	5013476
CP017307.1	3735944	CGCAAAGCCTGCGCGCGC	3735927
CP017307.1	168153	GCCTGCGCGCGCCACC	168168
CP017307.1	4183720	CGCAAAGCCTGCGCGC	4183735
CP017307.1	3028346	GCCTGCGCGCGCCAC	3028360
CP017307.1	4394654	GCAAAGCCTGCGCGC	4394640
CP017307.1	5099751	CGCAAAGCCTGCGCG	5099765
CP001655.1	844379	CGCAAAGCCTGCGCGCGC	844396
CP023344.1	3068088	CGCAAAGCCTGCGCGCGC	3068071
CP023344.1	2906494	GCCTGCGCGCGCCACC	2906479
CP023344.1	3417709	AAAGCCTGCGCGCGC	3417695
CP014136.1	4374229	CGCAAAGCCTGCGCGCGC	4374246
CP014136.1	1630069	CGCAAAGCCTGCGCGCG	1630085
CP014136.1	1471713	AGCCTGCGCGCGCCA	1471699
CP021424.1	1534736	CGCAAAGCCTGCGCGCGC	1534753
CP021424.1	2086388	AAAGCCTGCGCGCGC	2086374
CP018812.1	1322521	CGCAAAGCCTGCGCGCGC	1322504
CP018812.1	1215529	CGCAAAGCCTGCGCGCG	1215513
CP018812.1	1937311	AAGCCTGCGCGCGCC	1937325
CP018812.1	2034019	AAAGCCTGCGCGCGC	2034033
CP017433.1	5327546	GCAAGGCCTGCGCGCGCCAC	5327565
CP017433.1	5035039	CGCAAAGCCTGCGCGCGC	5035022
CP017433.1	4041323	AAAGCCTGCGCGCGCC	4041308
CP017433.1	3345473	GCCTGCGCGCGCCAC	3345487
CP012661.1	3626542	GCAAAGCCTGCGCGCGCC	3626559
CP014646.1	1933296	GCAAAGCCTGCGCGCGCC	1933313
CP014646.1	404329	GCCTGCGCGCGCCAC	404315
CP014646.1	3524267	AAAGCCTGCGCGCGC	3524281
CP014505.1	2846840	CGCAAAGCCTGCGCGCGC	2846823
CP014306.1	765148	CGCAAAGCCTGCGCGCGC	765131
CP013730.1	1114084	AAGCCTGCGCGCGCCACC	1114101
CP013730.1	485586	CGCAAAGCCTGCGCG	485572
CP011946.1	1158162	CGCAAAGCCTGCGCGCGC	1158179
CP011946.1	136547	GCCTGCGCGCGCCACC	136562
CP011946.1	2287402	GCAAAGCCTGCGCGC	2287416
CP011946.1	4346739	GCAAAGCCTGCGCGC	4346725
CP011946.1	4972787	CGCAAAGCCTGCGCG	4972801
CP012146.1	3640701	CGCAAAGCCTGCGCGCGC	3640684
CP012146.1	146313	GCCTGCGCGCGCCACC	146328
CP012146.1	2521792	GCAAAGCCTGCGCGC	2521778
CP012146.1	4260207	GCAAAGCCTGCGCGC	4260193
CP012146.1	4886813	CGCAAAGCCTGCGCG	4886827
CP012145.1	3630270	CGCAAAGCCTGCGCGCGC	3630253
CP012145.1	137595	GCCTGCGCGCGCCACC	137610
CP012145.1	4046421	CGCAAAGCCTGCGCGC	4046436
CP012145.1	2937583	GCCTGCGCGCGCCAC	2937597
CP012145.1	4260084	GCAAAGCCTGCGCGC	4260070
CP012145.1	4889287	CGCAAAGCCTGCGCG	4889301
CP012046.1	5512652	GCAAGGCCTGCGCGCGCCAC	5512671

CP012046.1	5220813	CGCAAAGCCTGCGCGCGC	5220796
CP012046.1	4253926	AAAGCCTGCGCGCGCC	4253911
CP012046.1	3524234	GCCTGCGCGCGCCAC	3524248
AP014568.1	2191961	CGCAAAGCCTGCGCGCGC	2191944
XM_005849537.1	411	CGCAAAGCCTGCGCCCGCCACC	390
CP004886.1	3609442	CGCAAAGCCCGCGCGCGCCACC	3609463
CP003740.1	3869610	GCAAAGCCTGCGCGCGCC	3869593
AP024109.1	1908047	CGCAAAGCCTGCGCGCGC	1908064
AP024109.1	118685	AAAGCCTGCGCGCGC	118699
AP024109.1	717368	GCAAAGCCTGCGCGC	717382
XM_037898474.1	57	AAGCCTGCGCGCGCCACC	40
XM_037898464.1	57	AAGCCTGCGCGCGCCACC	40
XM_007068977.3	56	AAGCCTGCGCGCGCCACC	39
LR884483.1	6407061	CGCAAAGCCTGCGCGCGC	6407044
AP024096.1	6396520	CGCAAAGCCTGCGCGCGC	6396503
CP062066.1	3767676	CGCAAAGCCTGCGCGCGC	3767659
CP062066.1	144567	GCCTGCGCGCGCCACC	144582
CP062066.1	4207108	CGCAAAGCCTGCGCGC	4207123
CP062066.1	2614591	GCAAAGCCTGCGCGC	2614577
CP062066.1	3072312	GCCTGCGCGCGCCAC	3072326
CP062066.1	4420098	GCAAAGCCTGCGCGC	4420084
CP062066.1	5080952	CGCAAAGCCTGCGCG	5080966
CP060008.1	2102525	CGCAAAGCCTGCGCGCGC	2102508
CP058243.1	3652732	CGCAAAGCCTGCGCGCGC	3652715
CP058243.1	136197	GCCTGCGCGCGCCACC	136212
CP058243.1	4081788	CGCAAAGCCTGCGCGC	4081803
CP058243.1	4292203	GCAAAGCCTGCGCGC	4292189
CP058243.1	4923527	CGCAAAGCCTGCGCG	4923541
CP002789.1	1248491	CGCAAAGCCTGCGCGCGC	1248508
CP002789.1	818572	CGCAAAGCCTGCGCGC	818557
CP002789.1	609183	GCAAAGCCTGCGCGC	609197
CP002789.1	4918339	CGCAAAGCCTGCGCG	4918353
CP002038.1	2882472	CGCAAAGCCTGCGCGCGC	2882455
CP002038.1	697082	CGCAAAGCCTGCGCGC	697067
FN563999.1	441	AAGCCTGCGCGCGCCACC	424
FN563998.1	441	AAGCCTGCGCGCGCCACC	424
AE008922.1	1184098	CGCAAAGCCTGCGCGCGC	1184115
AE008922.1	138222	GCCTGCGCGCGCCACC	138237
AE008922.1	767969	CGCAAAGCCTGCGCGC	767954
AE008922.1	1919634	GCCTGCGCGCGCCAC	1919620
AE008922.1	2363298	GCAAAGCCTGCGCGC	2363312
AE008922.1	4430498	GCAAAGCCTGCGCGC	4430484
AE008922.1	5054392	CGCAAAGCCTGCGCG	5054406
AM920689.1	3790946	CGCAAAGCCTGCGCGCGC	3790929
AM920689.1	137129	GCCTGCGCGCGCCACC	137144
AM920689.1	4213655	CGCAAAGCCTGCGCGC	4213670
AM920689.1	4425418	GCAAAGCCTGCGCGC	4425404
AM920689.1	5056686	CGCAAAGCCTGCGCG	5056700
AM167904.1	2733000	CGCAAAGCCTGCGCGCGC	2733017
AM167904.1	1453432	AAGCCTGCGCGCGCC	1453446
CP000050.1	3856451	CGCAAAGCCTGCGCGCGC	3856434
CP000050.1	137818	GCCTGCGCGCGCCACC	137833
CP000050.1	4272579	CGCAAAGCCTGCGCGC	4272594
CP000050.1	2616238	GCAAAGCCTGCGCGC	2616224
CP000050.1	3120500	GCCTGCGCGCGCCAC	3120514
CP000050.1	4488081	GCAAAGCCTGCGCGC	4488067
CP000050.1	5126916	CGCAAAGCCTGCGCG	5126930
BA000012.4	3988442	AAGCCTGCGCGCGCCACC	3988459

SEQUENCE N° 5/2 (T=3; C=1; G=2; A=0)

Query: **CGCAAAGCCTGCGCGCGCCACC**

Search Parameters	
Program	blastn
Word size	7
Expect value	1000
Hitlist size	**1000**
Match/Mismatch scores	1,-3
Gapcosts	5,2
Filter string	F
Genetic Code	1

1000 sequences selected:

Accession Seq	Description	Comomm Name	Max Score	Total Score	Query C.	E value	Per. Id.	Acc. Len
LR131983.1	Gouania willdenowi genome assembly, chromosome: 14	blunt-snouted clingfish	34.2	34.2	77%	36	100.00%	36801689
XM_036991933.1	PREDICTED: Manis javanica forkhead box L2 (FOXL2), mRNA	Malayan pangolin	34.2	34.2	77%	36	100.00%	4467
XM_036418791.1	*PREDICTED: Pipistrellus kuhlii E2F transcription factor 1 (E2F1), mRNA*	Kuhl's pipistrelle	34.2	34.2	77%	36	100.00%	2857
LR877215.1	Acomys russatus genome assembly, chromosome: 4	golden spiny mouse	34.2	34.2	77%	36	100.00%	85305823
LR862363.1	*Pipistrellus pipistrellus genome assembly, chromosome: 7*	common pipistrelle	34.2	34.2	77%	36	100.00%	84548104
LR584428.1	Salmo trutta genome assembly, chromosome: 13	river trout	32.2	32.2	90%	142	95.00%	91488822
LR131997.1	Gouania willdenowi genome assembly, chromosome: 9	blunt-snouted clingfish	32.2	32.2	72%	142	100.00%	43149712
OB794647.1	3_Tms_b3v08	Timema monikensis	32.2	32.2	72%	142	100.00%	146912
XM_036877357.1	PREDICTED: Manis pentadactyla forkhead box L2 (FOXL2), mRNA	Chinese pangolin	32.2	32.2	72%	142	100.00%	3254
OB794647.1	3_Tms_b3v08	Timema monikensis	32.2	32.2	72%	142	100.00%	146912

SEQUENCE N° 6/1

(T=2; C=3; G=0; A=1)

SEQUENCE N° 6/1 (T=2; C=3; G=0; A=1)

Original Sequence	**G**	**T**	**G**	**A**	**A**	**A**	**T**	**G**	**G**	**T**	**C**	**A**	**T**	**G**	**T**	**G**	**T**	**G**	**G**	**C**	**G**	**G**
Trend n° 6 (T=2; C=3; G=0; A=1)	**0**	**2**	**0**	**1**	**1**	**1**	**2**	**0**	**0**	**2**	**3**	**1**	**2**	**0**	**2**	**0**	**2**	**0**	**0**	**3**	**0**	**0**
Trend n° 6 (I i D d)	**-**	**I**	**D**	**I**	**i**	**i**	**I**	**D**	**d**	**I**	**I**	**D**	**I**	**D**	**I**	**D**	**I**	**D**	**d**	**I**	**D**	**d**
Trend n° 6/1	**1**	**3**	**1**	**2**	**2**	**2**	**3**	**1**	**1**	**2**	**3**	**0**	**1**	**0**	**3**	**1**	**3**	**1**	**1**	**2**	**1**	**1**
Sequence n° 6/1 (86,36%)	**A**	**C**	**A**	**T**	**T**	**T**	**C**	**A**	**A**	**T**	**C**	**G**	**A**	**G**	**C**	**A**	**C**	**A**	**A**	**T**	**A**	**A**

Query: **ACATTTCAATCGAGCACAATAA**

Search Parameters	
Program	blastn
Word size	7
Expect value	1000
Hitlist size	100
Match/Mismatch scores	1,-3
Gapcosts	5,2
Filter string	F
Genetic Code	1

99 sequences selected:

Accession Seq	Description	Comomm Name	Max Score	Total Score	Query C.	E value	Per. Id.	Acc. Len
CP017414.1	Halobacteriovorax marinus strain BE01 chromosome, complete genome	Halobacteriovorax marinus	38.2	38.2	86%	2.2	100.00%	3393238
XM_012937820.2	Schistosoma haematobium putative cys-loop ligand gated ion channel subunit (MS3_0016276), partial mRNA	Schistosoma haematobium	38.2	38.2	86%	2.2	100.00%	1829
FQ312005.1	Bacteriovorax marinus SJ genome	Halobacteriovorax marinus SJ	38.2	38.2	86%	2.2	100.00%	3435933
CP039349.1	Vigna unguiculata cultivar Xiabao 2 chromosome Vu04	cowpea	38.2	128	95%	2.2	100.00%	55796738
CP053108.1	Halomonas sp. TA22 chromosome	Halomonas sp. TA22	36.2	36.2	81%	8.7	100.00%	3732597
CP053087.1	Halomonas sp. TA6 chromosome	Halomonas sp. TA6	36.2	36.2	81%	8.7	100.00%	3732597
CP053025.1	Halomonas sp. PA5 chromosome	Halomonas sp. PA5	36.2	36.2	81%	8.7	100.00%	3732595
XM_032002300.1	Kwoniella shandongensis uncharacterized protein (CI109_001110), partial mRNA	Kwoniella shandongensis	36.2	36.2	81%	8.7	100.00%	1983
CP032556.1	Gossypium raimondii isolate D5-4 chromosome D5_04	Gossypium raimondii	36.2	36.2	100%	8.7	95.45%	49886343
LR131966.1	Parambassis ranga genome assembly, chromosome: 16	Indian glassy fish	36.2	36.2	81%	8.7	100.00%	22465611
CP021731.1	Azoarcus sp. DN11 chromosome, complete genome	Azoarcus sp. DN11	36.2	36.2	81%	8.7	100.00%	4956835
LK052913.1	Cyberlindnera fabianii strain YJS4271 genomic scaffold, scaffold CYFA0S28	Cyberlindnera fabianii	36.2	36.2	81%	8.7	100.00%	102844
AP019873.1	Marchantia polymorpha subsp. ruderalis Tak-1 DNA, chromosome: 8	Marchantia polymorpha subsp. ruderalis	36.2	66.4	81%	8.7	100.00%	21314552
CP032574.1	Gossypium turneri isolate D10-2 chromosome D10_04	Gossypium turneri	36.2	36.2	100%	8.7	95.45%	52521518
CP026254.1	Scophthalmus maximus chromosome 12	turbot	36.2	36.2	81%	8.7	100.00%	24811384
CU571057.2	M.truncatula DNA sequence from clone MTH2-52O4 on chromosome 3, complete sequence	barrel medic	36.2	36.2	81%	8.7	100.00%	136599
AC144482.13	Medicago truncatula clone mth2-10p1, complete sequence	barrel medic	36.2	36.2	81%	8.7	100.00%	121112
AC126786.22	Medicago truncatula clone mth2-8c2, complete sequence	barrel medic	36.2	36.2	81%	8.7	100.00%	115134
LR735560.1	Sarcophilus harrisii genome assembly, chromosome: X	Tasmanian devil	34.2	34.2	77%	35	100.00%	83081154
CP045339.1	Vibrio sp. THAF190c plasmid pTHAF190c_a, complete sequence	Vibrio sp. THAF190c	34.2	34.2	77%	35	100.00%	1722714

LR537134.1	**Sparus aurata genome assembly, chromosome: 14**	gilthead seabream	34.2	34.2	77%	35	100.00%	25861820
XM_004516808.3	PREDICTED: Cicer arietinum pentatricopeptide repeat-containing protein At1g18485 (LOC101509188), transcript variant X1, mRNA	chickpea	34.2	34.2	77%	35	100.00%	3685
XR_003335731.1	PREDICTED: Papaver somniferum uncharacterized LOC113300348 (LOC113300348), transcript variant X7, ncRNA	opium poppy	34.2	34.2	77%	35	100.00%	4266
XR_003335730.1	PREDICTED: Papaver somniferum uncharacterized LOC113300348 (LOC113300348), transcript variant X6, ncRNA	opium poppy	34.2	34.2	77%	35	100.00%	4281
XR_003335729.1	PREDICTED: Papaver somniferum uncharacterized LOC113300348 (LOC113300348), transcript variant X5, ncRNA	opium poppy	34.2	34.2	77%	35	100.00%	4248
XR_003335728.1	PREDICTED: Papaver somniferum uncharacterized LOC113300348 (LOC113300348), transcript variant X4, ncRNA	opium poppy	34.2	34.2	77%	35	100.00%	4264
XR_003335727.1	PREDICTED: Papaver somniferum uncharacterized LOC113300348 (LOC113300348), transcript variant X3, ncRNA	opium poppy	34.2	34.2	77%	35	100.00%	4165
XR_003335726.1	PREDICTED: Papaver somniferum uncharacterized LOC113300348 (LOC113300348), transcript variant X2, ncRNA	opium poppy	34.2	34.2	77%	35	100.00%	4509
XR_003335725.1	PREDICTED: Papaver somniferum uncharacterized LOC113300348 (LOC113300348), transcript variant X1, ncRNA	opium poppy	34.2	34.2	77%	35	100.00%	4270
XM_008140785.2	PREDICTED: Eptesicus fuscus homeobox C6 (HOXC6), transcript variant X1, mRNA	big brown bat	34.2	34.2	77%	35	100.00%	10240
CP032299.1	Anopheles stephensi strain Indian chromosome 2R	Asian malaria mosquito	34.2	126	90%	35	100.00%	58051481
CP032232.1	Anopheles stephensi strain SDA-500 chromosome 2R	Asian malaria mosquito	34.2	64.4	90%	35	100.00%	62568927
XM_025214551.1	PREDICTED: Alligator sinensis cystathionine beta-synthase (LOC102379548), transcript variant X9, mRNA	Chinese alligator	34.2	34.2	77%	35	100.00%	4432
XM_025214550.1	PREDICTED: Alligator sinensis cystathionine beta-synthase (LOC102379548), transcript variant X8, mRNA	Chinese alligator	34.2	34.2	77%	35	100.00%	4437
XM_025214549.1	PREDICTED: Alligator sinensis cystathionine beta-synthase (LOC102379548), transcript variant X7, mRNA	Chinese alligator	34.2	34.2	77%	35	100.00%	4660
XM_025214548.1	PREDICTED: Alligator sinensis cystathionine beta-synthase (LOC102379548), transcript variant X6, mRNA	Chinese alligator	34.2	34.2	77%	35	100.00%	4505
XM_025214547.1	PREDICTED: Alligator sinensis cystathionine beta-synthase (LOC102379548), transcript variant X5, mRNA	Chinese alligator	34.2	34.2	77%	35	100.00%	4544
XM_025214546.1	PREDICTED: Alligator sinensis cystathionine beta-synthase (LOC102379548), transcript variant X4, mRNA	Chinese alligator	34.2	34.2	77%	35	100.00%	4507
XM_025214545.1	PREDICTED: Alligator sinensis cystathionine beta-synthase (LOC102379548), transcript variant X3, mRNA	Chinese alligator	34.2	34.2	77%	35	100.00%	4538
XM_025214544.1	PREDICTED: Alligator sinensis cystathionine beta-synthase (LOC102379548), transcript variant X2, mRNA	Chinese alligator	34.2	34.2	77%	35	100.00%	4528
XM_025214543.1	PREDICTED: Alligator sinensis cystathionine beta-synthase (LOC102379548), transcript variant X1, mRNA	Chinese alligator	34.2	34.2	77%	35	100.00%	4767
AC277723.1 **1/2 2/2 3/2 3/3 5/1 10/1**	Homo sapiens chromosome 7 clone VMRC62-424L4, complete sequence	human	34.2	34.2	77%	35	100.00%	162561
CP020671.1	Oryzias latipes strain Hd-rR	Japanese medaka	34.2	34.2	77%	35	100.00%	34573382

	chromosome 7 sequence								
CP019094.1	Spirodela polyrhiza strain 9509 chromosome 2 sequence	great duckweed	34.2	34.2	77%	35	100.00%	11374105
XM_019596250.1	PREDICTED: Lupinus angustifolius pre-mRNA cleavage factor Im 25 kDa subunit 2-like (LOC109353889), mRNA	narrow-leaved blue lupine	34.2	34.2	77%	35	100.00%	618
XM_019487198.1	PREDICTED: Alligator mississippiensis cystathionine beta-synthase (LOC102560906), transcript variant X4, mRNA	American alligator	34.2	34.2	77%	35	100.00%	8490
XM_019487197.1	PREDICTED: Alligator mississippiensis cystathionine beta-synthase (LOC102560906), transcript variant X3, mRNA	American alligator	34.2	34.2	77%	35	100.00%	8722
XM_019487196.1	PREDICTED: Alligator mississippiensis cystathionine beta-synthase (LOC102560906), transcript variant X2, mRNA	American alligator	34.2	34.2	77%	35	100.00%	8647
XM_006260474.3	PREDICTED: Alligator mississippiensis cystathionine beta-synthase (LOC102560906), transcript variant X1, mRNA	American alligator	34.2	34.2	77%	35	100.00%	8711
XM_014378881.1	PREDICTED: Trichogramma pretiosum HMG box-containing protein 4-like (LOC106657379), mRNA	Trichogramma pretiosum	34.2	34.2	77%	35	100.00%	2138
XM_013452927.1	Necator americanus hypothetical protein mRNA	Necator americanus	34.2	34.2	77%	35	100.00%	321
XM_009826460.1	Aphanomyces astaci hypothetical protein mRNA	Aphanomyces astaci	34.2	34.2	77%	35	100.00%	1469
XM_007940675.1	PREDICTED: Orycteropus afer afer DEAD (Asp-Glu-Ala-Asp) box helicase 21 (DDX21), mRNA	Orycteropus afer afer	34.2	34.2	77%	35	100.00%	2413
XM_006835838.1	PREDICTED: Chrysochloris asiatica DEAD (Asp-Glu-Ala-Asp) box helicase 21 (DDX21), mRNA	Cape golden mole	34.2	34.2	77%	35	100.00%	2413
AC162031.12	Medicago truncatula chromosome 8 clone mth2-194a22, complete sequence	barrel medic	34.2	34.2	77%	35	100.00%	123712
XM_034868321.1	PREDICTED: Etheostoma cragini peptidyl arginine deiminase, type II (padi2), transcript variant X2, mRNA	Arkansas darter	34.2	34.2	77%	35	100.00%	3052
XM_034868320.1	PREDICTED: Etheostoma cragini peptidyl arginine deiminase, type II (padi2), transcript variant X1, mRNA	Arkansas darter	34.2	34.2	77%	35	100.00%	3067
LR792568.1	Pseudochaenichthys georgianus genome assembly, chromosome: 23	South Georgia icefish	34.2	64.4	95%	35	100.00%	25555492
XM_032513447.1	PREDICTED: Etheostoma spectabile peptidyl arginine deiminase 2 (padi2), mRNA	orangethroat darter	34.2	34.2	77%	35	100.00%	3066
LR738563.1	Neostethus bicornis genome assembly, chromosome: 2	Neostethus bicornis	34.2	34.2	77%	35	100.00%	16327240
LR736850.1	Pecten maximus genome assembly, chromosome: 13	Pecten maximus	34.2	34.2	77%	35	100.00%	39606779
CP032376.1	Trichinella spiralis isolate ISS534 chromosome X	Trichinella spiralis	34.2	34.2	77%	35	100.00%	15367684
CP032373.1	Trichinella spiralis isolate TY2 chromosome X	Trichinella spiralis	34.2	34.2	77%	35	100.00%	15373592
CP032367.1	Trichinella spiralis isolate g279 chromosome X	Trichinella spiralis	34.2	34.2	77%	35	100.00%	15349119
LR597437.1	Salarias fasciatus genome assembly, chromosome: 2	jewelled blenny	34.2	34.2	77%	35	100.00%	35747288
CP032252.1	Gossypioides kirkii chromosome KI_10	Gossypioides kirkii	34.2	34.2	77%	35	100.00%	47257441
AP019376.1	Thermosporothrix sp. COM3 DNA, nearly complete genome	Thermosporothrix sp. COM3	34.2	34.2	77%	35	100.00%	7674705
CP025674.1	**Ipomoea triloba cultivar NCNSP0323 chromosome 15**	trilobed morning glory	34.2	34.2	77%	35	100.00%	27111266
XM_011477429.3	PREDICTED: Oryzias latipes synaptotagmin-2 (LOC101161806), transcript variant X3, mRNA	Japanese medaka	34.2	34.2	77%	35	100.00%	7081
XM_011477428.3	PREDICTED: Oryzias latipes synaptotagmin-2 (LOC101161806), transcript variant X2, mRNA	Japanese medaka	34.2	34.2	77%	35	100.00%	7016
XM_004070826.4	PREDICTED: Oryzias latipes synaptotagmin-2 (LOC101161806), transcript variant X1, mRNA	Japanese medaka	34.2	34.2	77%	35	100.00%	7090

XM_023062592.1	PREDICTED: Onthophagus taurus uncharacterized LOC111427443 (LOC111427443), transcript variant X2, mRNA	Onthophagus taurus	34.2	34.2	77%	35	100.00%	2585
XM_023062591.1	PREDICTED: Onthophagus taurus uncharacterized LOC111427443 (LOC111427443), transcript variant X1, mRNA	Onthophagus taurus	34.2	34.2	77%	35	100.00%	2735
CP023120.1	Lupinus angustifolius cultivar Tanjil chromosome LG-08	narrow-leaved blue lupine	34.2	34.2	77%	35	100.00%	30636785
XR_002168388.1	PREDICTED: Hippocampus comes uncharacterized LOC109516675 (LOC109516675), transcript variant X1, ncRNA	tiger tail seahorse	34.2	34.2	77%	35	100.00%	966
JX899680.1	Saccoglossus kowalevskii scaffold_166 genomic sequence	Saccoglossus kowalevskii	34.2	34.2	77%	35	100.00%	950760
CP055244.1								
1/2 2/2 3/2 3/3 3/5 8/1 9/1 10/1	Solanum tuberosum cultivar Solyntus chromosome 11	potato	34.2	34.2	77%	35	100.00%	57022023
LR884470.1	Acanthopagrus latus isolate v.2019 genome assembly, chromosome: 11	yellowfin seabream	34.2	34.2	77%	35	100.00%	31321781
LR812569.1	Danio rerio strain Nadia (NA) genome assembly, chromosome: 1	zebrafish	34.2	34.2	95%	35	95.24%	58221445
LR812544.1	Danio rerio strain Cooch Behar (CB) genome assembly, chromosome: 1	zebrafish	34.2	34.2	95%	35	95.24%	54094920
AC073347.3 **1/2 2/2 3/2 3/3 5/1 10/1**	Homo sapiens BAC clone RP11-775L16 from 7, complete sequence	human	34.2	34.2	77%	35	100.00%	197144
XM_028886192.2	PREDICTED: Peromyscus leucopus DExD-box helicase 21 (Ddx21), transcript variant X2, mRNA	white-footed mouse	32.2	32.2	72%	137	100.00%	3521
XM_037200273.1	PREDICTED: Peromyscus leucopus DExD-box helicase 21 (Ddx21), transcript variant X1, mRNA	white-footed mouse	32.2	32.2	72%	137	100.00%	3525
XM_037233253.1	PREDICTED: Pollicipes pollicipes glycerol-3-phosphate acyltransferase 3-like (LOC119109572), transcript variant X3, mRNA	Pollicipes pollicipes	32.2	32.2	72%	137	100.00%	4432
XM_037233252.1	PREDICTED: Pollicipes pollicipes glycerol-3-phosphate acyltransferase 3-like (LOC119109572), transcript variant X2, mRNA	Pollicipes pollicipes	32.2	32.2	72%	137	100.00%	4195
XM_037233251.1	PREDICTED: Pollicipes pollicipes glycerol-3-phosphate acyltransferase 3-like (LOC119109572), transcript variant X1, mRNA	Pollicipes pollicipes	32.2	32.2	72%	137	100.00%	4198
XM_037232415.1	PREDICTED: Pollicipes pollicipes glycerol-3-phosphate acyltransferase 3-like (LOC119108870), transcript variant X2, mRNA	Pollicipes pollicipes	32.2	32.2	72%	137	100.00%	4190
XM_037232414.1	PREDICTED: Pollicipes pollicipes glycerol-3-phosphate acyltransferase 3-like (LOC119108870), transcript variant X1, mRNA	Pollicipes pollicipes	32.2	32.2	72%	137	100.00%	4193
CP062047.1	Macrobrachium nipponense isolate FS-2020 chromosome 34	Macrobrachium nipponense	32.2	96.6	72%	137	100.00%	71390351
CP055245.1								
1/2 2/2 3/2 3/3 3/5 8/1 9/1 10/1	Solanum tuberosum cultivar Solyntus chromosome 12	potato	32.2	62.4	90%	137	100.00%	68020514
LR884475.1	Acanthopagrus latus isolate v.2019 genome assembly, chromosome: 16	yellowfin seabream	32.2	32.2	72%	137	100.00%	25616698
XM_028194100.2	PREDICTED: Monomorium pharaonis eukaryotic translation initiation factor 5B (LOC105830051), mRNA	pharaoh ant	32.2	32.2	90%	137	95.00%	1818
XM_037532010.1	PREDICTED: Pygocentrus nattereri dynein, axonemal, heavy chain 1	red-bellied piranha	32.2	32.2	72%	137	100.00%	13212

	(dnah1), mRNA							
CP062010.1	Macrobrachium nipponense isolate FS-2020 chromosome 44	Macrobrachium nipponense	32.2	62.4	81%	137	100.00%	52992041
CP062040.1	Macrobrachium nipponense isolate FS-2020 chromosome 19	Macrobrachium nipponense	32.2	32.2	90%	137	95.00%	86821439
CP055236.1 1/2 2/2 3/2 3/3 3/5 8/1 9/1 10/1	Solanum tuberosum cultivar Solyntus chromosome 3	potato	32.2	32.2	72%	137	100.00%	63701590
CP055235.1 1/2 2/2 3/2 3/3 3/5 8/1 9/1 10/1	Solanum tuberosum cultivar Solyntus chromosome 2	potato	32.2	62.4	72%	137	100.00%	42706079
CP053152.1	Carposina sasakii isolate BJYQ chromosome 05	Carposina sasakii	32.2	32.2	72%	137	100.00%	15951206
XM_036036937.1	PREDICTED: Anopheles stephensi ankyrin repeat and LEM domain-containing protein 1 (LOC118503538), mRNA	Asian malaria mosquito	32.2	32.2	72%	137	100.00%	3090

ALIGNMENT VIEW FLAT QUERY-ANCHORED WITH LETTERS FOR IDENTITIES

Query range 1: 1 to 22

Query	1	ACATTTCAATCGAGCACAATAA	22
CP017414.1	3009550	TTTCAATCGAGCACAATAA	3009532
XM_012937820.2	953	ACATTTCAATCGAGCACAA	935
FQ312005.1	3039136	TTTCAATCGAGCACAATAA	3039118
CP039349.1	37482378	ATTTCAATCGAGCACAATA	37482396
CP039349.1	20575420	CAATCGAGCACAATA	20575406
CP039349.1	20576953	CAATCGAGCACAATA	20576939
CP039349.1	44282803	ACATTTCAATCGAGC	44282789
CP053108.1	3244107	TTTCAATCGAGCACAATA	3244090
CP053087.1	2270311	TTTCAATCGAGCACAATA	2270294
CP053025.1	1157315	TTTCAATCGAGCACAATA	1157298
XM_032002300.1	754	ATTTCAATCGAGCACAAT	737
CP032556.1	38143954	ACATTTCAATCTAGCACAATAA	38143933
LR131966.1	16203278	ACATTTCAATCGAGCACA	16203261
CP021731.1	3204783	CATTTCAATCGAGCACAA	3204766
LK052913.1	4606	CATTTCAATCGAGCACAA	4589
AP019873.1	5235583	TTCAATCGAGCACAATAA	5235600
AP019873.1	6577719	TCAATCGAGCACAAT	6577705
CP032574.1	12923002	ACATTTCAATCTAGCACAATAA	12923023
CP026254.1	19605221	ACATTTCAATCGAGCACA	19605238
CU571057.2	86562	ATTTCAATCGAGCACAAT	86579
AC144482.13	71081	ATTTCAATCGAGCACAAT	71098
AC126786.22	62624	ATTTCAATCGAGCACAAT	62641
LR735560.1	14668288	TTCAATCGAGCACAATA	14668272
CP045339.1	501666	TCAATCGAGCACAATAA	501650
LR537134.1	15016187	ATTTCAATCGAGCACAA	15016171
XM_004516808.3	1822	ATTTCAATCGAGCACAA	1806
XR_003335731.1	1679	TCAATCGAGCACAATAA	1663
XR_003335730.1	1679	TCAATCGAGCACAATAA	1663
XR_003335729.1	1679	TCAATCGAGCACAATAA	1663
XR_003335728.1	1679	TCAATCGAGCACAATAA	1663
XR_003335727.1	1679	TCAATCGAGCACAATAA	1663
XR_003335726.1	1679	TCAATCGAGCACAATAA	1663
XR_003335725.1	1679	TCAATCGAGCACAATAA	1663
XM_008140785.2	1155	ACATTTCAATCGAGCAC	1139
CP032299.1	10003448	TTTCAATCGAGCACCATAA	10003430

CP032299.1	22162784	ATTTCAATCGAGCACAA	22162768
CP032299.1	43636053	CAATCGAGCACAATAA	43636068
CP032299.1	57624346	AATCGAGCACAATAA	57624360
CP032232.1	10453175	TTTCAATCGAGCACCATAA	10453157
CP032232.1	22474271	ATTTCAATCGAGCACAA	22474255
XM_025214551.1	2673	CATTTCAATCGAGCACA	2657
XM_025214550.1	2678	CATTTCAATCGAGCACA	2662
XM_025214549.1	2901	CATTTCAATCGAGCACA	2885
XM_025214548.1	2746	CATTTCAATCGAGCACA	2730
XM_025214547.1	2785	CATTTCAATCGAGCACA	2769
XM_025214546.1	2748	CATTTCAATCGAGCACA	2732
XM_025214545.1	2779	CATTTCAATCGAGCACA	2763
XM_025214544.1	2769	CATTTCAATCGAGCACA	2753
XM_025214543.1	3008	CATTTCAATCGAGCACA	2992
AC277723.1	162338	ATTTCAATCGAGCACAA	162322
CP020671.1	20151957	ACATTTCAATCGAGCAC	20151941
CP019094.1	6869168	TTCAATCGAGCACAATA	6869152
XM_019596250.1	130	ATTTCAATCGAGCACAA	114
XM_019487198.1	2871	CATTTCAATCGAGCACA	2855
XM_019487197.1	3103	CATTTCAATCGAGCACA	3087
XM_019487196.1	3028	CATTTCAATCGAGCACA	3012
XM_006260474.3	3092	CATTTCAATCGAGCACA	3076
XM_014378881.1	395	TTTCAATCGAGCACAAT	379
XM_013452927.1	225	TTTCAATCGAGCACAAT	209
XM_009826460.1	1377	ACATTTCAATCGAGCAC	1361
XM_007940675.1	1997	ACATTTCAATCGAGCAC	1981
XM_006835838.1	1997	ACATTTCAATCGAGCAC	1981
AC162031.12	28360	ATTTCAATCGAGCACAA	28344
XM_034868321.1	2593	CATTTCAATCGAGCACA	2609
XM_034868320.1	2608	CATTTCAATCGAGCACA	2624
LR792568.1	10557225	ACATTTTAATCGAGCACAA	10557207
LR792568.1	8265588	TTCAATCGAGCACAATA	8265604
XM_032513447.1	2612	CATTTCAATCGAGCACA	2628
LR738563.1	8015903	TTTCAATCGAGCACAAT	8015919
LR736850.1	25837080	TTTCAATCGAGCACAAT	25837096
CP032376.1	11155525	TTTCAATCGAGCACAAT	11155541
CP032373.1	11159719	TTTCAATCGAGCACAAT	11159735
CP032367.1	11143339	TTTCAATCGAGCACAAT	11143355
LR597437.1	20047958	CATTTCAATCGAGCACA	20047974
CP032252.1	33076510	ATTTCAATCGAGCACAA	33076526
AP019376.1	4612470	CATTTCAATCGAGCACA	4612486
CP025674.1	2139338	CATTTCAATCGAGCACA	2139354
XM_011477429.3	3178	ACATTTCAATCGAGCAC	3194
XM_011477428.3	3113	ACATTTCAATCGAGCAC	3129
XM_004070826.4	3187	ACATTTCAATCGAGCAC	3203
XM_023062592.1	1380	TTTCAATCGAGCACAAT	1396
XM_023062591.1	1530	TTTCAATCGAGCACAAT	1546
CP023120.1	9676200	ATTTCAATCGAGCACAA	9676216
XR_002168388.1	71	ACATTTCAATCGAGCAC	87
JX899680.1	727976	TTTCAATCGAGCACAAT	727992
CP055244.1	20094247	TCAATCGAGCACAATAA	20094263
LR884470.1	10579535	CATTTCAATCGAGCACA	10579551
LR812569.1	39184610	ACATTTCAACCGAGCACAATA	39184630
LR812544.1	36644576	ACATTTCAACCGAGCACAATA	36644596
AC073347.3	193276	ATTTCAATCGAGCACAA	193292
XM_028886192.2	2032	CATTTCAATCGAGCAC	2017
XM_037200273.1	2036	CATTTCAATCGAGCAC	2021
XM_037233253.1	3654	TTTCAATCGAGCACAA	3639
XM_037233252.1	3417	TTTCAATCGAGCACAA	3402
XM_037233251.1	3420	TTTCAATCGAGCACAA	3405

XM_037232415.1	3412	TTTCAATCGAGCACAA	3397
XM_037232414.1	3415	TTTCAATCGAGCACAA	3400
CP062047.1	43036499	TTCAATCGAGCACAAT	43036484
CP062047.1	44290346	TTCAATCGAGCACAAT	44290331
CP062047.1	44483553	TTCAATCGAGCACAAT	44483538
CP055245.1	522194	CAATCGAGCACAATAA	522179
CP055245.1	33882478	ATTTCAATCGAGCAC	33882492
LR884475.1	6417364	ACATTTCAATCGAGCA	6417349
XM_028194100.2	1792	CATTTCAATCGATCACAATA	1773
XM_037532010.1	3252	ATTTCAATCGAGCACA	3267
CP062010.1	15890455	TCAATCGAGCACAATA	15890470
CP062010.1	32773415	TTTCAATCGAGCACA	32773401
CP062040.1	26550593	ATTTCAATCGTGCACAATAA	26550612
CP055236.1	30893849	TTTCAATCGAGCACAA	30893864
CP055235.1	28026296	ATTTCAATCGAGCACA	28026311
CP055235.1	18352690	ATTTCAATCGAGCAC	18352704
CP053152.1	11536670	TCAATCGAGCACAATA	11536685
XM_036036937.1	1933	ATTTCAATCGAGCACA	1948

PRODUCTS OF SIGNIFICANT ALIGNMENTS (CDS FEATURE)

- **Ribose 5-phosphate isomerase B** GenBank; GenBank
- **Putative cys-loop ligand gated ion channel subunit** GenBank
- **Uncharacterized protein** GenBank; GenBank; GenBank
- **Acyl-CoA dehydrogenase** GenBank
- **CYFA0S28e00122g1_1** GenBank
- **14481 bp at 5' side: putative cytoplasmic polyadenylation element-binding prot...ein 2-like**
- **5167 bp at 3' side: putative protein NDNF**
- **Pentatricopeptide repeat-containing protein At1g18485** GenBank
- **Pre-mRNA cleavage factor Im 25 kDa subunit 2-like** GenBank
- **Hypothetical protein** GenBank
- **Nucleolar RNA helicase 2** GenBank; GenBank; GenBank; GenBank
- **Isocitrate lyase** GenBank
- **Dynein heavy chain 1, axonemal** GenBank
- **Ankyrin repeat and LEM domain-containing protein** GenBank

SEQUENCE N° 6/1 (T=2; C=3; G=0; A=1)

Query: **ACATTTCAATCGAGCACAATAA**

Search Parameters	
Program	blastn
Word size	7
Expect value	1000
Hitlist size	250
Match/Mismatch scores	1,-3
Gapcosts	5,2
Filter string	F
Genetic Code	1

250 sequences selected:

Accession Seq	Description	Comomm Name	Max Score	Total Score	Query C.	E value	Per. Id.	Acc. Len
.....	.....	.....	.....	.....	.....	.....	.....	.....
LR862361.1 1/2 2/2 3/2 3/3 3/5	***Pipistrellus pipistrellus genome assembly, chromosome: 5***	common pipistrelle	32.2	32.2	72%	138	100.00%	94929986
.....	.....	.....	.....	.....	.....	.....	.....	.....

SEQUENCE N° 6/2

(T=2; C=3; G=0; A=1)

SEQUENCE N° 6/2 (T=2; C=3; G=0; A=1)

Original Sequence	G	T	G	A	A	A	T	G	G	T	C	A	T	G	T	G	T	G	G	C	G	G
Trend n° 6 (T=2; C=3; G=0; A=1)	0	2	0	1	1	1	2	0	0	2	3	1	2	0	2	0	2	0	0	3	0	0
Trend n° 6 (I i D d)	-	I	D	I	i	i	I	D	d	I	I	D	I	D	I	D	I	D	d	I	D	d
Trend n° 6/2	0	3	1	2	2	2	3	0	0	1	2	0	3	1	3	1	3	1	1	2	1	1
Sequence n° 6/2 (86,36%)	G	C	A	T	T	T	C	G	G	A	T	G	C	A	C	A	C	A	A	T	A	A

Query: **GCATTTCGGATGCACACAATAA**

Search Parameters	
Program	blastn
Word size	7
Expect value	1000
Hitlist size	100
Match/Mismatch scores	1,-3
Gapcosts	5,2
Filter string	F
Genetic Code	1

100 sequences selected:

Accession Seq	Description	Comomm Name	Max Score	Total Score	Query C.	E value	Per. Id.	Acc. Len
LR884471.1	Acanthopagrus latus isolate v.2019 genome assembly, chromosome: 12	yellowfin seabream	38.2	38.2	86%	2.3	100.00%	26679254
LR812078.1	Danio rerio genome assembly, chromosome: 16	zebrafish	36.2	36.2	81%	8.9	100.00%	55350661
LR792548.1	Pseudochaenichthys georgianus genome assembly, chromosome: 3	South Georgia icefish	36.2	36.2	100%	8.9	95.45%	53425585
XM_031276121.1	**PREDICTED: Ipomoea triloba FAD synthetase 1, chloroplastic-like (LOC116033344), partial mRNA**	trilobed morning glory	36.2	36.2	81%	8.9	100.00%	1153
XM_030786843.1	**PREDICTED: Chanos chanos V-type proton ATPase subunit B, brain isoform-like (LOC115822892), mRNA**	milkfish	36.2	36.2	81%	8.9	100.00%	2465
LR697115.1	**Chanos chanos genome assembly, chromosome: 10**	milkfish	36.2	36.2	81%	8.9	100.00%	40347274
CP025660.1	**Ipomoea triloba cultivar NCNSP0323 chromosome 1**	trilobed morning glory	36.2	36.2	81%	8.9	100.00%	38093535
CP032595.1	Lateolabrax maculatus linkage group 2 sequence	spotted sea bass	36.2	36.2	81%	8.9	100.00%	29991467
CP027278.1	Lateolabrax maculatus chromosome Lm17	spotted sea bass	36.2	36.2	81%	8.9	100.00%	22884195
XM_019320759.1	PREDICTED: Ipomoea nil FAD synthetase 2, chloroplastic-like (LOC109171689), transcript variant X3, mRNA	Japanese morning glory	36.2	36.2	81%	8.9	100.00%	1421
XM_019320758.1	PREDICTED: Ipomoea nil FAD synthetase 2, chloroplastic-like (LOC109171689), transcript variant X2, mRNA	Japanese morning glory	36.2	36.2	81%	8.9	100.00%	1401
XM_019320757.1	PREDICTED: Ipomoea nil FAD synthetase 2, chloroplastic-like (LOC109171689), transcript variant X1, mRNA	Japanese morning glory	36.2	36.2	81%	8.9	100.00%	1405
XM_013256488.1	PREDICTED: Stomoxys calcitrans homocysteine S-methyltransferase-like (LOC106090333), mRNA	stable fly	36.2	36.2	81%	8.9	100.00%	1230
CP025644.1	**Ipomoea trifida cultivar NCNSP0306 chromosome 1**	Ipomoea trifida	36.2	36.2	81%	8.9	100.00%	32247286
LT594792.1	Theobroma cacao genome assembly, chromosome: V	cacao	36.2	66.4	90%	8.9	100.00%	39371879
JN949980.1	Mus musculus targeted non-conditional, lacZ-tagged mutant allele Ggt6:tm1e(KOMP)Wtsi; transgenic	house mouse	36.2	36.2	81%	8.9	100.00%	39835

JN949979.1	Mus musculus targeted KO-first, conditional ready, lacZ-tagged mutant allele Ggt6:tm1a(KOMP)Wtsi; transgenic	house mouse	36.2	36.2	81%	8.9	100.00%	39885
JN949026.1	Mus musculus targeted non-conditional, lacZ-tagged mutant allele Mybbp1a:tm1e(EUCOMM)Hmgu; transgenic	house mouse	36.2	36.2	81%	8.9	100.00%	38333
JN949025.1	Mus musculus targeted KO-first, conditional ready, lacZ-tagged mutant allele Mybbp1a:tm1a(EUCOMM)Hmgu; transgenic	house mouse	36.2	36.2	81%	8.9	100.00%	38404
AF345929.1	Mus musculus Myb-binding protein p160 1a (Mybbp1a) gene, promoter region and partial cds	house mouse	36.2	36.2	81%	8.9	100.00%	549
AL662812.12	Mouse DNA sequence from clone RP23-48A2 on chromosome 11, complete sequence	house mouse	36.2	36.2	81%	8.9	100.00%	258521
LR792833.1	Digitaria exilis annotation	Digitaria exilis	34.2	34.2	77%	35	100.00%	29094342
LR761618.1	Digitaria exilis genome assembly, chromosome: 7B	Digitaria exilis	34.2	34.2	77%	35	100.00%	29094342
XM_017289635.2	PREDICTED: Drosophila miranda zinc finger protein 622 (LOC108157528), mRNA	Drosophila miranda	34.2	34.2	77%	35	100.00%	1577
XM_001355143.4	PREDICTED: Drosophila pseudoobscura zinc finger protein 622 (LOC4815046), mRNA	Drosophila pseudoobscura	34.2	34.2	77%	35	100.00%	1573
XR_004425352.1	**PREDICTED: Rhinolophus ferrumequinum serine/arginine repetitive matrix protein 1-like (LOC117036333), transcript variant X5, misc_RNA**	greater horseshoe bat	34.2	34.2	77%	35	100.00%	2643
XM_033131157.1	**PREDICTED: Rhinolophus ferrumequinum serine/arginine repetitive matrix protein 1-like (LOC117036333), transcript variant X4, mRNA**	greater horseshoe bat	34.2	34.2	77%	35	100.00%	2775
XM_033131156.1	**PREDICTED: Rhinolophus ferrumequinum serine/arginine repetitive matrix protein 1-like (LOC117036333), transcript variant X3, mRNA**	greater horseshoe bat	34.2	34.2	77%	35	100.00%	2844
XM_033131154.1	**PREDICTED: Rhinolophus ferrumequinum serine/arginine repetitive matrix protein 1-like (LOC117036333), transcript variant X1, mRNA**	greater horseshoe bat	34.2	34.2	77%	35	100.00%	2758
LR778290.1	Coregonus sp. 'balchen' genome assembly, chromosome: 38	Coregonus sp. 'balchen'	34.2	132	77%	35	100.00%	36774138
LR778289.1	Coregonus sp. 'balchen' genome assembly, chromosome: 37	Coregonus sp. 'balchen'	34.2	98.6	100%	35	100.00%	43663377
LR584434.1	**Salmo trutta genome assembly, chromosome: 34**	river trout	34.2	34.2	77%	35	100.00%	42900288
LR584426.1	**Salmo trutta genome assembly, chromosome: 17**	river trout	34.2	68.4	95%	35	95.24%	59764774
LR584421.1	**Salmo trutta genome assembly, chromosome: 23**	river trout	34.2	34.2	95%	35	95.24%	51488243
LR584414.1	**Salmo trutta genome assembly, chromosome: 33**	river trout	34.2	34.2	77%	35	100.00%	44893998
LR584412.1	**Salmo trutta genome assembly, chromosome: 24**	river trout	34.2	64.4	95%	35	100.00%	50333583
XM_002022755.2	PREDICTED: Drosophila persimilis zinc finger protein 622 (LOC6597574), mRNA	Drosophila persimilis	34.2	34.2	77%	35	100.00%	1532
CP032234.1	Anopheles stephensi strain SDA-500 chromosome 3R	Asian malaria mosquito	34.2	64.4	77%	35	100.00%	51436555
CP026254.1	Scophthalmus maximus chromosome 12	turbot	34.2	34.2	95%	35	95.24%	24811384
XM_002071777.3	PREDICTED: Drosophila willistoni chromosome-associated kinesin KIF4A (LOC6648886), mRNA	Drosophila willistoni	34.2	34.2	77%	35	100.00%	4044
CP023332.1	Drosophila melanogaster strain sitter (fors) chromosome 2R	fruit fly	34.2	34.2	77%	35	100.00%	21146707
AE013599.5	Drosophila melanogaster chromosome 2R	fruit fly	34.2	34.2	77%	35	100.00%	25286936
XM_007136828.1	Phaseolus vulgaris hypothetical protein (PHAVU_009G082600g) mRNA, complete cds	Phaseolus vulgaris	34.2	34.2	77%	35	100.00%	5319

OA888277.1	Notodromas monacha	Notodromas monacha	34.2	34.2	77%	35	100.00%	13315
CP002563.1	Carnobacterium sp. 17-4, complete genome	Carnobacterium sp. 17-4	34.2	34.2	77%	35	100.00%	2635294
AC146941.8	Medicago truncatula clone mth2-133o10, complete sequence	barrel medic	34.2	34.2	77%	35	100.00%	134901
LR812071.1	Danio rerio genome assembly, chromosome: 9	zebrafish	34.2	64.4	77%	35	100.00%	56856035
XM_034397081.1	PREDICTED: Thrips palmi uncharacterized LOC117652271 (LOC117652271), transcript variant X4, mRNA	Thrips palmi	34.2	34.2	77%	35	100.00%	910
XM_034397072.1	PREDICTED: Thrips palmi uncharacterized LOC117652271 (LOC117652271), transcript variant X3, mRNA	Thrips palmi	34.2	34.2	77%	35	100.00%	910
XM_034397054.1	PREDICTED: Thrips palmi uncharacterized LOC117652271 (LOC117652271), transcript variant X1, mRNA	Thrips palmi	34.2	34.2	77%	35	100.00%	931
LR778269.1	Coregonus sp. 'balchen' genome assembly, chromosome: 17	Coregonus sp. 'balchen'	34.2	124	95%	35	100.00%	54216998
LR778267.1	Coregonus sp. 'balchen' genome assembly, chromosome: 15	Coregonus sp. 'balchen'	34.2	94.6	95%	35	95.24%	55641933
LR778260.1	Coregonus sp. 'balchen' genome assembly, chromosome: 8	Coregonus sp. 'balchen'	34.2	124	86%	35	100.00%	68138733
LR738627.1	Sciurus vulgaris genome assembly, chromosome: 16	Eurasian red squirrel	34.2	64.4	95%	35	100.00%	79696672
LR738607.1	Sciurus carolinensis genome assembly, chromosome: 17	gray squirrel	34.2	64.4	95%	35	100.00%	66045118
CP035804.1	Haemonchus contortus strain NZ_Hco_NP chromosome 4	barber pole worm	34.2	64.4	100%	35	100.00%	92053285
XM_029729683.1	**PREDICTED: Salmo trutta coiled-coil domain-containing protein 85C-A-like (LOC115172326), transcript variant X3, mRNA**	river trout	34.2	34.2	77%	35	100.00%	3497
XM_029729682.1	**PREDICTED: Salmo trutta coiled-coil domain-containing protein 85C-A-like (LOC115172326), transcript variant X2, mRNA**	river trout	34.2	34.2	77%	35	100.00%	3601
XM_029729681.1	**PREDICTED: Salmo trutta coiled-coil domain-containing protein 85C-A-like (LOC115172326), transcript variant X1, mRNA**	river trout	34.2	34.2	77%	35	100.00%	3604
LR584435.1	**Salmo trutta genome assembly, chromosome: 18**	river trout	34.2	94.6	95%	35	95.24%	59142494
LR584433.1	**Salmo trutta genome assembly, chromosome: 5**	river trout	34.2	34.2	95%	35	95.24%	67763595
LR584423.1	**Salmo trutta genome assembly, chromosome: 31**	river trout	34.2	34.2	77%	35	100.00%	45589842
LR131966.1	Parambassis ranga genome assembly, chromosome: 16	Indian glassy fish	34.2	34.2	77%	35	100.00%	22465611
CP025673.1	**Ipomoea triloba cultivar NCNSP0323 chromosome 14**	trilobed morning glory	34.2	94.6	100%	35	95.24%	23651743
LS997565.1	Haemonchus contortus, ISE/inbred ISE, WGS project CAVP01000000 data, chromosome: 4	barber pole worm	34.2	64.4	77%	35	100.00%	51826579
LT594791.1	Theobroma cacao genome assembly, chromosome: IV	cacao	34.2	34.2	77%	35	100.00%	31929104
LT594790.1	Theobroma cacao genome assembly, chromosome: III	cacao	34.2	34.2	77%	35	100.00%	36364294
XM_014194909.1	PREDICTED: Salmo salar coiled-coil domain-containing protein 85C-A-like (LOC106602325), transcript variant X3, mRNA	Atlantic salmon	34.2	34.2	77%	35	100.00%	5155
XM_014194899.1	PREDICTED: Salmo salar coiled-coil domain-containing protein 85C-A-like (LOC106602325), transcript variant X2, mRNA	Atlantic salmon	34.2	34.2	77%	35	100.00%	5260
XM_014194890.1	PREDICTED: Salmo salar coiled-coil domain-containing protein 85C-A-like (LOC106602325), transcript variant X1, mRNA	Atlantic salmon	34.2	34.2	77%	35	100.00%	5261
CP012024.1	Bacillus smithii strain DSM 4216, complete genome	Bacillus smithii	34.2	34.2	77%	35	100.00%	3368778
LK066269.1	Apteryx australis mantelli genome assembly AptMant0, scaffold scaffold1697	Apteryx mantelli mantelli	34.2	102	77%	35	100.00%	483239
LN013793.1	Spirometra erinaceieuropaei genome assembly S_erinaceieuropaei, scaffold SPER_contig0000183	Spirometra erinaceieuropaei	34.2	34.2	77%	35	100.00%	11008

LR862152.1	Ananas comosus var. bracteatus genome assembly, chromosome: 24	Ananas comosus var. bracteatus	34.2	34.2	77%	35	100.00%	21245074
AP023178.1	Paraburkholderia sp. PGU16 plasmid PPGU16_p3 DNA, complete sequence	Paraburkholderia sp. PGU16	34.2	34.2	77%	35	100.00%	58022
LR812046.1	Danio rerio genome assembly, chromosome: 9	zebrafish	34.2	34.2	77%	35	100.00%	57070137
HE601631.2	Schistosoma mansoni strain Puerto Rico genome assembly, chromosome: ZW	Schistosoma mansoni	34.2	96.6	86%	35	100.00%	88385488
LR812602.1	Danio rerio genome assembly, chromosome: 9	zebrafish	34.2	34.2	77%	35	100.00%	56263738
LR812577.1	Danio rerio strain Nadia (NA) genome assembly, chromosome: 9	zebrafish	34.2	64.4	77%	35	100.00%	57525603
LR812552.1	Danio rerio strain Cooch Behar (CB) genome assembly, chromosome: 9	zebrafish	34.2	64.4	77%	35	100.00%	54069560
LR812514.1	Danio aesculapii genome assembly, chromosome: 18	Danio aesculapii	34.2	34.2	77%	35	100.00%	56436977
AC136507.19	Medicago truncatula clone mth2-24o19, complete sequence	barrel medic	34.2	34.2	77%	35	100.00%	131025
CR391919.5	Zebrafish DNA sequence from clone DKEY-174K18 in linkage group 9, complete sequence	zebrafish	34.2	34.2	77%	35	100.00%	179764
AC006247.12	Drosophila melanogaster, chromosome 2R, region 49E6-49F8, BAC clone BACR48I10, complete sequence	fruit fly	34.2	34.2	77%	35	100.00%	163162
XM_038023528.1	PREDICTED: Glossina fuscipes uncharacterized LOC119631376 (LOC119631376), transcript variant X3, mRNA	Glossina fuscipes	32.2	32.2	72%	139	100.00%	2035
XM_038023527.1	PREDICTED: Glossina fuscipes uncharacterized LOC119631376 (LOC119631376), transcript variant X2, mRNA	Glossina fuscipes	32.2	32.2	72%	139	100.00%	2097
XM_038023526.1	PREDICTED: Glossina fuscipes uncharacterized LOC119631376 (LOC119631376), transcript variant X1, mRNA	Glossina fuscipes	32.2	32.2	72%	139	100.00%	2092
CP065149.1	**Bacillus paranthracis strain Gxun-30 chromosome, complete genome**	Bacillus paranthracis	32.2	32.2	72%	139	100.00%	5149464
CP065161.1	Rhodobacterales bacterium strain HKCCA1288 chromosome, complete genome	Rhodobacterales bacterium	32.2	32.2	72%	139	100.00%	2663031
OA565374.1	**1_Tdi_b3v08**	Timema douglasi	32.2	32.2	90%	139	95.00%	181011
XM_037692747.1	PREDICTED: Nematolebias whitei nuclear factor, erythroid 2-like 3 (nfe2l3), mRNA	Rio pearlfish	32.2	32.2	72%	139	100.00%	2051
XM_037667276.1	PREDICTED: Rhipicephalus sanguineus transmembrane protein 114-like (LOC119400263), mRNA	brown dog tick	32.2	32.2	72%	139	100.00%	5227
XM_037611556.1	PREDICTED: Triticum dicoccoides uncharacterized LOC119339539 (LOC119339539), mRNA	Triticum dicoccoides	32.2	32.2	72%	139	100.00%	373
XM_037610901.1	PREDICTED: Triticum dicoccoides uncharacterized LOC119338597 (LOC119338597), mRNA	Triticum dicoccoides	32.2	32.2	72%	139	100.00%	1260
XM_038048380.1	PREDICTED: Hermetia illucens uncharacterized LOC119647451 (LOC119647451), transcript variant X2, mRNA	Hermetia illucens	32.2	32.2	72%	139	100.00%	4697
XM_038048379.1	PREDICTED: Hermetia illucens uncharacterized LOC119647451 (LOC119647451), transcript variant X1, mRNA	Hermetia illucens	32.2	32.2	72%	139	100.00%	4137
XM_038017770.1	PREDICTED: Bombyx mori guanine nucleotide-binding protein subunit beta-5 (LOC119629989), mRNA	domestic silkworm	32.2	32.2	72%	139	100.00%	3489
OA884554.1	Notodromas monacha	Notodromas monacha	32.2	32.2	72%	139	100.00%	38594
LR899427.1	Corylus avellana genome assembly, chromosome: ca5	Corylus avellana	32.2	32.2	72%	139	100.00%	36653616
CP046680.1	Solanum tuberosum cultivar P8 chromosome 2	potato	30.2	60.5	68%	549	100.00%	48611948

ALIGNMENT VIEW FLAT QUERY-ANCHORED WITH LETTERS FOR IDENTITIES

Query range 1: 1 to 22

Query	1	GCATTTCGGATGCACACAATAA	22
LR884471.1	5872730	CATTTCGGATGCACACAAT	5872748
LR812078.1	6416688	CATTTCGGATGCACACAA	6416671
LR792548.1	27700847	GCATTTCGGAAGCACACAATAA	27700826
XM_031276121.1	386	ATTTCGGATGCACACAAT	369
XM_030786843.1	1748	CATTTCGGATGCACACAA	1731
LR697115.1	19324110	CATTTCGGATGCACACAA	19324093
CP025660.1	13048241	ATTTCGGATGCACACAAT	13048224
CP032595.1	11460181	CATTTCGGATGCACACAA	11460164
CP027278.1	9140189	CATTTCGGATGCACACAA	9140172
XM_019320759.1	469	ATTTCGGATGCACACAAT	452
XM_019320758.1	449	ATTTCGGATGCACACAAT	432
XM_019320757.1	453	ATTTCGGATGCACACAAT	436
XM_013256488.1	851	ATTTCGGATGCACACAAT	834
CP025644.1	8402783	ATTTCGGATGCACACAAT	8402800
LT594792.1	27997305	ATTTCGGATGCACACAAT	27997322
LT594792.1	7456216	GGATGCACACAATAA	7456202
JN949980.1	27127	TTCGGATGCACACAATAA	27144
JN949979.1	27177	TTCGGATGCACACAATAA	27194
JN949026.1	11815	TTCGGATGCACACAATAA	11832
JN949025.1	11815	TTCGGATGCACACAATAA	11832
AF345929.1	185	TTCGGATGCACACAATAA	202
AL662812.12	58729	TTCGGATGCACACAATAA	58746
LR792833.1	3481679	TTCGGATGCACACAATA	3481663
LR761618.1	3481679	TTCGGATGCACACAATA	3481663
XM_017289635.2	164	CATTTCGGATGCACACA	148
XM_001355143.4	170	CATTTCGGATGCACACA	154
XR_004425352.1	869	CATTTCGGATGCACACA	853
XM_033131157.1	886	CATTTCGGATGCACACA	870
XM_033131156.1	955	CATTTCGGATGCACACA	939
XM_033131154.1	869	CATTTCGGATGCACACA	853
LR778290.1	30972461	CATTTCGGATGCACACA	30972445
LR778290.1	31640047	CATTTCGGATGCACACA	31640031
LR778290.1	31724491	CATTTCGGATGCACACA	31724507
LR778290.1	25939133	CATTTCGGATGCACA	25939147
LR778289.1	37891224	GCATTTGGGATGCACACAATA	37891244
LR778289.1	38138938	CATTTCGGATGCACACA	38138922
LR778289.1	1178049	GGATGCACACAATAA	1178063
LR584434.1	4737613	CATTTCGGATGCACACA	4737597
LR584426.1	40716693	CATTTCGGATGCACACTATAA	40716673
LR584426.1	44568833	CATTTCGGATGCACACA	44568849
LR584421.1	6926956	CATTTGGGATGCACACAATAA	6926936
LR584414.1	8526326	CATTTCGGATGCACACA	8526310
LR584412.1	27486011	GCATTTCGGATGCACAC	27485995
LR584412.1	30270180	CGGATGCACACAATA	30270194
XM_002022755.2	129	CATTTCGGATGCACACA	113
CP032234.1	29284202	CATTTCGGATGCACACA	29284186
CP032234.1	30390915	TTTCGGATGCACACA	30390901
CP026254.1	7772701	CATTGCGGATGCACACAATAA	7772681
XM_002071777.3	408	TTTCGGATGCACACAAT	392
CP023332.1	8950321	TTTCGGATGCACACAAT	8950305
AE013599.5	13062817	TTTCGGATGCACACAAT	13062801
XM_007136828.1	1139	TTTCGGATGCACACAAT	1123

OA888277.1	4987	ATTTCGGATGCACACAA	4971
CP002563.1	1553457	TTCGGATGCACACAATA	1553441
AC146941.8	83883	CATTTCGGATGCACACA	83867
LR812071.1	20700303	TTTCGGATGCACACAAT	20700319
LR812071.1	7146821	TTCGGATGCACACAA	7146835
XM_034397081.1	827	CATTTCGGATGCACACA	843
XM_034397072.1	827	CATTTCGGATGCACACA	843
XM_034397054.1	848	CATTTCGGATGCACACA	864
LR778269.1	2268914	CATTTCGGATGCAGACAAT	2268896
LR778269.1	12469841	CATTTGGGATGCACACAAT	12469823
LR778269.1	40386131	CATTTCGGATGCACACA	40386147
LR778269.1	21275431	GGATGCACACAATAA	21275417
LR778267.1	2833714	CATTTGGGATGCACACAATAA	2833734
LR778267.1	42529115	GGATGCACACAATAA	42529101
LR778267.1	12460471	GGATGCACACAATAA	12460485
LR778260.1	16977017	CATTTGGGATGCACACAAT	16976999
LR778260.1	43845202	CATTTGGGATGCACACAAT	43845220
LR778260.1	18485245	CATTTCGGATGCACACA	18485261
LR778260.1	13433975	ATTTCGGATGCACAC	13433961
LR738627.1	9776345	CATTTCGGATGCACACA	9776361
LR738627.1	43382214	GGATGCACACAATAA	43382200
LR738607.1	12873655	CATTTCGGATGCACACA	12873671
LR738607.1	65672302	GGATGCACACAATAA	65672316
CP035804.1	68968159	GCATTTCGGATGCACAC	68968175
CP035804.1	74077465	GGATGCACACAATAA	74077451
XM_029729683.1	3338	CATTTCGGATGCACACA	3354
XM_029729682.1	3442	CATTTCGGATGCACACA	3458
XM_029729681.1	3445	CATTTCGGATGCACACA	3461
LR584435.1	4663289	CATTTAGGATGCACACAATAA	4663309
LR584435.1	56328727	CATTTGGGATGCACACAAT	56328709
LR584435.1	50726731	CATTTCGGATGCACA	50726717
LR584433.1	2698528	CATTTAGGATGCACACAATAA	2698548
LR584423.1	23636461	CATTTCGGATGCACACA	23636477
LR131966.1	1611448	TCGGATGCACACAATAA	1611464
CP025673.1	18984285	CATTTCGGATGCACAGAATAA	18984305
CP025673.1	18456469	GCATTTCGGATGCAC	18456483
CP025673.1	21343206	GCATTTCGGATGCAC	21343220
LS997565.1	39271780	GCATTTCGGATGCACAC	39271796
LS997565.1	16214664	CATTTCGGATGCACA	16214650
LT594791.1	27812647	TTTCGGATGCACACAAT	27812663
LT594790.1	5296331	TTCGGATGCACACAATA	5296347
XM_014194909.1	3809	CATTTCGGATGCACACA	3825
XM_014194899.1	3914	CATTTCGGATGCACACA	3930
XM_014194890.1	3915	CATTTCGGATGCACACA	3931
CP012024.1	588716	TTCGGATGCACACAATA	588732
LK066269.1	301368	TCGGATGCACACAATAA	301384
LK066269.1	313490	TCGGATGCACACAATAA	313506
LK066269.1	325800	TCGGATGCACACAATAA	325816
LN013793.1	6993	CATTTCGGATGCACACA	7009
LR862152.1	6253400	TTTCGGATGCACACAAT	6253416
AP023178.1	48147	GCATTTCGGATGCACAC	48163
LR812046.1	20731743	TTTCGGATGCACACAAT	20731759
HE601631.2	58795411	CATTTCGTATGCACACAAT	58795393
HE601631.2	55394958	TTTCGGATGCACACAAT	55394974
HE601631.2	17951684	CATTTCGGATGCACAC	17951669
LR812602.1	20725446	TTTCGGATGCACACAAT	20725462
LR812577.1	21923491	TTTCGGATGCACACAAT	21923507
LR812577.1	7565339	TTCGGATGCACACAA	7565353
LR812552.1	20433661	TTTCGGATGCACACAAT	20433677
LR812552.1	5522106	TTCGGATGCACACAA	5522120

LR812514.1	16160608	CATTTCGGATGCACACA	16160624
AC136507.19	74014	CATTTCGGATGCACACA	74030
CR391919.5	156569	TTTCGGATGCACACAAT	156585
AC006247.12	139248	TTTCGGATGCACACAAT	139264
XM_038023528.1	299	TTCGGATGCACACAAT	284
XM_038023527.1	304	TTCGGATGCACACAAT	289
XM_038023526.1	299	TTCGGATGCACACAAT	284
CP065149.1	2963910	TTCGGATGCACACAAT	2963895
CP065161.1	1067558	CATTTCGGATGCACAC	1067543
OA565374.1	4441	ATTTCGAATGCACACAATAA	4422
XM_037692747.1	307	CATTTCGGATGCACAC	292
XM_037667276.1	646	TTTCGGATGCACACAA	631
XM_037611556.1	314	TTTCGGATGCACACAA	299
XM_037610901.1	627	TCGGATGCACACAATA	612
XM_038048380.1	3624	TTCGGATGCACACAAT	3639
XM_038048379.1	3064	TTCGGATGCACACAAT	3079
XM_038017770.1	1589	TTTCGGATGCACACAA	1604
OA884554.1	13328	TTTCGGATGCACACAA	13343
LR899427.1	14480734	CATTTCGGATGCACAC	14480749
CP046680.1	41617088	GGATGCACACAATAA	41617102
CP046680.1	41620099	GGATGCACACAATAA	41620113

SEQUENCE N° 6/2 (T=2; C=3; G=0; A=1)

Query: **GCATTTCGGATGCACACAATAA**

Search Parameters	
Program	blastn
Word size	7
Expect value	1000
Hitlist size	250
Match/Mismatch scores	1,-3
Gapcosts	5,2
Filter string	F
Genetic Code	1

248 sequences selected:

Accession Seq	Description	Comomm Name	Max Score	Total Score	Query C.	E value	Per. Id.	Acc. Len
XM_030786843.1	*PREDICTED: Chanos chanos V-type proton ATPase subunit B, brain isoform-like (LOC115822892), mRNA*	milkfish	36.2	36.2	81%	8.9	100.00%	2465
LR697115.1	*Chanos chanos genome assembly, chromosome: 10*	milkfish	36.2	36.2	81%	8.9	100.00%	40347274
.....	.....	.....	.....	.....	.....	....	.....	.....
LR584434.1	*Salmo trutta genome assembly, chromosome: 34*	river trout	34.2	34.2	77%	35	100.00%	42900288
LR584426.1	*Salmo trutta genome assembly, chromosome: 17*	river trout	34.2	68.4	95%	35	95.24%	59764774
LR584421.1	*Salmo trutta genome assembly, chromosome: 23*	river trout	34.2	34.2	95%	35	95.24%	51488243
LR584414.1	*Salmo trutta genome assembly, chromosome: 33*	river trout	34.2	34.2	77%	35	100.00%	44893998
LR584412.1	*Salmo trutta genome assembly, chromosome: 24*	river trout	34.2	64.4	95%	35	100.00%	50333583
.....	.....	.....	.....	.....	.....	....	.....	.....
XM_029729683.1	*PREDICTED: Salmo trutta coiled-coil domain-containing protein 85C-A-like (LOC115172326), transcript variant X3, mRNA*	river trout	34.2	34.2	77%	35	100.00%	3497
XM_029729682.1	*PREDICTED: Salmo trutta coiled-coil domain-containing protein 85C-A-like (LOC115172326), transcript variant X2, mRNA*	river trout	34.2	34.2	77%	35	100.00%	3601
XM_029729681.1	*PREDICTED: Salmo trutta coiled-coil domain-containing protein 85C-A-like (LOC115172326), transcript variant X1, mRNA*	river trout	34.2	34.2	77%	35	100.00%	3604
LR584435.1	*Salmo trutta genome assembly, chromosome: 18*	river trout	34.2	94.6	95%	35	95.24%	59142494
LR584433.1	*Salmo trutta genome assembly, chromosome: 5*	river trout	34.2	34.2	95%	35	95.24%	67763595
LR584423.1	*Salmo trutta genome assembly, chromosome: 31*	river trout	34.2	34.2	77%	35	100.00%	45589842

.....	.....	.....	.....	.....	.....	....	.....	.....
XM_014194909.1	PREDICTED: Salmo salar coiled-coil domain-containing protein 85C-A-like (LOC106602325), transcript variant X3, mRNA	Atlantic salmon	34.2	34.2	77%	35	100.00%	5155
XM_014194899.1	PREDICTED: Salmo salar coiled-coil domain-containing protein 85C-A-like (LOC106602325), transcript variant X2, mRNA	Atlantic salmon	34.2	34.2	77%	35	100.00%	5260
XM_014194890.1	PREDICTED: Salmo salar coiled-coil domain-containing protein 85C-A-like (LOC106602325), transcript variant X1, mRNA	Atlantic salmon	34.2	34.2	77%	35	100.00%	5261
.....	.....	.....	.....	.....	.....	....	.....	.....
CP054816.1	Bacillus anthracis strain FDAARGOS_695 chromosome	Bacillus anthracis	32.2	32.2	72%	139	100.00%	5135792
CP054797.1	Bacillus anthracis strain FDAARGOS_703 chromosome	Bacillus anthracis	32.2	32.2	72%	139	100.00%	5261520
.....	.....	.....	.....	.....	.....	....	.....	.....
CP050971.1	Bacillus anthracis strain FDAARGOS_696 chromosome	Bacillus anthracis	32.2	32.2	72%	139	100.00%	5255974
CP050970.1	Bacillus anthracis strain FDAARGOS_694 chromosome	Bacillus anthracis	32.2	32.2	72%	139	100.00%	5251490
.....	.....	.....	.....	.....	.....	....	.....	.....
CP041981.1	Bacillus paranthracis strain NCCP 15910 chromosome, complete genome	Bacillus paranthracis	32.2	32.2	72%	139	100.00%	5208566
CP047131.1	Bacillus anthracis str. BF1 chromosome, complete genome	Bacillus anthracis str. BF1	32.2	32.2	72%	139	100.00%	5213532
.....	.....	.....	.....	.....	.....	....	.....	.....
CP047104.1	Bacillus anthracis strain FDAARGOS_706 chromosome, complete genome	Bacillus anthracis	32.2	32.2	72%	139	100.00%	5227454
CP047099.1	Bacillus anthracis strain FDAARGOS_700 chromosome, complete genome	Bacillus anthracis	32.2	32.2	72%	139	100.00%	5227419
CP047097.1	Bacillus anthracis strain FDAARGOS_698 chromosome, complete genome	Bacillus anthracis	32.2	32.2	72%	139	100.00%	5225746
.....	.....	.....	.....	.....	.....	....	.....	.....
CP065149.1	Bacillus paranthracis strain Gxun-30 chromosome, complete genome	Bacillus paranthracis	32.2	32.2	72%	139	100.00%	5149464
.....	.....	.....	.....	.....	.....	....	.....	.....
OA565374.1	1_Tdi_b3v08	Timema douglasi	32.2	32.2	90%	139	95.00%	181011
.....	.....	.....	.....	.....	.....	....	.....	.....
CP060194.1	Bacillus anthracis strain Kanchipuram chromosome	Bacillus anthracis	32.2	32.2	72%	139	100.00%	5227419
.....	.....	.....	.....	.....	.....	....	.....	.....

LR862368.1	***Pipistrellus pipistrellus genome assembly, chromosome: 12***	common pipistrelle	32.2	32.2	90%	139	95.00%	58291996
LR862364.1	***Pipistrellus pipistrellus genome assembly, chromosome: 8***	common pipistrelle	32.2	32.2	90%	139	95.00%	83522093
.....	.....	.....	.....	.....	.....	....	.....	.....
CP050973.1	***Bacillus anthracis strain FDAARGOS_699 chromosome***	Bacillus anthracis	32.2	32.2	72%	139	100.00%	5227345
CP050972.1	***Bacillus anthracis strain FDAARGOS_697 chromosome***	Bacillus anthracis	32.2	32.2	72%	139	100.00%	5240765
.....	.....	.....	.....	.....	.....	....	.....	.....
CP047111.1	***Bacillus anthracis strain FDAARGOS_704 chromosome, complete genome***	Bacillus anthracis	32.2	32.2	72%	139	100.00%	5227450
CP047107.1	***Bacillus anthracis strain FDAARGOS_705 chromosome, complete genome***	Bacillus anthracis	32.2	32.2	72%	139	100.00%	5227452
CP047098.1	***Bacillus anthracis strain FDAARGOS_701 chromosome, complete genome***	Bacillus anthracis	32.2	32.2	72%	139	100.00%	5227419
.....	.....	.....	.....	.....	.....	....	.....	.....
LR597569.1	***Myripristis murdjan genome assembly, chromosome: 20***	pinecone soldierfish	32.2	32.2	72%	139	100.00%	29743470
.....	.....	.....	.....	.....	.....	....	.....	.....

SEQUENCE N° 7/1

(T=2; C=1; G=0; A=3)

SEQUENCE N° 7/1 (T=2; C=1; G=0; A=3)

Original Sequence	**G**	**T**	**G**	**A**	**A**	**A**	**T**	**G**	**G**	**T**	**C**	**A**	**T**	**G**	**T**	**G**	**T**	**G**	**G**	**C**	**G**	**G**
Trend n° 7 (T=2; C=1; G=0; A=3)	**0**	**2**	**0**	**3**	**3**	**3**	**2**	**0**	**0**	**2**	**1**	**3**	**2**	**0**	**2**	**0**	**2**	**0**	**0**	**1**	**0**	**0**
Trend n° 7 (I i D d)	**-**	**I**	**D**	**I**	**i**	**i**	**D**	**D**	**d**	**I**	**D**	**I**	**D**	**D**	**I**	**D**	**I**	**D**	**d**	**I**	**D**	**d**
Trend n° 7/1	**1**	**3**	**1**	**3**	**3**	**3**	**1**	**0**	**0**	**1**	**0**	**2**	**1**	**0**	**3**	**1**	**3**	**1**	**1**	**2**	**1**	**1**
Sequence n° 7/1 (72,73%)	**C**	**A**	**C**	**A**	**A**	**A**	**C**	**G**	**G**	**C**	**G**	**T**	**C**	**G**	**A**	**C**	**A**	**C**	**C**	**T**	**C**	**C**

Query: **CACAAACGGCGTCGACACCTCC**

Search Parameters	
Program	blastn
Word size	7
Expect value	1000
Hitlist size	100
Match/Mismatch scores	1,-3
Gapcosts	5,2
Filter string	F
Genetic Code	1

100 sequences selected:

Accession Seq	Description	Comomm Name	Max Score	Total Score	Query C.	E value	Per. Id.	Acc. Len
MT310855.1	Microbacterium phage Rasovi, complete genome	Microbacterium phage Rasovi	40.1	40.1	90%	0.56	100.00%	42917
XM_013906883.1	Thecamonas trahens ATCC 50062 hypothetical protein partial mRNA	Thecamonas trahens ATCC 50062	40.1	72.4	90%	0.56	100.00%	49860
XM_033910218.1	Saccharomyces paradoxus Dug1 (SPAR_F01050), partial mRNA	Saccharomyces paradoxus	38.2	38.2	86%	2.2	100.00%	1446
CP015039.1	Rhodovulum sp. P5, complete genome	Rhodovulum sp. P5	38.2	38.2	86%	2.2	100.00%	4137334
CP020281.1	Saccharomyces paradoxus strain YPS138 chromosome VI, complete sequence	Saccharomyces paradoxus	38.2	38.2	86%	2.2	100.00%	296071
CP020264.1	Saccharomyces paradoxus strain N44 chromosome VI, complete sequence	Saccharomyces paradoxus	38.2	38.2	86%	2.2	100.00%	296174
CP020247.1	Saccharomyces paradoxus strain CBS432 chromosome VI, complete sequence	Saccharomyces paradoxus	38.2	38.2	86%	2.2	100.00%	296034
LN169883.1	Spirometra erinaceieuropaei genome assembly S_erinaceieuropaei, scaffold SPER_scaffold0150729	Spirometra erinaceieuropaei	38.2	38.2	86%	2.2	100.00%	2171
LR792826.1	Digitaria exilis annotation	Digitaria exilis	36.2	36.2	81%	8.7	100.00%	27476960
LR761611.1	Digitaria exilis genome assembly, chromosome: 4A	Digitaria exilis	36.2	36.2	81%	8.7	100.00%	27476960
AP018441.1	Undibacterium sp. YM2 DNA, complete genome	Undibacterium sp. YM2	36.2	180	81%	8.7	100.00%	6484812
AP018439.1	Undibacterium sp. KW1 DNA, complete genome	Undibacterium sp. KW1	36.2	108	81%	8.7	100.00%	6557365
CP023747.1	Streptomyces nodosus strain ATCC 14899 chromosome, complete genome	Streptomyces nodosus	36.2	98.6	100%	8.7	95.45%	7772587
CP023697.1	Streptomyces prasinus strain ATCC 13879 chromosome, complete genome	Streptomyces prasinus	36.2	36.2	81%	8.7	100.00%	7647592
CP036291.1	Planctomycetes bacterium Pla175 chromosome, complete genome	Planctomycetes bacterium Pla175	36.2	66.4	81%	8.7	100.00%	6618662
XM_030095728.1	PREDICTED: Salarias fasciatus septin-7-like (LOC115391457), transcript variant X4, mRNA	jewelled blenny	36.2	36.2	81%	8.7	100.00%	4270
XM_030095727.1	PREDICTED: Salarias fasciatus septin-7-like (LOC115391457), transcript variant X3, mRNA	jewelled blenny	36.2	36.2	81%	8.7	100.00%	4273

XM_030095726.1	PREDICTED: Salarias fasciatus septin-7-like (LOC115391457), transcript variant X2, mRNA	jewelled blenny	36.2	36.2	81%	8.7	100.00%	4282
XM_030095725.1	PREDICTED: Salarias fasciatus septin-7-like (LOC115391457), transcript variant X1, mRNA	jewelled blenny	36.2	36.2	81%	8.7	100.00%	4285
XM_030095379.1	PREDICTED: Salarias fasciatus septin-7-like (LOC115391256), transcript variant X5, mRNA	jewelled blenny	36.2	36.2	81%	8.7	100.00%	1397
XM_030095378.1	PREDICTED: Salarias fasciatus septin-7-like (LOC115391256), transcript variant X4, mRNA	jewelled blenny	36.2	36.2	81%	8.7	100.00%	1653
XM_030095377.1	PREDICTED: Salarias fasciatus septin-7-like (LOC115391256), transcript variant X3, mRNA	jewelled blenny	36.2	36.2	81%	8.7	100.00%	1656
XM_030095376.1	PREDICTED: Salarias fasciatus septin-7-like (LOC115391256), transcript variant X2, mRNA	jewelled blenny	36.2	36.2	81%	8.7	100.00%	1666
XM_030095375.1	PREDICTED: Salarias fasciatus septin-7-like (LOC115391256), transcript variant X1, mRNA	jewelled blenny	36.2	36.2	81%	8.7	100.00%	1669
LR597442.1	Salarias fasciatus genome assembly, chromosome: 7	jewelled blenny	36.2	72.4	81%	8.7	100.00%	33379828
CP032382.1	Chryseolinea soli strain KIS68-18, complete genome	Chryseolinea soli	36.2	66.4	81%	8.7	100.00%	8241389
CP031455.1	Streptomyces olivoreticuli subsp. olivoreticuli strain ATCC 31159 chromosome, complete genome	Streptomyces olivoreticuli subsp. olivoreticuli	36.2	66.4	81%	8.7	100.00%	8809793
CP025803.1	Sulfitobacter sp. SK011 chromosome, complete genome	Sulfitobacter sp. SK011	36.2	36.2	81%	8.7	100.00%	4190786
CP030989.1	Arachis hypogaea cultivar Shitouqi chromosome A07	peanut	36.2	36.2	81%	8.7	100.00%	83117376
CP017316.1	Streptomyces rubrolavendulae strain MJM4426, complete genome	Streptomyces rubrolavendulae	36.2	159	81%	8.7	100.00%	6543262
CP009313.1	Streptomyces nodosus strain ATCC 14899 genome	Streptomyces nodosus	36.2	98.6	100%	8.7	95.45%	7714110
CP005286.1	Corynebacterium humireducens NBRC 106098 = DSM 45392, complete genome	Corynebacterium humireducens NBRC 106098 = DSM 45392	36.2	66.4	81%	8.7	100.00%	2681312
XM_008910664.1	Phytophthora parasitica INRA-310 hypothetical protein mRNA	Phytophthora parasitica INRA-310	36.2	36.2	81%	8.7	100.00%	6776
XM_003660782.1	Thermothelomyces thermophilus ATCC 42464 uncharacterized protein (MYCTH_97448), partial mRNA	Thermothelomyces thermophilus ATCC 42464	36.2	36.2	81%	8.7	100.00%	1335
XM_036686787.1	Fusarium subglutinans uncharacterized protein (FSUBG_8384), partial mRNA	Fusarium subglutinans	36.2	36.2	100%	8.7	95.45%	1008
CP048289.1	Streptacidiphilus sp. P02-A3a chromosome, complete genome	Streptacidiphilus sp. P02-A3a	36.2	161	81%	8.7	100.00%	9330052
CP058334.1	Natronomonas sp. C90 chromosome, complete genome	Natronomonas sp. C90	36.2	126	81%	8.7	100.00%	3211682
CP003002.1	Myceliophthora thermophila ATCC 42464 chromosome 1, complete sequence	Thermothelomyces thermophilus ATCC 42464	36.2	36.2	81%	8.7	100.00%	10931058
XM_002998468.1	Phytophthora infestans T30-4 conserved hypothetical protein (PITG_06383) mRNA, complete cds	Phytophthora infestans T30-4	36.2	36.2	81%	8.7	100.00%	2769
CP002049.1	Truepera radiovictrix DSM 17093, complete genome	Truepera radiovictrix DSM 17093	36.2	36.2	81%	8.7	100.00%	3260398
CP053450.1	Planctomycetes bacterium isolate PNC22 chromosome	Planctomycetes bacterium	34.2	34.2	95%	35	95.24%	4124050
CP054038.1	Microbacterium hominis strain PA2F3 chromosome, complete genome	Microbacterium hominis	34.2	64.4	77%	35	100.00%	3908277
AP019869.1	Marchantia polymorpha subsp. ruderalis Tak-1 DNA, chromosome: 4	Marchantia polymorpha subsp. ruderalis	34.2	34.2	77%	35	100.00%	26988051
AP019867.1	Marchantia polymorpha subsp. ruderalis Tak-1 DNA, chromosome: 2	Marchantia polymorpha subsp. ruderalis	34.2	34.2	77%	35	100.00%	29643427
CP053625.1	Apiotrichum mycotoxinovorans strain CICC 1454 chromosome 6	Apiotrichum mycotoxinivorans	34.2	34.2	77%	35	100.00%	1638155
CP053621.1	Apiotrichum mycotoxinovorans strain CICC 1454 chromosome	Apiotrichum mycotoxinivorans	34.2	34.2	77%	35	100.00%	10254415

	2							
CP053032.1	Halomonas sp. PGE1 chromosome, complete genome	Halomonas sp. PGE1	34.2	34.2	77%	35	100.00%	3905792
CP052860.1 1/2 5/1 7/2 9/1	Pseudomonas sp. ADAK13 chromosome, complete genome	Pseudomonas sp. ADAK13	34.2	64.4	95%	35	100.00%	7300098
XM_033681904.1	Eremomyces bilateralis CBS 781.70 DEAD/DEAH box helicase (P152DRAFT_482706), mRNA	Eremomyces bilateralis CBS 781.70	34.2	34.2	77%	35	100.00%	2540
CP051181.1	Thalassobius gelatinovorus strain NEB572 chromosome, complete genome	Thalassobius gelatinovorus	34.2	34.2	77%	35	100.00%	3908020
CP051010.1	Streptomyces sp. S1D4-11 chromosome, complete genome	Streptomyces sp. S1D4-11	34.2	157	81%	35	100.00%	12276515
CP050974.1	Streptomyces sp. RLB1-33 chromosome, complete genome	Streptomyces sp. RLB1-33	34.2	157	81%	35	100.00%	12127650
CP050975.1	Streptomyces sp. RPA4-2 chromosome, complete genome	Streptomyces sp. RPA4-2	34.2	128	81%	35	100.00%	9856149
CP046172.1	Nocardia arthritidis strain AUSMDU00012717 chromosome, complete genome	Nocardia arthritidis	34.2	34.2	77%	35	100.00%	10014015
XM_032945539.1	PREDICTED: Petromyzon marinus sestrin-1-like (LOC116938425), mRNA	sea lamprey	34.2	34.2	77%	35	100.00%	1755
CP049825.1	Apiotrichum mycotoxinovorans strain GMU1709 chromosome V	Apiotrichum mycotoxinivorans	34.2	34.2	77%	35	100.00%	1646260
CP049824.1	Apiotrichum mycotoxinovorans strain GMU1709 chromosome IV	Apiotrichum mycotoxinivorans	34.2	34.2	77%	35	100.00%	10238178
XM_032809184.1	PREDICTED: Odontomachus brunneus alpha-mannosidase 2 (LOC116841363), transcript variant X2, mRNA	Odontomachus brunneus	34.2	34.2	77%	35	100.00%	6205
XM_032809183.1	PREDICTED: Odontomachus brunneus alpha-mannosidase 2 (LOC116841363), transcript variant X1, mRNA	Odontomachus brunneus	34.2	34.2	77%	35	100.00%	6209
CP049199.1	Caulobacter sp. Ji-3-8 chromosome	Caulobacter sp. Ji-3-8	34.2	64.4	77%	35	100.00%	5256397
CP041650.2	Streptomyces sp. RLB1-8 chromosome, complete genome	Streptomyces sp. RLB1-8	34.2	185	81%	35	100.00%	11765395
CP048397.1	Streptomyces sp. S4.7 chromosome, complete genome	Streptomyces sp. S4.7	34.2	96.6	81%	35	100.00%	7920066
CP041612.2	Streptomyces sp. S1A1-8 chromosome, complete genome	Streptomyces sp. S1A1-8	34.2	185	81%	35	100.00%	12037094
CP041607.2	Streptomyces sp. S1D4-14 chromosome, complete genome	Streptomyces sp. S1D4-14	34.2	185	81%	35	100.00%	11723552
CP041609.2	Streptomyces sp. S1D4-20 chromosome	Streptomyces sp. S1D4-20	34.2	185	81%	35	100.00%	11851273
CP041613.2	Streptomyces sp. S1D4-23 chromosome	Streptomyces sp. S1D4-23	34.2	215	81%	35	100.00%	12057750
CP041604.2	Streptomyces sp. S1A1-7 chromosome	Streptomyces sp. S1A1-7	34.2	245	81%	35	100.00%	11713216
CP041602.2	Streptomyces sp. RLB3-6 chromosome	Streptomyces sp. RLB3-6	34.2	215	81%	35	100.00%	12338170
CP041610.2	Streptomyces sp. RLB3-17 chromosome, complete genome	Streptomyces sp. RLB3-17	34.2	185	81%	35	100.00%	12022941
AP022600.1	Mycolicibacterium tokaiense JCM 6373 DNA, nearly complete genome	Mycolicibacterium tokaiense	34.2	98.6	77%	35	100.00%	6328149
AP022561.1	Mycolicibacterium aichiense JCM 6376 DNA, complete genome	Mycolicibacterium aichiense	34.2	94.6	81%	35	100.00%	5925482
CP047385.1	Pandoraea fibrosis strain 6399 chromosome, complete genome	Pandoraea fibrosis	34.2	34.2	77%	35	100.00%	5592065
CP047386.1	Pandoraea fibrosis strain 7641 chromosome, complete genome	Pandoraea fibrosis	34.2	34.2	77%	35	100.00%	5592064

CP047183.1	Rathayibacter sp. VKM Ac-2801 chromosome, complete genome	Rathayibacter sp. VKM Ac-2801	34.2	64.4	81%	35	100.00%	3577774
CP043930.1	Gimesia benthica strain E7 chromosome, complete genome	Gimesia sp. E7	34.2	34.2	77%	35	100.00%	8027492
XM_031649147.1	PREDICTED: Nymphaea colorata 30S ribosomal protein S5, chloroplastic (LOC116267408), mRNA	Nymphaea colorata	34.2	34.2	77%	35	100.00%	1361
CP045643.1	Streptomyces sp. QMT-28 chromosome, complete genome	Streptomyces sp. QMT-28	34.2	96.6	81%	35	100.00%	8880330
XM_031503155.1	PREDICTED: Photinus pyralis neurogenic locus Notch protein-like (LOC116182610), mRNA	common eastern firefly	34.2	34.2	77%	35	100.00%	6755
XM_031478322.1	PREDICTED: Photinus pyralis neurogenic locus Notch protein-like (LOC116164181), mRNA	common eastern firefly	34.2	34.2	77%	35	100.00%	8468
CP023690.1	Streptomyces spectabilis strain ATCC 27465 chromosome, complete genome	Streptomyces spectabilis	34.2	64.4	81%	35	100.00%	9807160
CP029194.1	Streptomyces venezuelae strain ATCC 15068 chromosome, complete genome	Streptomyces venezuelae	34.2	66.4	77%	35	100.00%	8558202
CP043534.1	[Candida] auris strain B11220 chromosome 4	[Candida] auris	34.2	34.2	77%	35	100.00%	1318327
CP043448.1	[Candida] auris strain B11245 chromosome 7	[Candida] auris	34.2	34.2	77%	35	100.00%	781361
XM_030681488.1	PREDICTED: Rhodamnia argentea macrophage migration inhibitory factor homolog (LOC115745871), transcript variant X2, mRNA	Rhodamnia argentea	34.2	34.2	77%	35	100.00%	720
XM_030681487.1	PREDICTED: Rhodamnia argentea macrophage migration inhibitory factor homolog (LOC115745871), transcript variant X1, mRNA	Rhodamnia argentea	34.2	34.2	77%	35	100.00%	779
XM_030588374.1	PREDICTED: Syzygium oleosum macrophage migration inhibitory factor homolog (LOC115666694), transcript variant X3, mRNA	Syzygium oleosum	34.2	34.2	77%	35	100.00%	746
XM_030588373.1	PREDICTED: Syzygium oleosum macrophage migration inhibitory factor homolog (LOC115666694), transcript variant X2, mRNA	Syzygium oleosum	34.2	34.2	77%	35	100.00%	829
XM_030588372.1	PREDICTED: Syzygium oleosum macrophage migration inhibitory factor homolog (LOC115666694), transcript variant X1, mRNA	Syzygium oleosum	34.2	34.2	77%	35	100.00%	805
CP032580.1	Gossypium turneri isolate D10-2 chromosome D10_10	Gossypium turneri	34.2	34.2	77%	35	100.00%	63881757
CP032571.1	Gossypium turneri isolate D10-2 chromosome D10_01	Gossypium turneri	34.2	34.2	77%	35	100.00%	60445394
XM_037413552.1	PREDICTED: Rhipicephalus microplus uncharacterized LOC119161190 (LOC119161190), mRNA	southern cattle tick	34.2	34.2	77%	35	100.00%	6047
CP063451.1	Rhodococcus pyridinivorans strain 5Ap plasmid pNAPH, complete sequence	Rhodococcus pyridinivorans	34.2	34.2	77%	35	100.00%	207815
AP024094.1	Jeongeupia sp. HS-3 DNA, complete genome	Jeongeupia sp. HS-3	34.2	64.4	81%	35	100.00%	3396565
CP061167.1	[Candida] auris strain CA-AM1 chromosome VII	[Candida] auris	34.2	34.2	77%	35	100.00%	779240
CP060496.1	Brachybacterium sp. Z12 chromosome, complete genome	Brachybacterium sp. Z12	34.2	34.2	77%	35	100.00%	3511914
CP048835.1	Streptomyces sp. So13.3 chromosome, complete genome	Streptomyces sp. So13.3	34.2	64.4	77%	35	100.00%	9564008
CP041141.1	[Candida] auris isolate A1 chromosome VII	[Candida] auris	34.2	34.2	77%	35	100.00%	748923
CP059696.1	Streptomyces pristinaespiralis strain PR11 chromosome, complete genome	Streptomyces pristinaespiralis	34.2	126	81%	35	100.00%	8112061
XM_035673227.1	PREDICTED: Morone saxatilis protein asteroid homolog 1-like	striped sea-bass	34.2	34.2	77%	35	100.00%	2394

	(LOC118336560), mRNA							
CP058693.1	Streptomyces sp. Rer75 chromosome	Streptomyces sp. Rer75	34.2	34.2	77%	35	100.00%	10474412

ALIGNMENT VIEW FLAT QUERY-ANCHORED WITH LETTERS FOR IDENTITIES

Query range 1: 1 to 22

Query	1	CACAAACGGCGTCGACACCTCC	22
MT310855.1	30662	CAAACGGCGTCGACACCTCC	30681
XM_013906883.1	48951	CACAAACGGCGTCGACACCT	48970
XM_013906883.1	30727	AACGGCGTCGACACCT	30742
XM_033910218.1	591	ACAAACGGCGTCGACACCT	573
CP015039.1	4066839	AAACGGCGTCGACACCTCC	4066857
CP020281.1	256166	ACAAACGGCGTCGACACCT	256184
CP020264.1	256268	ACAAACGGCGTCGACACCT	256286
CP020247.1	254431	ACAAACGGCGTCGACACCT	254449
LN169883.1	892	ACAAACGGCGTCGACACCT	874
LR792826.1	7762633	AACGGCGTCGACACCTCC	7762650
LR761611.1	7762633	AACGGCGTCGACACCTCC	7762650
AP018441.1	698901	CACAAACGGCGTCGACAC	698918
AP018441.1	3921576	CACAAACGGCGTCGACAC	3921559
AP018441.1	5973565	CACAAACGGCGTCGACAC	5973548
AP018441.1	6131915	CACAAACGGCGTCGACAC	6131932
AP018441.1	6468273	CACAAACGGCGTCGACAC	6468290
AP018439.1	561375	CACAAACGGCGTCGACAC	561358
AP018439.1	1342166	CACAAACGGCGTCGACAC	1342183
AP018439.1	3878339	CACAAACGGCGTCGACAC	3878322
CP023747.1	4327056	CACAAACGGCGGCGACACCTCC	4327077
CP023747.1	4069082	CGGCGTCGACACCTCC	4069097
CP023747.1	1717733	AACGGCGTCGACACC	1717747
CP023697.1	6464346	AACGGCGTCGACACCTCC	6464363
CP036291.1	3551990	CAAACGGCGTCGACACCT	3552007
CP036291.1	3520870	AACGGCGTCGACACC	3520884
XM_030095728.1	1177	AACGGCGTCGACACCTCC	1194
XM_030095727.1	1180	AACGGCGTCGACACCTCC	1197
XM_030095726.1	1189	AACGGCGTCGACACCTCC	1206
XM_030095725.1	1192	AACGGCGTCGACACCTCC	1209
XM_030095379.1	922	AACGGCGTCGACACCTCC	939
XM_030095378.1	1178	AACGGCGTCGACACCTCC	1195
XM_030095377.1	1181	AACGGCGTCGACACCTCC	1198
XM_030095376.1	1190	AACGGCGTCGACACCTCC	1207
XM_030095375.1	1193	AACGGCGTCGACACCTCC	1210
LR597442.1	24088064	AACGGCGTCGACACCTCC	24088081
LR597442.1	24476302	AACGGCGTCGACACCTCC	24476319
CP032382.1	7967126	CAAACGGCGTCGACACCT	7967143
CP032382.1	3980627	CAAACGGCGTCGACA	3980641
CP031455.1	4584638	AACGGCGTCGACACCTCC	4584621
CP031455.1	2710076	AACGGCGTCGACACC	2710090
CP025803.1	2236918	AAACGGCGTCGACACCTC	2236901
CP030989.1	67044159	AACGGCGTCGACACCTCC	67044176
CP017316.1	12679	AACGGCGTCGACACCTCC	12696
CP017316.1	2170361	CGGCGTCGACACCTCC	2170376
CP017316.1	241740	GGCGTCGACACCTCC	241754

CP017316.1	2341321	GGCGTCGACACCTCC	2341307
CP017316.1	3640586	GGCGTCGACACCTCC	3640572
CP009313.1	4252661	CACAAACGGCGGCGACACCTCC	4252682
CP009313.1	3995346	CGGCGTCGACACCTCC	3995361
CP009313.1	1646973	AACGGCGTCGACACC	1646987
CP005286.1	1169796	AACGGCGTCGACACCTCC	1169813
CP005286.1	2111537	AACGGCGTCGACACC	2111523
XM_008910664.1	1121	AACGGCGTCGACACCTCC	1138
XM_003660782.1	1057	AACGGCGTCGACACCTCC	1074
XM_036686787.1	798	CACAAACGACGTCGACACCTCC	819
CP048289.1	2823686	AACGGCGTCGACACCTCC	2823703
CP048289.1	5599762	AACGGCGTCGACACCT	5599747
CP048289.1	9091151	CGGCGTCGACACCTCC	9091136
CP048289.1	728680	GGCGTCGACACCTCC	728694
CP048289.1	2908602	AACGGCGTCGACACC	2908588
CP058334.1	918529	AACGGCGTCGACACCTCC	918512
CP058334.1	858919	AACGGCGTCGACACC	858933
CP058334.1	1177828	ACGGCGTCGACACCT	1177842
CP058334.1	1665270	CGGCGTCGACACCTC	1665284
CP003002.1	9999161	AACGGCGTCGACACCTCC	9999144
XM_002998468.1	1754	CACAAACGGCGTCGACAC	1737
CP002049.1	836152	AACGGCGTCGACACCTCC	836135
CP053450.1	3761022	CACAAACGGCGTCGCCACCTC	3761002
CP054038.1	2316437	ACGGCGTCGACACCTCC	2316421
CP054038.1	719854	ACGGCGTCGACACCT	719868
AP019869.1	17134560	ACGGCGTCGACACCTCC	17134576
AP019867.1	23100797	CACAAACGGCGTCGACA	23100781
CP053625.1	1429148	AACGGCGTCGACACCTC	1429132
CP053621.1	314245	ACGGCGTCGACACCTCC	314229
CP053032.1	3358899	ACGGCGTCGACACCTCC	3358915
CP052860.1	5581283	AACGGCGTCGACACCTC	5581267
CP052860.1	6399354	CACAAACGGCGTCGA	6399368
XM_033681904.1	950	AACGGCGTCGACACCTC	934
CP051181.1	1971308	CACAAACGGCGTCGACA	1971292
CP051010.1	7300134	ACGGCGTCGACACCTCC	7300150
CP051010.1	6353150	CGGCGTCGACACCTCC	6353165
CP051010.1	2899305	AACGGCGTCGACACC	2899319
CP051010.1	9622143	AACGGCGTCGACACC	9622129
CP051010.1	10787014	AACGGCGTCGACACC	10787028
CP050974.1	7139722	ACGGCGTCGACACCTCC	7139738
CP050974.1	6222402	CGGCGTCGACACCTCC	6222417
CP050974.1	2945818	AACGGCGTCGACACC	2945832
CP050974.1	9426651	AACGGCGTCGACACC	9426637
CP050974.1	10760095	AACGGCGTCGACACC	10760109
CP050975.1	5759656	ACGGCGTCGACACCTCC	5759672
CP050975.1	3089795	CGGCGTCGACACCTCC	3089780
CP050975.1	5005684	CGGCGTCGACACCTCC	5005699
CP050975.1	2084240	AACGGCGTCGACACC	2084254
CP046172.1	8207102	ACGGCGTCGACACCTCC	8207086
XM_032945539.1	1495	ACGGCGTCGACACCTCC	1511
CP049825.1	207129	AACGGCGTCGACACCTC	207145
CP049824.1	9970787	ACGGCGTCGACACCTCC	9970803
XM_032809184.1	2446	ACAAACGGCGTCGACAC	2430
XM_032809183.1	2450	ACAAACGGCGTCGACAC	2434
CP049199.1	2654895	AACGGCGTCGACACCTC	2654911
CP049199.1	3239860	CGGCGTCGACACCTC	3239846
CP041650.2	7554117	ACGGCGTCGACACCTCC	7554133
CP041650.2	1161359	CGGCGTCGACACCTC	1161373
CP041650.2	2673368	AACGGCGTCGACACC	2673382
CP041650.2	3305153	AACGGCGTCGACACC	3305167

CP041650.2	9576160	AACGGCGTCGACACC	9576146
CP041650.2	11021158	AACGGCGTCGACACC	11021172
CP048397.1	4900581	ACGGCGTCGACACCTCC	4900597
CP048397.1	1645955	AACGGCGTCGACACCT	1645940
CP048397.1	1267077	AACGGCGTCGACACC	1267091
CP041612.2	7317910	ACGGCGTCGACACCTCC	7317926
CP041612.2	858628	CGGCGTCGACACCTC	858642
CP041612.2	2496028	AACGGCGTCGACACC	2496042
CP041612.2	3098102	AACGGCGTCGACACC	3098116
CP041612.2	9486014	AACGGCGTCGACACC	9486000
CP041612.2	10881723	AACGGCGTCGACACC	10881737
CP041607.2	7106408	ACGGCGTCGACACCTCC	7106424
CP041607.2	782328	CGGCGTCGACACCTC	782342
CP041607.2	2292309	AACGGCGTCGACACC	2292323
CP041607.2	2924127	AACGGCGTCGACACC	2924141
CP041607.2	9147840	AACGGCGTCGACACC	9147826
CP041607.2	10593155	AACGGCGTCGACACC	10593169
CP041609.2	7199676	ACGGCGTCGACACCTCC	7199692
CP041609.2	784863	CGGCGTCGACACCTC	784877
CP041609.2	2298219	AACGGCGTCGACACC	2298233
CP041609.2	2930021	AACGGCGTCGACACC	2930035
CP041609.2	9277199	AACGGCGTCGACACC	9277185
CP041609.2	10723930	AACGGCGTCGACACC	10723944
CP041613.2	7552518	ACGGCGTCGACACCTCC	7552534
CP041613.2	1277247	CGGCGTCGACACCTC	1277261
CP041613.2	1314664	GGCGTCGACACCTCC	1314650
CP041613.2	2673915	AACGGCGTCGACACC	2673929
CP041613.2	3279128	AACGGCGTCGACACC	3279142
CP041613.2	9933257	AACGGCGTCGACACC	9933243
CP041613.2	11319675	AACGGCGTCGACACC	11319689
CP041604.2	7471604	ACGGCGTCGACACCTCC	7471620
CP041604.2	1103167	CGGCGTCGACACCTC	1103181
CP041604.2	1122128	CGGCGTCGACACCTC	1122114
CP041604.2	1246344	CGGCGTCGACACCTC	1246358
CP041604.2	2626808	AACGGCGTCGACACC	2626822
CP041604.2	3229941	AACGGCGTCGACACC	3229955
CP041604.2	9585963	AACGGCGTCGACACC	9585949
CP041604.2	10977721	AACGGCGTCGACACC	10977735
CP041602.2	7193261	ACGGCGTCGACACCTCC	7193277
CP041602.2	877655	CGGCGTCGACACCTC	877669
CP041602.2	915072	GGCGTCGACACCTCC	915058
CP041602.2	2332815	AACGGCGTCGACACC	2332829
CP041602.2	2938025	AACGGCGTCGACACC	2938039
CP041602.2	9589792	AACGGCGTCGACACC	9589778
CP041602.2	10981344	AACGGCGTCGACACC	10981358
CP041610.2	7104883	ACGGCGTCGACACCTCC	7104899
CP041610.2	809048	CGGCGTCGACACCTC	809062
CP041610.2	2275886	AACGGCGTCGACACC	2275900
CP041610.2	2880087	AACGGCGTCGACACC	2880101
CP041610.2	9351182	AACGGCGTCGACACC	9351168
CP041610.2	10818242	AACGGCGTCGACACC	10818256
AP022600.1	6097870	ACGGCGTCGACACCTCC	6097854
AP022600.1	57054	CGGCGTCGACACCTCC	57039
AP022600.1	368925	CGGCGTCGACACCTCC	368940
AP022561.1	5804988	ACGGCGTCGACACCTCC	5804972
AP022561.1	4225043	CGGCGTCGACACCTC	4225057
AP022561.1	4288025	AACGGCGTCGACACC	4288039
CP047385.1	4262943	CAAACGGCGTCGACACC	4262927
CP047386.1	2084339	CAAACGGCGTCGACACC	2084355
CP047183.1	1084993	ACGGCGTCGACACCTCC	1085009

CP047183.1	236126	AACGGCGTCGACACC	236112
CP043930.1	7805105	AAACGGCGTCGACACCT	7805089
XM_031649147.1	331	ACGGCGTCGACACCTCC	347
CP045643.1	3681359	ACGGCGTCGACACCTCC	3681343
CP045643.1	4463765	CGGCGTCGACACCTCC	4463750
CP045643.1	7339390	AACGGCGTCGACACC	7339376
XM_031503155.1	5971	AAACGGCGTCGACACCT	5955
XM_031478322.1	7684	AAACGGCGTCGACACCT	7668
CP023690.1	3643320	ACGGCGTCGACACCTCC	3643336
CP023690.1	2238430	AACGGCGTCGACACC	2238444
CP029194.1	6347693	ACGGCGTCGACACCTCC	6347709
CP029194.1	5353612	CGGCGTCGACACCTCC	5353627
CP043534.1	1232091	ACGGCGTCGACACCTCC	1232075
CP043448.1	102288	ACGGCGTCGACACCTCC	102304
XM_030681488.1	153	ACGGCGTCGACACCTCC	169
XM_030681487.1	153	ACGGCGTCGACACCTCC	169
XM_030588374.1	154	ACGGCGTCGACACCTCC	170
XM_030588373.1	154	ACGGCGTCGACACCTCC	170
XM_030588372.1	154	ACGGCGTCGACACCTCC	170
CP032580.1	54132560	CACAAACGGCGTCGACA	54132576
CP032571.1	204763	ACGGCGTCGACACCTCC	204779
XM_037413552.1	19	ACGGCGTCGACACCTCC	3
CP063451.1	27415	CAAACGGCGTCGACACC	27399
AP024094.1	2513669	CAAACGGCGTCGACACC	2513653
AP024094.1	318121	ACGGCGTCGACACCT	318107
CP061167.1	110524	ACGGCGTCGACACCTCC	110540
CP060496.1	3316655	ACGGCGTCGACACCTCC	3316671
CP048835.1	2484608	AACGGCGTCGACACCTC	2484624
CP048835.1	2633466	AACGGCGTCGACACC	2633480
CP041141.1	111235	ACGGCGTCGACACCTCC	111251
CP059696.1	1534571	AACGGCGTCGACACCTC	1534587
CP059696.1	3230849	CGGCGTCGACACCTCC	3230834
CP059696.1	637027	ACGGCGTCGACACCT	637013
CP059696.1	1664580	AACGGCGTCGACACC	1664594
XM_035673227.1	851	ACGGCGTCGACACCTCC	835
CP058693.1	9175631	ACGGCGTCGACACCTCC	9175615

PRODUCTS OF SIGNIFICANT ALIGNMENTS (CDS FEATURE)

- **RecA-like DNA recombinase** GenBank
- **Hypothetical protein** GenBank; GenBank
- **Cys-Gly metallo-di-peptidase; similar to YFR044C** GenBank; GenBank
- **NADH-quinone oxidoreductase subunit C** GenBank
- **Adenylosuccinate synthase** GenBank; GenBank; GenBank; GenBank; GenBank; GenBank; GenBank
- **Proteasome subunit alpha** GenBank; GenBank
- **Carbohydrate kinase family protein** GenBank; GenBank; GenBank
- **Autotransporter-associated beta strand repeat protein** GenBank
- **PEP-CTERM motif protein** GenBank
- **Septin-7-like** GenBank; GenBank; GenBank; GenBank; GenBank; GenBank; GenBank; GenBank; GenBank
- **RagB/SusD family nutrient uptake outer membrane protein** GenBank
- **Adenylate/guanylate cyclase domain-containing protein** GenBank
- **Helix-turn-helix protein** GenBank
- **2,4-dienoyl-CoA reductase [NADPH]** GenBank
- **NAD-specific glutamate dehydrogenase** GenBank; GenBank; GenBank
- **D-alanyl-D-alanine carboxypeptidase DacC precursor** GenBank
- **NADH dehydrogenase** GenBank
- **Sugar transferase** GenBank
- **Uncharacterized protein** GenBank; GenBank
- **Extracellular solute-binding protein** GenBank
- **FtsX-like permease family protein** GenBank
- **Glycosyltransferase** GenBank
- **Ammonium transporter** GenBank
- **Excinuclease ABC subunit C** GenBank
- **PfkB domain protein** GenBank
- **M56 family metallopeptidase** GenBank
- **(d)CMP kinase** GenBank
- **SDR family oxidoreductase** GenBank
- **Alpha/beta fold hydrolase** GenBank; GenBank
- **DEAD/DEAH box helicase** GenBank
- **Calcium-binding protein** GenBank
- **Sugar kinase** GenBank; GenBank; GenBank; GenBank; GenBank; GenBank; GenBank; GenBank; GenBank; GenBank; GenBank; GenBank; GenBank; GenBank; GenBank; GenBank
- **DUF3533 domain-containing protein** GenBank; GenBank; GenBank; GenBank; GenBank; GenBank; GenBank; GenBank; GenBank; GenBank
- **HAMP domain-containing histidine kinase** GenBank; GenBank; GenBank; GenBank; GenBank; GenBank; GenBank; GenBank; GenBank; GenBank
- **Sestrin-1-like** GenBank
- **Alpha-mannosidase 2** GenBank; GenBank
- **TIGR03620 family F420-dependent LLM class oxidoreductase** GenBank; GenBank; GenBank; GenBank; GenBank; GenBank; GenBank; GenBank
- **Esterase-like activity of phytase family protein** GenBank; GenBank; GenBank; GenBank; GenBank; GenBank; GenBank; GenBank
- **2-dehydro-3-deoxygluconokinase** GenBank
- **DUF3152 domain-containing protein** GenBank
- **DUF3472 domain-containing protein** GenBank; GenBank
- **Non-ribosomal peptide synthetase** GenBank
- **Sarcosine oxidase subunit beta** GenBank

- **Transcriptional regulator** GenBank
- **Microsomal epoxide hydrolase** GenBank
- **Cytochrome c** GenBank; GenBank
- **D-alanyl-D-alanine carboxypeptidase/D-alanyl-D-alanine-endopeptidase** GenBank
- **Bifunctional acetate--CoA ligase family protein/GNAT family N-acetyltransferase** GenBank
- **30S ribosomal protein S5, chloroplastic** GenBank
- **Neurogenic locus Notch protein-like** GenBank; GenBank
- **PAS domain-containing sensor histidine kinase** GenBank
- **Macrophage migration inhibitory factor** GenBank; GenBank; GenBank; GenBank; GenBank
- **LOW QUALITY PROTEIN: uncharacterized protein LOC119161190** GenBank
- **Relaxase domain-containing protein** GenBank
- **Murein hydrolase effector protein LrgB** GenBank
- **Protein asteroid homolog 1-like** GenBank

SEQUENCE N° 7/2

(T=2; C=1; G=0; A=3)

SEQUENCE N° 7/2 (T=2; C=1; G=0; A=3)

Original Sequence	G	T	G	A	A	A	T	G	G	T	C	A	T	G	T	G	T	G	G	C	G	G
Trend n° 7 (T=2; C=1; G=0; A=3)	0	2	0	3	3	3	2	0	0	2	1	3	2	0	2	0	2	0	0	1	0	0
Trend n° 7 (I i D d)	-	I	D	I	i	i	D	D	d	I	D	I	D	D	I	D	I	D	d	I	D	d
Trend n° 7/2	1	2	1	2	2	2	1	0	0	3	0	3	1	0	1	0	3	2	2	3	2	2
Sequence n° 7/2 (72,73%)	C	T	C	T	T	T	C	G	G	A	G	A	C	G	C	G	A	T	T	A	T	T

Query: **C T C T T T C G G A G A C G C G A T T A T T**

Search Parameters	
Program	blastn
Word size	7
Expect value	1000
Hitlist size	100
Match/Mismatch scores	1,-3
Gapcosts	5,2
Filter string	F
Genetic Code	1

100 sequences selected:

Accession Seq	Description	Comomm Name	Max Score	Total Score	Query C.	E value	Per. Id.	Acc. Len
MH155284.1	Caloplaca epithallina voucher Urbanavichus_sn2 internal transcribed spacer 1, partial sequence; 5.8S ribosomal RNA gene and internal transcribed spacer 2, complete sequence; and large subunit ribosomal RNA gene, partial sequence	Caloplaca epithallina	40.1	40.1	90%	0.57	100.00%	721
MH155283.1	Caloplaca epithallina voucher Urbanavichus_sn1 small subunit ribosomal RNA gene, partial sequence; internal transcribed spacer 1, 5.8S ribosomal RNA gene, and internal transcribed spacer 2, complete sequence; and large subunit ribosomal RNA gene, partial sequence	Caloplaca epithallina	40.1	40.1	90%	0.57	100.00%	838
CP062260.1	Paenibacillus sp. JNUCC-32 chromosome, complete genome	Paenibacillus sp. JNUCC-32	38.2	38.2	86%	2.2	100.00%	6941504
XM_025088965.1	PREDICTED: Cephus cinctus uncharacterized LOC107271692 (LOC107271692), transcript variant X25, mRNA	wheat stem sawfly	38.2	38.2	86%	2.2	100.00%	17927
XM_025088964.1	PREDICTED: Cephus cinctus uncharacterized LOC107271692 (LOC107271692), transcript variant X24, mRNA	wheat stem sawfly	38.2	38.2	86%	2.2	100.00%	17933
XM_015747967.2	PREDICTED: Cephus cinctus uncharacterized LOC107271692 (LOC107271692), transcript variant X23, mRNA	wheat stem sawfly	38.2	38.2	86%	2.2	100.00%	19153
XM_015747962.2	PREDICTED: Cephus cinctus uncharacterized LOC107271692 (LOC107271692), transcript variant X22, mRNA	wheat stem sawfly	38.2	38.2	86%	2.2	100.00%	15913
XM_015747961.2	PREDICTED: Cephus cinctus uncharacterized LOC107271692 (LOC107271692), transcript variant X21, mRNA	wheat stem sawfly	38.2	38.2	86%	2.2	100.00%	16292
XM_015747963.2	PREDICTED: Cephus cinctus uncharacterized LOC107271692 (LOC107271692), transcript variant X20, mRNA	wheat stem sawfly	38.2	38.2	86%	2.2	100.00%	16473
XM_025088962.1	PREDICTED: Cephus cinctus uncharacterized LOC107271692 (LOC107271692), transcript	wheat stem sawfly	38.2	38.2	86%	2.2	100.00%	18282

	variant X19, mRNA							
XM_025088961.1	PREDICTED: Cephus cinctus uncharacterized LOC107271692 (LOC107271692), transcript variant X18, mRNA	wheat stem sawfly	38.2	38.2	86%	2.2	100.00%	18114
XM_015747960.2	PREDICTED: Cephus cinctus uncharacterized LOC107271692 (LOC107271692), transcript variant X17, mRNA	wheat stem sawfly	38.2	38.2	86%	2.2	100.00%	19334
XM_025088960.1	PREDICTED: Cephus cinctus uncharacterized LOC107271692 (LOC107271692), transcript variant X16, mRNA	wheat stem sawfly	38.2	38.2	86%	2.2	100.00%	16569
XM_025088959.1	PREDICTED: Cephus cinctus uncharacterized LOC107271692 (LOC107271692), transcript variant X15, mRNA	wheat stem sawfly	38.2	38.2	86%	2.2	100.00%	16572
XM_025088958.1	PREDICTED: Cephus cinctus uncharacterized LOC107271692 (LOC107271692), transcript variant X14, mRNA	wheat stem sawfly	38.2	38.2	86%	2.2	100.00%	15997
XM_015747954.2	PREDICTED: Cephus cinctus uncharacterized LOC107271692 (LOC107271692), transcript variant X13, mRNA	wheat stem sawfly	38.2	38.2	86%	2.2	100.00%	17919
XM_025088957.1	PREDICTED: Cephus cinctus uncharacterized LOC107271692 (LOC107271692), transcript variant X12, mRNA	wheat stem sawfly	38.2	38.2	86%	2.2	100.00%	17153
XM_025088956.1	PREDICTED: Cephus cinctus uncharacterized LOC107271692 (LOC107271692), transcript variant X11, mRNA	wheat stem sawfly	38.2	38.2	86%	2.2	100.00%	19318
XM_015747952.2	PREDICTED: Cephus cinctus uncharacterized LOC107271692 (LOC107271692), transcript variant X10, mRNA	wheat stem sawfly	38.2	38.2	86%	2.2	100.00%	17737
XM_025088955.1	PREDICTED: Cephus cinctus uncharacterized LOC107271692 (LOC107271692), transcript variant X9, mRNA	wheat stem sawfly	38.2	38.2	86%	2.2	100.00%	19324
XM_025088954.1	PREDICTED: Cephus cinctus uncharacterized LOC107271692 (LOC107271692), transcript variant X8, mRNA	wheat stem sawfly	38.2	38.2	86%	2.2	100.00%	18629
XM_025088953.1	PREDICTED: Cephus cinctus uncharacterized LOC107271692 (LOC107271692), transcript variant X7, mRNA	wheat stem sawfly	38.2	38.2	86%	2.2	100.00%	18643
XM_025088952.1	PREDICTED: Cephus cinctus uncharacterized LOC107271692 (LOC107271692), transcript variant X6, mRNA	wheat stem sawfly	38.2	38.2	86%	2.2	100.00%	18648
XM_025088951.1	PREDICTED: Cephus cinctus uncharacterized LOC107271692 (LOC107271692), transcript variant X5, mRNA	wheat stem sawfly	38.2	38.2	86%	2.2	100.00%	18291
XM_025088950.1	PREDICTED: Cephus cinctus uncharacterized LOC107271692 (LOC107271692), transcript variant X4, mRNA	wheat stem sawfly	38.2	38.2	86%	2.2	100.00%	16623
XM_025088949.1	PREDICTED: Cephus cinctus uncharacterized LOC107271692 (LOC107271692), transcript variant X3, mRNA	wheat stem sawfly	38.2	38.2	86%	2.2	100.00%	16889
XM_025088948.1	PREDICTED: Cephus cinctus uncharacterized LOC107271692 (LOC107271692), transcript variant X2, mRNA	wheat stem sawfly	38.2	38.2	86%	2.2	100.00%	16641
XM_025088947.1	PREDICTED: Cephus cinctus uncharacterized LOC107271692 (LOC107271692), transcript variant X1, mRNA	wheat stem sawfly	38.2	38.2	86%	2.2	100.00%	16647
CP001793.1	Paenibacillus sp. Y412MC10, complete genome	Paenibacillus sp. Y412MC10	38.2	38.2	86%	2.2	100.00%	7121665
LR584071.1	Scleropages formosus genome assembly, chromosome: 6	Asian bonytongue	36.2	66.4	86%	8.9	100.00%	38292110
LR902823.1	Darwinula stevensoni	Darwinula stevensoni	36.2	72.4	81%	8.9	100.00%	32304
XM_033363171.1	PREDICTED: Belonocnema treatae uncharacterized LOC117174255 (LOC117174255),	Belonocnema treatae	36.2	36.2	81%	8.9	100.00%	3089

	mRNA							
LR134318.1 1/2 5/1 7/1 9/1	Pseudomonas fluorescens strain NCTC9428 genome assembly, chromosome: 1	Pseudomonas fluorescens	36.2	36.2	81%	8.9	100.00%	6033966
CP027082.1 1/2 3/6 10/1	Bos mutus isolate yakQH1 chromosome 14	wild yak	36.2	36.2	81%	8.9	100.00%	81354091
XM_019138661.1	Cryptococcus amylolentus CBS 6039 chitin synthase 1 (L202_04556), mRNA	Cryptococcus amylolentus CBS 6039	36.2	36.2	81%	8.9	100.00%	3964
XM_033324960.1	PREDICTED: Bombus impatiens uncharacterized LOC100748080 (LOC100748080), transcript variant X10, mRNA	common eastern bumble bee	34.2	34.2	77%	35	100.00%	6059
XM_033324959.1	PREDICTED: Bombus impatiens uncharacterized LOC100748080 (LOC100748080), transcript variant X9, mRNA	common eastern bumble bee	34.2	34.2	77%	35	100.00%	6140
XM_024372643.2	PREDICTED: Bombus impatiens uncharacterized LOC100748080 (LOC100748080), transcript variant X8, mRNA	common eastern bumble bee	34.2	34.2	77%	35	100.00%	6143
XM_033324958.1	PREDICTED: Bombus impatiens uncharacterized LOC100748080 (LOC100748080), transcript variant X7, mRNA	common eastern bumble bee	34.2	34.2	77%	35	100.00%	6616
XM_033324957.1	PREDICTED: Bombus impatiens uncharacterized LOC100748080 (LOC100748080), transcript variant X6, mRNA	common eastern bumble bee	34.2	34.2	77%	35	100.00%	6111
XM_024372641.2	PREDICTED: Bombus impatiens uncharacterized LOC100748080 (LOC100748080), transcript variant X5, mRNA	common eastern bumble bee	34.2	34.2	77%	35	100.00%	6619
XM_012394472.3	PREDICTED: Bombus impatiens uncharacterized LOC100748080 (LOC100748080), transcript variant X4, mRNA	common eastern bumble bee	34.2	34.2	77%	35	100.00%	6195
XM_024372640.2	PREDICTED: Bombus impatiens uncharacterized LOC100748080 (LOC100748080), transcript variant X3, mRNA	common eastern bumble bee	34.2	34.2	77%	35	100.00%	7680
XM_012394470.3	PREDICTED: Bombus impatiens uncharacterized LOC100748080 (LOC100748080), transcript variant X2, mRNA	common eastern bumble bee	34.2	34.2	77%	35	100.00%	6671
XM_012394469.3	PREDICTED: Bombus impatiens uncharacterized LOC100748080 (LOC100748080), transcript variant X1, mRNA	common eastern bumble bee	34.2	34.2	77%	35	100.00%	7684
MG592473.1	Vibrio phage 1.101.O._10N.261.45.C6, partial genome	Vibrio phage 1.101.O._10N.261.45.C6	34.2	34.2	77%	35	100.00%	130250
AP018207.1	Calothrix brevissima NIES-22 DNA, nearly complete genome	Calothrix brevissima NIES-22	34.2	34.2	77%	35	100.00%	8527291
LT629756.1 1/2 5/1 7/1 9/1	Pseudomonas sp. Z003-0.4C(8344-21) genome assembly, chromosome: I	Pseudomonas sp. Z003-0.4C(8344-21)	34.2	34.2	77%	35	100.00%	6113481
CP013264.1	Sphingobium baderi strain DE-13, complete genome	Sphingobium baderi	34.2	34.2	77%	35	100.00%	4107398
XM_012422314.1	PREDICTED: Orussus abietinus zinc finger protein 236 (LOC105698248), mRNA	Orussus abietinus	34.2	34.2	77%	35	100.00%	1906
AP023273.1	Agrobacterium vitis VAR06-30 DNA, chromosome 2, complete sequence	Agrobacterium vitis	34.2	34.2	77%	35	100.00%	1162882
XM_032820819.1	PREDICTED: Odontomachus brunneus beta-galactosidase (LOC116846673), transcript variant X3, mRNA	Odontomachus brunneus	34.2	34.2	77%	35	100.00%	2422
XM_032820818.1	PREDICTED: Odontomachus brunneus beta-galactosidase (LOC116846673), transcript variant X2, mRNA	Odontomachus brunneus	34.2	34.2	77%	35	100.00%	2161
XM_032820817.1	PREDICTED: Odontomachus brunneus beta-galactosidase (LOC116846673), transcript variant X1, mRNA	Odontomachus brunneus	34.2	34.2	77%	35	100.00%	2501
XM_032807220.1	PREDICTED: Odontomachus brunneus uncharacterized LOC116840469 (LOC116840469), transcript variant X2, mRNA	Odontomachus brunneus	34.2	34.2	95%	35	95.24%	7299

XM_032807219.1	PREDICTED: Odontomachus brunneus uncharacterized LOC116840469 (LOC116840469), transcript variant X1, mRNA	Odontomachus brunneus	34.2	34.2	95%	35	95.24%	7326
CP044989.1	Deinococcus sp. AJ005 plasmid p380k, complete sequence	Deinococcus sp. AJ005	34.2	34.2	77%	35	100.00%	380021
CP042997.1	Aquisphaera giovannonii strain OJF2 chromosome, complete genome	Aquisphaera giovannonii	34.2	34.2	77%	35	100.00%	10370956
LR584241.1	Takifugu rubripes genome assembly, chromosome: 11	torafugu	34.2	34.2	77%	35	100.00%	16367601
CP023741.1	Sphingobium yanoikuyae strain S72 chromosome, complete genome	Sphingobium yanoikuyae	34.2	34.2	77%	35	100.00%	5532633
XM_007513213.1	Bathycoccus prasinos predicted protein (Bathy05g03570), partial mRNA	Bathycoccus prasinos	34.2	34.2	77%	35	100.00%	2850
FO082274.1	Bathycoccus prasinos genomic : chromosome_5	Bathycoccus prasinos	34.2	34.2	77%	35	100.00%	1019276
XM_003113858.1	Caenorhabditis remanei hypothetical protein (CRE_26188) mRNA, complete cds	Caenorhabditis remanei	34.2	34.2	77%	35	100.00%	4470
FP929003.1	Candidatus Nitrospira defluvii chromosome, complete genome	Nitrospira defluvii	34.2	34.2	77%	35	100.00%	4317083
CP000634.1	Agrobacterium vitis S4 chromosome 2, complete sequence	Agrobacterium vitis S4	34.2	34.2	77%	35	100.00%	1283187
OA882732.1	Notodromas monacha	Notodromas monacha	32.2	32.2	72%	138	100.00%	85654
XM_037544960.1	PREDICTED: Pygocentrus nattereri glutamate receptor, ionotropic, N-methyl D-aspartate 2A, a (grin2aa), mRNA	red-bellied piranha	32.2	32.2	72%	138	100.00%	7232
XM_037258044.1	PREDICTED: Syngnathus acus heat shock protein b8 (hspb8), transcript variant X2, mRNA	greater pipefish	32.2	32.2	72%	138	100.00%	853
XM_037258043.1	PREDICTED: Syngnathus acus heat shock protein b8 (hspb8), transcript variant X1, mRNA	greater pipefish	32.2	32.2	72%	138	100.00%	980
CP062022.1	Macrobrachium nipponense isolate FS-2020 chromosome 18	Macrobrachium nipponense	32.2	32.2	72%	138	100.00%	90060326
CP062049.1	Macrobrachium nipponense isolate FS-2020 chromosome 31	Macrobrachium nipponense	32.2	32.2	72%	138	100.00%	73203386
CP043893.1	Leptospira interrogans serovar Bataviae strain 1489 chromosome 1, complete sequence	Leptospira interrogans serovar Bataviae	32.2	32.2	72%	138	100.00%	4431091
CP043891.1	Leptospira interrogans serovar Icterohaemorrhagiae strain 898 chromosome 1, complete sequence	Leptospira interrogans serovar Icterohaemorrhagiae	32.2	32.2	72%	138	100.00%	4280412
CP043884.1	Leptospira interrogans serovar Canicola strain 782 chromosome 1, complete sequence	Leptospira interrogans serovar Canicola	32.2	32.2	72%	138	100.00%	4556406
CP043880.1	Leptospira interrogans serovar Bataviae strain 1548 chromosome, complete genome	Leptospira interrogans serovar Bataviae	32.2	32.2	72%	138	100.00%	4354848
CP043876.1	Leptospira interrogans serovar Icterohaemorrhagiae strain Langkawi chromosome 1, complete sequence	Leptospira interrogans serovar Icterohaemorrhagiae	32.2	32.2	72%	138	100.00%	4367462
CP053164.1	Carposina sasakii isolate BJYQ chromosome 17	Carposina sasakii	32.2	32.2	72%	138	100.00%	13242407
CP061007.1	Saccharopolyspora spinosa strain CCTCC M206084 chromosome, complete genome	Saccharopolyspora spinosa	32.2	32.2	72%	138	100.00%	8876435
CP060577.1	Pyrenophora teres f. maculata isolate P-A14 chromosome 6	Pyrenophora teres f. maculata	32.2	32.2	72%	138	100.00%	2338523
CP060561.1	Pyrenophora teres f. maculata isolate FGOB10Ptm-1 chromosome 6	Pyrenophora teres f. maculata	32.2	32.2	72%	138	100.00%	2295326
CP060552.1	Pyrenophora teres f. maculata isolate NZKF2 chromosome 6	Pyrenophora teres f. maculata	32.2	32.2	72%	138	100.00%	2316045
CP060540.1	Pyrenophora teres f. maculata isolate DEN2.6 chromosome 6	Pyrenophora teres f. maculata	32.2	32.2	72%	138	100.00%	2451318
LR877232.1	**Acomys russatus genome assembly, chromosome: 21**	golden spiny mouse	32.2	32.2	72%	138	100.00%	61132863
OB736593.1	Cyprideis torosa	Cyprideis torosa	32.2	32.2	72%	138	100.00%	564
OB665008.1	Cyprideis torosa	Cyprideis torosa	32.2	32.2	72%	138	100.00%	12539
LR901227.1	Darwinula stevensoni	Darwinula stevensoni	32.2	32.2	72%	138	100.00%	59429

XM_037540769.1	PREDICTED: Pygocentrus nattereri T cell immune regulator 1, ATPase H+ transporting V0 subunit a3b (tcirg1b), mRNA	red-bellied piranha	32.2	32.2	90%	138	95.00%	3779
XR_005138601.1	PREDICTED: Triticum dicoccoides uncharacterized LOC119282005 (LOC119282005), ncRNA	Triticum dicoccoides	32.2	32.2	72%	138	100.00%	527
CP062018.1	Macrobrachium nipponense isolate FS-2020 chromosome 17	Macrobrachium nipponense	32.2	32.2	72%	138	100.00%	89928525
XR_004986512.1 1/2 2/2 3/2 3/3 3/5 6/1 8/1 9/1 10/1	***PREDICTED: Pipistrellus kuhlii Rap guanine nucleotide exchange factor 5 (RAPGEF5), transcript variant X7, misc_RNA***	Kuhl's pipistrelle	32.2	32.2	72%	138	100.00%	2448
XM_036429202.1 1/2 2/2 3/2 3/3 3/5 6/1 8/1 9/1 10/1	***PREDICTED: Pipistrellus kuhlii Rap guanine nucleotide exchange factor 5 (RAPGEF5), transcript variant X5, mRNA***	Kuhl's pipistrelle	32.2	32.2	72%	138	100.00%	2517
XM_036429201.1 1/2 2/2 3/2 3/3 3/5 6/1 8/1 9/1 10/1	***PREDICTED: Pipistrellus kuhlii Rap guanine nucleotide exchange factor 5 (RAPGEF5), transcript variant X4, mRNA***	Kuhl's pipistrelle	32.2	32.2	72%	138	100.00%	2640
XM_036429200.1 1/2 2/2 3/2 3/3 3/5 6/1 8/1 9/1 10/1	***PREDICTED: Pipistrellus kuhlii Rap guanine nucleotide exchange factor 5 (RAPGEF5), transcript variant X3, mRNA***	Kuhl's pipistrelle	32.2	32.2	72%	138	100.00%	2667
XM_036429199.1 1/2 2/2 3/2 3/3 3/5 6/1 8/1 9/1 10/1	***PREDICTED: Pipistrellus kuhlii Rap guanine nucleotide exchange factor 5 (RAPGEF5), transcript variant X2, mRNA***	Kuhl's pipistrelle	32.2	32.2	72%	138	100.00%	2786
XM_036429198.1 1/2 2/2 3/2 3/3 3/5 6/1 8/1 9/1 10/1	***PREDICTED: Pipistrellus kuhlii Rap guanine nucleotide exchange factor 5 (RAPGEF5), transcript variant X1, mRNA***	Kuhl's pipistrelle	32.2	32.2	72%	138	100.00%	2789
CP061876.1	Alternaria alternata strain Z7 chromosome 2, complete sequence	Alternaria alternata	32.2	32.2	72%	138	100.00%	5542836
LR880646.1	Poecilia reticulata genome assembly, chromosome: 2	guppy	32.2	32.2	72%	138	100.00%	45373629
XM_035929535.1	PREDICTED: Anopheles albimanus protein tumorous imaginal discs, mitochondrial-like (LOC118463153), transcript variant X2, Mrna	Anopheles albimanus	32.2	32.2	72%	138	100.00%	1793
XM_035929534.1	PREDICTED: Anopheles albimanus protein tumorous imaginal discs, mitochondrial-like (LOC118463153), transcript variant X1, mRNA	Anopheles albimanus	32.2	32.2	72%	138	100.00%	2018
LR862366.1 1/2 2/2 3/2 3/3 3/5 6/1 8/1 9/1 10/1	***Pipistrellus pipistrellus genome assembly, chromosome: 10***	common pipistrelle	32.2	32.2	72%	138	100.00%	71504835

PREDICTED: Pipistrellus kuhlii Rap guanine nucleotide exchange factor 5 (RAPGEF5), transcript variant X1, mRNA

NCBI Reference Sequence: XM_036429198.1

FASTA Graphics

Go to:

LOCUS XM_036429198 2789 bp mRNA linear MAM **01-OCT-2020**

DEFINITION PREDICTED: Pipistrellus kuhlii Rap guanine nucleotide exchange factor 5 (RAPGEF5), transcript variant X1, mRNA.

ACCESSION XM_036429198

VERSION XM_036429198.1

DBLINK BioProject: PRJNA666119

KEYWORDS RefSeq.

SOURCE Pipistrellus kuhlii (Kuhl's pipistrelle)

ORGANISM Pipistrellus kuhlii

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Laurasiatheria; Chiroptera; Microchiroptera; Vespertilionidae; Pipistrellus.

COMMENT MODEL REFSEQ: This record is predicted by automated computational analysis. This record is derived from a genomic sequence (NW_023425405.1) annotated using gene prediction method: Gnomon.

Also see:

Documentation of NCBI's Annotation Process

##Genome-Annotation-Data-START##

Annotation Provider :: NCBI

Annotation Status :: Full annotation

Annotation Name :: Pipistrellus kuhlii Annotation Release 100

Annotation Version :: 100

Annotation Pipeline :: NCBI eukaryotic genome annotation pipeline

Annotation Software Version :: 8.5

Annotation Method :: Best-placed RefSeq; Gnomon

Features Annotated :: Gene; mRNA; CDS; ncRNA

##Genome-Annotation-Data-END##

FEATURES Location/Qualifiers

source 1..2789

/organism="Pipistrellus kuhlii"

/mol_type="mRNA"

/isolate="mPipKuh1"

/db_xref="taxon:59472"

/chromosome="Unknown"

/sex="male"

/tissue_type="flight muscle, heart, brain, liver, kidney"

/dev_stage="adult"

/country="Italy: Bergamo"
/lat_lon="45.7430 N 9.5831 E"
/collection_date="2017"
/collected_by="Emma Teeling, Andrea Locatelli"

gene 1..2789
/gene="RAPGEF5"
/note="Derived by automated computational analysis using gene prediction method: Gnomon. Supporting evidence includes similarity to: 2 Proteins, and 100% coverage of the annotated genomic feature by RNAseq alignments, including 14 samples with support for all annotated introns"
/db_xref="GeneID:118714204"

CDS 53..2713
/gene="RAPGEF5"
/codon_start=1
/product="rap guanine nucleotide exchange factor 5 isoform X1"
/protein_id="XP_036285091.1"
/db_xref="GeneID:118714204"
/translation="MRMAVGSVKMQPPPCESPALARAVAAAAADSALRRSPSAREPEP
EPEQPPAPPRPRLRDLPALLRSGLTLRRKRSVGGGGRTLQRRISNPYLEHTSSQIYGE
NSSCAGRALRNIIIVQAADLIKDRVNLKGFYRRSCVGSELVDWLLEHCPFIQCRSTAI
GVWQLLLDMGILSSVDQHLYFQDTYVFYQFSSDECSYLYCEFEREEEWQNGVKMLLQL
VPVIRTRAGLCELSHRKIEDSEESSDEILARLTSAVQRELAAVIALKAKKSAIEQDEE
SSDKHIPTAEADSDPDPQAGVMCKLQGREDIGRIELVQKLARENCQFLQTDRKEQDKS
EQQDGEVTTVREQEQSVLVLRKVQRRGPAPPGGSADSDWRYVVVSGTPEKILEHLLND
LHLEEVQDKETETLLDDFLLTYTVFMTTDDLCQALLRHYSAKKYQGKEENSDVPRRKR
KVLHLVSQWISLYKDWLHEDEHSKMFLKTIYRNVLDDVYEYPILEKELKEFQKILGMH
RRHTVDEYSPQRKNKALFHQFSLKENWLQHRGAVTETEEIFCHVYITEHSYVSVKAKV
SSTAQEVLRVVAEKIQHAEEDLALVAVSFSGGKRELHPNDLAISKSLEASGRIYVYRK
NLADTLNPFAENEESQQRPLRILGMNTWDLALELMNFDWSLFNSIHEQELIYFTFSRQ
GSGEHTANLSLLLQRCNEVQLWVATEILLCSQLGKRVQLVKKFIKIAAHCKAQRNLNS
FFAIVMGLNTASVSRLSQTWEKIPGKFKKLFSELESLTDPSLNHKAYRDTFKKMKPPK
IPFMPLLLKDVTFIHEGNKTFLDNLVNFEKLHMIADTVRTLRHCRANQFGGDASPKEH
HELKSYVTHLLVIDSQQALFELSHRIEPRA"

ORIGIN

```
        1 cgctgccggg agaagccacg ggagctgcgg gcgccgcggc cgccgccggg gcatgaggat
       61 ggccgtgggc tccgtcaaga tgcagccgcc gccgtgcgag agcccggcgc tggcccgggc
      121 ggtggcggcg gcggcggcgg acagcgccct gcgccgcagc cccagcgccc gcgagcccga
      181 gcccgagccc gagcagccgc ccgccccgcc gcggccgcgg ctgcgggacc tgccggcgct
      241 gctgcggagc gggctcacgc tgcggaggaa gcggagcgtc ggcggcggcg gccggactct
      301 ccagagaagg atttccaatc catacctcga gcacacttct tcccagattt atggggagaa
```

```
 361 ttcttcttgt gcaggaagag cactgaggaa tatcattatc gtccaggcag ctgatctgat
 421 caaggacaga gtgaacctca agggatttta caggcgaagc tgtgttggat cagaactggt
 481 cgactggctt ctggaacact gtcctttcat ccagtgcaga tccacggcca taggggtctg
 541 gcagctccta ctggacatgg ggattctgtc atcagtggat cagcatttat actttcagga
 601 cacctacgtt ttctatcagt tttcatctga tgaatgtagc tatctatact gtgagttcga
 661 acgagaagaa gaatggcaaa atggtgtcaa aatgttactg cagcttgtgc ctgtcattcg
 721 caccagagct ggcctctgtg aactgtctca tcggaaaatc gaagactcgg aagaaagcag
 781 tgatgaaatt cttgcgcgtc taacgtctgc tgtgcagaga gagctagcag ctgtcattgc
 841 tttgaaagca aagaagtctg caattgaaca agacgaggag agcagtgaca agcacatacc
 901 cacggcagaa gctgacagtg atccagatcc tcaggcaggg gtgatgtgca agcttcaggg
 961 aagagaggac atcgggcgca ttgagctggt ccagaagctg gcgcgggaga actgtcagtt
1021 tttgcagacg gacagaaaag aacaggacaa atctgaacag caggatggtg aggtgaccac
1081 agtcagggag caggagcaga gcgtcctggt gctgaggaag gtgcagcgca gaggcccggc
1141 ccccccagga gggagcgctg acagcgattg gagatacgtg gtcgtgtccg ggacccccga
1201 gaagatcttg gagcatcttt tgaatgactt gcacctggaa gaagtccagg acaaagaaac
1261 agagaccctc ctggatgact tcctcctcac ctacaccgtc ttcatgacca cggacgactt
1321 gtgccaggcg ctgctgagac actattctgc caagaagtat caaggcaaag aagaaaactc
1381 agatgttcct cgtcggaaac gcaaagtctt gcatcttgtc tcccagtgga tttctctgta
1441 caaagactgg ctacacgaag acgagcactc caaaatgttc ttaaagacca tatacaggaa
1501 cgttctggat gatgtctatg aatacccaat acttgaaaaa gaattgaagg aatttcaaaa
1561 gatacttgga atgcaccgtc gtcacaccgt tgatgagtac tctccacaga ggaagaataa
1621 ggccctttc caccaattca gtcttaagga gaactggctc cagcaccgag gcgctgtgac
1681 tgaaacagag gaaatttttt gccatgtgta catcacggag cactcgtacg tcagcgtgaa
1741 ggcgaaagtt tccagtacgg ctcaggaggt cctaagagtc gtggcggaga agatccagca
1801 cgcggaagag gacctggccc tggtggccgt ctccttctct gggggaaagc gtgaacttca
1861 tcccaatgat ctagccatct ccaaatccct ggaagcatct ggtcggatat atgtctatcg
1921 gaaaaacctg gctgacactt tgaacccttt tgccgaaaac gaggagtccc agcaaaggcc
1981 gttgaggatc ttgggaatga acacttggga tcttgctctg gaattaatga attttgattg
2041 gagtctcttc aattcgattc acgagcaaga gctgatctac ttcacattca gcagacaagg
2101 cagtggggag cacaccgcga acctcagcct cctgctccag cgctgcaatg aggtccagct
2161 ctgggtggcc acggagattc tgctctgcag ccagctgggc aagcgggtgc agctggtgaa
2221 gaagttcatc aagatcgccg cccactgcaa agcccagagg aacctgaact ctttctttgc
2281 tatcgtgatg ggtctcaaca cggcttctgt cagccgcctg tcgcaaacct gggagaaaat
2341 ccctgggaaa tttaagaaac ttttctctga acttgaaagc ttaacagacc cttccctgaa
2401 ccacaaggcc tacagagata ccttcaaaaa gatgaagccg cccaagatcc ccttcatgcc
2461 cttactgctc aaagatgtga catttattca cgagggaaac aaaacatttt tggataatct
2521 cgtcaatttt gagaagctgc atatgattgc agacaccgtc cgaaccctga ggcactgcag
2581 ggccaaccag tttggagggg acgcgtctcc gaaagagcac cacgagctga agtcctacgt
2641 cactcacctg ctcgtcatcg acagccagca ggccctgttc gagctctccc atcggatcga
2701 gccccgggcc tgagccggcc gcgccccctg agcacctgca gctctcccgc ccgggatggc
2761 agtgcccggc atgcccctct cctgggaga
//
```

PREDICTED: Pipistrellus kuhlii Rap guanine nucleotide exchange factor 5 (RAPGEF5), transcript variant X1, mRNA

NCBI Reference Sequence: XM_036429198.1

FASTA Graphics

Go to:

LOCUS XM_036429198 16 bp mRNA linear MAM **01-OCT-2020**

DEFINITION PREDICTED: Pipistrellus kuhlii Rap guanine nucleotide exchange factor 5 (RAPGEF5), transcript variant X1, mRNA.

ACCESSION XM_036429198 REGION: 2602..2617

VERSION XM_036429198.1

DBLINK BioProject: PRJNA666119

KEYWORDS RefSeq.

SOURCE Pipistrellus kuhlii (Kuhl's pipistrelle)

ORGANISM Pipistrellus kuhlii

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Laurasiatheria; Chiroptera; Microchiroptera; Vespertilionidae; Pipistrellus.

COMMENT MODEL REFSEQ: This record is predicted by automated computational analysis. This record is derived from a genomic sequence (NW_023425405.1) annotated using gene prediction method: Gnomon.

Also see:

Documentation of NCBI's Annotation Process

##Genome-Annotation-Data-START##

Annotation Provider :: NCBI

Annotation Status :: Full annotation

Annotation Name :: Pipistrellus kuhlii Annotation Release 100

Annotation Version :: 100

Annotation Pipeline :: NCBI eukaryotic genome annotation pipeline

Annotation Software Version :: 8.5

Annotation Method :: Best-placed RefSeq; Gnomon

Features Annotated :: Gene; mRNA; CDS; ncRNA

##Genome-Annotation-Data-END##

FEATURES Location/Qualifiers

source 1..16

/organism="Pipistrellus kuhlii"

/mol_type="mRNA"

/isolate="mPipKuh1"

/db_xref="taxon:59472"

/chromosome="Unknown"

/sex="male"

/tissue_type="flight muscle, heart, brain, liver, kidney"

/dev_stage="adult"

/country="Italy: Bergamo"
/lat_lon="45.7430 N 9.5831 E"
/collection_date="2017"
/collected_by="Emma Teeling, Andrea Locatelli"
gene <1..>16
/gene="RAPGEF5"
/note="Derived by automated computational analysis using gene prediction method: Gnomon. Supporting evidence includes similarity to: 2 Proteins, and 100% coverage of the annotated genomic feature by RNAseq alignments, including 14 samples with support for all annotated introns"
/db_xref="GeneID:118714204"
CDS <1..>16
/gene="RAPGEF5"
/codon_start=2
/product="rap guanine nucleotide exchange factor 5 isoform X1"
/protein_id="XP_036285091.1"
/db_xref="GeneID:118714204"
/translation="MRMAVGSVKMQPPPCESPALARAVAAAAADSALRRSPSAREPEP
EPEQPPAPPRPRLRDLPALLRSGLTLRRKRSVGGGGRTLQRRISNPYLEHTSSQIYGE
NSSCAGRALRNIIIVQAADLIKDRVNLKGFYRRSCVGSELVDWLLEHCPFIQCRSTAI
GVWQLLLDMGILSSVDQHLYFQDTYVFYQFSSDECSYLYCEFEREEEWQNGVKMLLQL
VPVIRTRAGLCELSHRKIEDSEESSDEILARLTSAVQRELAAVIALKAKKSAIEQDEE
SSDKHIPTAEADSDPDPQAGVMCKLQGREDIGRIELVQKLARENCQFLQTDRKEQDKS
EQQDGEVTTVREQEQSVLVLRKVQRRGPAPPGGSADSDWRYVVVSGTPEKILEHLLND
LHLEEVQDKETETLLDDFLLTYTVFMTTDDLCQALLRHYSAKKYQGKEENSDVPRRKR
KVLHLVSQWISLYKDWLHEDEHSKMFLKTIYRNVLDDVYEYPILEKELKEFQKILGMH
RRHTVDEYSPQRKNKALFHQFSLKENWLQHRGAVTETEEIFCHVYITEHSYVSVKAKV
SSTAQEVLRVVAEKIQHAEEDLALVAVSFSGGKRELHPNDLAISKSLEASGRIYVYRK
NLADTLNPFAENEESQQRPLRILGMNTWDLALELMNFDWSLFNSIHEQELIYFTFSRQ
GSGEHTANLSLLLQRCNEVQLWVATEILLCSQLGKRVQLVKKFIKIAAHCKAQRNLNS
FFAIVMGLNTASVSRLSQTWEKIPGKFKKLFSELESLTDPSLNHKAYRDTFKKMKPPK
IPFMPLLLKDVTFIHEGNKTFLDNLVNFEKLHMIADTVRTLRHCRANQFGGDASPKEH
HELKSYVTHLLVIDSQQALFELSHRIEPRA"
ORIGIN
1 cgcgtctccg aaagag
//

ALIGNMENT VIEW FLAT QUERY-ANCHORED WITH LETTERS FOR IDENTITIES

Query range 1: 1 to 22

Query	1	CTCTTTCGGAGACGCGATTATT	22
MH155284.1	376	CTTTCGGAGACGCGATTATT	357
MH155283.1	493	CTTTCGGAGACGCGATTATT	474
CP062260.1	1929896	TTTCGGAGACGCGATTATT	1929914
XM_025088965.1	4832	CTCTTTCGGAGACGCGATT	4814
XM_025088964.1	4838	CTCTTTCGGAGACGCGATT	4820
XM_015747967.2	6058	CTCTTTCGGAGACGCGATT	6040
XM_015747962.2	2818	CTCTTTCGGAGACGCGATT	2800
XM_015747961.2	3197	CTCTTTCGGAGACGCGATT	3179
XM_015747963.2	3378	CTCTTTCGGAGACGCGATT	3360
XM_025088962.1	5187	CTCTTTCGGAGACGCGATT	5169
XM_025088961.1	5019	CTCTTTCGGAGACGCGATT	5001
XM_015747960.2	6239	CTCTTTCGGAGACGCGATT	6221
XM_025088960.1	3552	CTCTTTCGGAGACGCGATT	3534
XM_025088959.1	3552	CTCTTTCGGAGACGCGATT	3534
XM_025088958.1	2902	CTCTTTCGGAGACGCGATT	2884
XM_015747954.2	4824	CTCTTTCGGAGACGCGATT	4806
XM_025088957.1	4058	CTCTTTCGGAGACGCGATT	4040
XM_025088956.1	6223	CTCTTTCGGAGACGCGATT	6205
XM_015747952.2	4642	CTCTTTCGGAGACGCGATT	4624
XM_025088955.1	6229	CTCTTTCGGAGACGCGATT	6211
XM_025088954.1	5534	CTCTTTCGGAGACGCGATT	5516
XM_025088953.1	5548	CTCTTTCGGAGACGCGATT	5530
XM_025088952.1	5553	CTCTTTCGGAGACGCGATT	5535
XM_025088951.1	5196	CTCTTTCGGAGACGCGATT	5178
XM_025088950.1	3546	CTCTTTCGGAGACGCGATT	3528
XM_025088949.1	3794	CTCTTTCGGAGACGCGATT	3776
XM_025088948.1	3546	CTCTTTCGGAGACGCGATT	3528
XM_025088947.1	3552	CTCTTTCGGAGACGCGATT	3534
CP001793.1	3933978	TTTCGGAGACGCGATTATT	3933960
LR584071.1	31326350	CTTTCGGAGACGCGATTA	31326367
LR584071.1	27485450	TCTTTCGGAGACGCG	27485436
LR902823.1	4961	TCTTTCGGAGACGCGATT	4978
LR902823.1	8325	TCTTTCGGAGACGCGATT	8342
XM_033363171.1	1374	TTCGGAGACGCGATTATT	1357
LR134318.1	421823	TTCGGAGACGCGATTATT	421806
CP027082.1	80817758	TTCGGAGACGCGATTATT	80817741
XM_019138661.1	3041	TCTTTCGGAGACGCGATT	3024
XM_033324960.1	4758	CTTTCGGAGACGCGATT	4774
XM_033324959.1	4839	CTTTCGGAGACGCGATT	4855
XM_024372643.2	4842	CTTTCGGAGACGCGATT	4858
XM_033324958.1	5315	CTTTCGGAGACGCGATT	5331
XM_033324957.1	4810	CTTTCGGAGACGCGATT	4826
XM_024372641.2	5318	CTTTCGGAGACGCGATT	5334
XM_012394472.3	4894	CTTTCGGAGACGCGATT	4910
XM_024372640.2	6379	CTTTCGGAGACGCGATT	6395
XM_012394470.3	5370	CTTTCGGAGACGCGATT	5386
XM_012394469.3	6383	CTTTCGGAGACGCGATT	6399
MG592473.1	59030	TCTTTCGGAGACGCGAT	59046
AP018207.1	6429381	TCTTTCGGAGACGCGAT	6429397
LT629756.1	5906409	TCGGAGACGCGATTATT	5906425
CP013264.1	3247187	TCTTTCGGAGACGCGAT	3247203

XM_012422314.1	432	TCGGAGACGCGATTATT	448
AP023273.1	71221	TCTTTCGGAGACGCGAT	71237
XM_032820819.1	1673	CTTTCGGAGACGCGATT	1657
XM_032820818.1	1412	CTTTCGGAGACGCGATT	1396
XM_032820817.1	1123	CTTTCGGAGACGCGATT	1107
XM_032807220.1	6941	TCTTTCTGAGACGCGATTATT	6921
XM_032807219.1	6968	TCTTTCTGAGACGCGATTATT	6948
CP044989.1	123135	TCGGAGACGCGATTATT	123119
CP042997.1	7633838	CTTTCGGAGACGCGATT	7633822
LR584241.1	3301410	TCTTTCGGAGACGCGAT	3301394
CP023741.1	1579636	TTCGGAGACGCGATTAT	1579620
XM_007513213.1	2073	CTCTTTCGGAGACGCGA	2057
FO082274.1	648929	CTCTTTCGGAGACGCGA	648913
XM_003113858.1	685	TCGGAGACGCGATTATT	669
FP929003.1	2288837	TCTTTCGGAGACGCGAT	2288821
CP000634.1	709168	TCTTTCGGAGACGCGAT	709152
OA882732.1	38100	CTCTTTCGGAGACGCG	38115
XM_037544960.1	1451	CTCTTTCGGAGACGCG	1466
XM_037258044.1	165	CTTTCGGAGACGCGAT	180
XM_037258043.1	166	CTTTCGGAGACGCGAT	181
CP062022.1	54352183	CGGAGACGCGATTATT	54352198
CP062049.1	25998022	CTCTTTCGGAGACGCG	25998037
CP043893.1	3895188	TCGGAGACGCGATTAT	3895203
CP043891.1	3755873	TCGGAGACGCGATTAT	3755888
CP043884.1	4031545	TCGGAGACGCGATTAT	4031560
CP043880.1	3830381	TCGGAGACGCGATTAT	3830396
CP043876.1	3835920	TCGGAGACGCGATTAT	3835935
CP053164.1	9752269	TTCGGAGACGCGATTA	9752284
CP061007.1	7161695	TTTCGGAGACGCGATT	7161710
CP060577.1	345982	TTCGGAGACGCGATTA	345997
CP060561.1	322497	TTCGGAGACGCGATTA	322512
CP060552.1	279310	TTCGGAGACGCGATTA	279325
CP060540.1	421508	TTCGGAGACGCGATTA	421523
LR877232.1	16749084	CTCTTTCGGAGACGCG	16749099
OB736593.1	165	CTCTTTCGGAGACGCG	150
OB665008.1	10526	CTCTTTCGGAGACGCG	10511
LR901227.1	24185	CTCTTTCGGAGACGCG	24170
XM_037540769.1	584	CTCTTTCGGAGACGAGATTA	565
XR_005138601.1	369	TTTCGGAGACGCGATT	354
CP062018.1	56490497	CGGAGACGCGATTATT	56490482
XR_004986512.1	2388	CTCTTTCGGAGACGCG	2373
XM_036429202.1	2345	CTCTTTCGGAGACGCG	2330
XM_036429201.1	2468	CTCTTTCGGAGACGCG	2453
XM_036429200.1	2495	CTCTTTCGGAGACGCG	2480
XM_036429199.1	2614	CTCTTTCGGAGACGCG	2599
XM_036429198.1	2617	CTCTTTCGGAGACGCG	2602
CP061876.1	4788579	TCTTTCGGAGACGCGA	4788564
LR880646.1	25246344	CTCTTTCGGAGACGCG	25246329
XM_035929535.1	458	CTCTTTCGGAGACGCG	443
XM_035929534.1	461	CTCTTTCGGAGACGCG	446
LR862366.1	31329587	CTCTTTCGGAGACGCG	31329572

PRODUCTS OF SIGNIFICANT ALIGNMENTS (CDS FEATURE)

- **Epoxide hydrolase** GenBank; GenBank
- **Uncharacterized protein LOC107271692** GenBank; GenBank
- **Fimbrial assembly** GenBank
- **Chitin synthase 1** GenBank
- **Hypothetical protein** GenBank; GenBank; GenBank
- **Type IV pilus assembly protein PilN**GenBank
- **Metal-dependent hydrolase** GenBank
- **Beta-galactosidase** GenBank; GenBank; GenBank
- **Sugar ABC transporter ATP-binding protein** GenBank
- **Calcineurin-like phosphoesterase** GenBank
- **Predicted protein** GenBank; GenBank
- **UniProtKB/TrEMBL:**D8PFL3 GenBank
- **KluB [Agrobacterium vitis S4]** ACM38767.1
- **Glutamate receptor ionotropic, NMDA 2A** GenBank
- **Heat shock protein beta-8** GenBank; GenBank
- **Adenylate cyclase** GenBank; GenBank; GenBank
- **Guanylate cyclase** GenBank; GenBank
- **Rap guanine nucleotide exchange factor** GenBank; GenBank; GenBank; GenBank; GenBank
- **Protein tumorous imaginal discs, mitochondrial-like** GenBank; GenBank

SEQUENCE N° 9/1

(T=2; C=0; G=1; A=3)

SEQUENCE N° 9/1 (T=2; C=0; G=1; A=3)

Original Sequence	**G**	**T**	**G**	**A**	**A**	**A**	**T**	**G**	**G**	**T**	**C**	**A**	**T**	**G**	**T**	**G**	**T**	**G**	**G**	**C**	**G**	**G**
Trend n° 9 (T=2; C=0; G=1; A=3)	**1**	**2**	**1**	**3**	**3**	**3**	**2**	**1**	**1**	**2**	**0**	**3**	**2**	**1**	**2**	**1**	**2**	**1**	**1**	**0**	**1**	**1**
Trend n° 9 (I i D d)	**-**	**I**	**D**	**I**	**i**	**i**	**D**	**D**	**d**	**I**	**D**	**I**	**D**	**D**	**I**	**D**	**I**	**D**	**d**	**D**	**I**	**i**
Trend n° 9/1	**0**	**1**	**0**	**2**	**2**	**2**	**1**	**0**	**0**	**3**	**1**	**2**	**1**	**0**	**3**	**2**	**3**	**2**	**2**	**1**	**2**	**2**
Sequence n° 9/1 (100%)	**C**	**G**	**C**	**T**	**T**	**T**	**G**	**C**	**C**	**A**	**G**	**T**	**G**	**C**	**A**	**T**	**A**	**T**	**T**	**G**	**T**	**T**

Query: **CGCTTTGCCAGTGCATATTGTT**

Search Parameters	
Program	blastn
Word size	7
Expect value	1000
Hitlist size	100
Match/Mismatch scores	1,-3
Gapcosts	5,2
Filter string	F
Genetic Code	1

100 sequences selected:

Accession Seq	Description	Comomm Name	Max Score	Total Score	Query C.	E value	Per. Id.	Acc. Len
LR812067.1	Danio rerio genome assembly, chromosome: 5	zebrafish	40.1	70.4	90%	0.57	100.00%	72587816
NC_041308.1	Johnius borneensis isolate FJDS20160816 mitochondrion, complete genome	sharpnose hammer croaker	40.1	40.1	90%	0.57	100.00%	18630
LR812598.1	Danio rerio genome assembly, chromosome: 5	zebrafish	40.1	70.4	90%	0.57	100.00%	67558190
BX005265.6	Zebrafish DNA sequence from clone CH211-254O16 in linkage group 5, complete sequence	zebrafish	40.1	40.1	90%	0.57	100.00%	162599
LR877239.1	**Acomys russatus genome assembly, chromosome: 28**	golden spiny mouse	40.1	40.1	90%	0.57	100.00%	43538471
XM_027028742.2	PREDICTED: Electrophorus electricus homeobox and leucine zipper encoding a (homeza), mRNA	electric eel	40.1	40.1	90%	0.57	100.00%	3885
LR606196.1 **1/2 3/2 3/3 3/5**	Aquila chrysaetos chrysaetos genome assembly, chromosome: 16	Aquila chrysaetos chrysaetos	38.2	38.2	86%	2.2	100.00%	30610376
LR812527.1	Danio kyathit genome assembly, chromosome: 9	Danio kyathit	38.2	102	86%	2.2	100.00%	70176753
LR535835.1	Mastacembelus armatus genome assembly, chromosome: 3	zig-zag eel	38.2	38.2	86%	2.2	100.00%	28130073
AP018449.1	Methylomusa anaerophila DNA, complete genome	Methylomusa anaerophila	38.2	38.2	86%	2.2	100.00%	4781198
CP062041.1	Macrobrachium nipponense isolate FS-2020 chromosome 46	Macrobrachium nipponense	38.2	68.4	86%	2.2	100.00%	48001745
LR812525.1	Danio kyathit genome assembly, chromosome: 7	Danio kyathit	38.2	68.4	86%	2.2	100.00%	89469171
AP022687.1	Epinephelus fuscoguttatus DNA, LG13, complete sequence	brown-marbled grouper	36.2	36.2	81%	8.9	100.00%	43720812
AP022676.1	Epinephelus fuscoguttatus DNA, LG2, complete sequence	brown-marbled grouper	36.2	66.4	100%	8.9	100.00%	52288159
AP019298.1	Prunus dulcis DNA, pseudomolecule Pd02	almond	36.2	36.2	81%	8.9	100.00%	24875702
XM_002182111.1	Phaeodactylum tricornutum CCAP 1055/1 predicted protein partial mRNA	Phaeodactylum tricornutum CCAP 1055/1	36.2	36.2	81%	8.9	100.00%	1977
XM_001387892.1	Scheffersomyces stipitis CBS 6054 Bud Emergence Mediator (BEM1), partial mRNA	Scheffersomyces stipitis CBS 6054	36.2	36.2	81%	8.9	100.00%	2123
AC110241.17	Mus musculus chromosome 16, clone RP23-129F22, complete sequence	house mouse	36.2	36.2	81%	8.9	100.00%	256720
AC113180.5	Mus musculus chromosome 16, clone RP23-366B13, complete sequence	house mouse	36.2	36.2	81%	8.9	100.00%	215198
XM_034347292.1	PREDICTED: Prunus dulcis myosin-12	almond	36.2	36.2	81%	8.9	100.00%	4872

	(LOC117617764), transcript variant X3, mRNA							
XM_034347291.1	PREDICTED: Prunus dulcis myosin-12 (LOC117617764), transcript variant X2, mRNA	almond	36.2	36.2	81%	8.9	100.00%	4955
XM_034347290.1	PREDICTED: Prunus dulcis myosin-12 (LOC117617764), transcript variant X1, mRNA	almond	36.2	36.2	81%	8.9	100.00%	5000
LR792557.1	Pseudochaenichthys georgianus genome assembly, chromosome: 12	South Georgia icefish	36.2	36.2	81%	8.9	100.00%	30946200
CP029289.2	Acidianus brierleyi strain DSM 1651 chromosome, complete genome	Acidianus brierleyi	36.2	36.2	81%	8.9	100.00%	2947244
CP027284.1	Lateolabrax maculatus chromosome Lm23	spotted sea bass	36.2	96.6	95%	8.9	100.00%	28603024
XM_022004535.1	PREDICTED: Spinacia oleracea calcium-transporting ATPase 8, plasma membrane-type-like (LOC110799299), transcript variant X6, mRNA	spinach	36.2	36.2	81%	8.9	100.00%	3193
XM_022004534.1	PREDICTED: Spinacia oleracea calcium-transporting ATPase 8, plasma membrane-type-like (LOC110799299), transcript variant X5, mRNA	spinach	36.2	36.2	81%	8.9	100.00%	3313
XM_022004533.1	PREDICTED: Spinacia oleracea calcium-transporting ATPase 8, plasma membrane-type-like (LOC110799299), transcript variant X4, mRNA	spinach	36.2	36.2	81%	8.9	100.00%	3300
XM_022004532.1	PREDICTED: Spinacia oleracea calcium-transporting ATPase 8, plasma membrane-type-like (LOC110799299), transcript variant X3, mRNA	spinach	36.2	36.2	81%	8.9	100.00%	3712
XM_022004531.1	PREDICTED: Spinacia oleracea calcium-transporting ATPase 8, plasma membrane-type-like (LOC110799299), transcript variant X2, mRNA	spinach	36.2	36.2	81%	8.9	100.00%	3720
XM_022004530.1	PREDICTED: Spinacia oleracea calcium-transporting ATPase 8, plasma membrane-type-like (LOC110799299), transcript variant X1, mRNA	spinach	36.2	36.2	81%	8.9	100.00%	3811
XM_021963057.1	PREDICTED: Prunus avium myosin-12 (LOC110760740), mRNA	sweet cherry	36.2	36.2	81%	8.9	100.00%	4996
XM_020557865.1	PREDICTED: Prunus persica myosin-12 (LOC18786033), transcript variant X3, mRNA	peach	36.2	36.2	81%	8.9	100.00%	5238
XM_020557864.1	PREDICTED: Prunus persica myosin-12 (LOC18786033), transcript variant X2, mRNA	peach	36.2	36.2	81%	8.9	100.00%	5161
XM_020557863.1	PREDICTED: Prunus persica myosin-12 (LOC18786033), transcript variant X1, mRNA	peach	36.2	36.2	81%	8.9	100.00%	5203
XM_008234624.2	PREDICTED: Prunus mume myosin-12-like (LOC103331946), partial mRNA	Japanese apricot	36.2	36.2	81%	8.9	100.00%	2283
XM_033550415.1	Arthroderma uncinatum phosphoserine aminotransferase (GIQ15_02998), partial mRNA	Arthroderma uncinatum	34.2	34.2	77%	35	100.00%	1335
XR_004496365.1	PREDICTED: Parus major EPM2A glucan phosphatase, laforin (EPM2A), transcript variant X4, misc_RNA	Great Tit	34.2	34.2	77%	35	100.00%	1729
XR_004496364.1	PREDICTED: Parus major EPM2A glucan phosphatase, laforin (EPM2A), transcript variant X3, misc_RNA	Great Tit	34.2	34.2	77%	35	100.00%	2012
XM_033518452.1	PREDICTED: Parus major ubiquitin specific peptidase 25 (USP25), transcript variant X3, mRNA	Great Tit	34.2	34.2	77%	35	100.00%	3744
XM_033518447.1	PREDICTED: Parus major ubiquitin specific peptidase 25 (USP25), transcript variant X2, mRNA	Great Tit	34.2	34.2	77%	35	100.00%	4346
XM_033518444.1	PREDICTED: Parus major ubiquitin specific peptidase 25 (USP25), transcript variant X1, mRNA	Great Tit	34.2	34.2	77%	35	100.00%	4556
XM_033052136.1	PREDICTED: Catharus ustulatus ubiquitin specific peptidase 25 (USP25), transcript variant X2, mRNA	Swainson's thrush	34.2	34.2	77%	35	100.00%	4629
XM_033052135.1	PREDICTED: Catharus ustulatus ubiquitin specific peptidase 25 (USP25), transcript variant X1, mRNA	Swainson's thrush	34.2	34.2	77%	35	100.00%	4419
XM_032989801.1	PREDICTED: Tyto alba alba ubiquitin specific peptidase 25 (USP25), partial mRNA	Tyto alba alba	34.2	34.2	77%	35	100.00%	2487
XM_030262676.2	PREDICTED: Taeniopygia guttata ubiquitin specific peptidase 25 (USP25), transcript variant X3, mRNA	zebra finch	34.2	34.2	77%	35	100.00%	4464
XM_030262598.2	PREDICTED: Taeniopygia guttata	zebra finch	34.2	34.2	77%	35	100.00%	3423

XM_030262523.2	ubiquitin specific peptidase 25 (USP25), transcript variant X2, mRNA PREDICTED: Taeniopygia guttata ubiquitin specific peptidase 25 (USP25), transcript variant X1, mRNA	zebra finch	34.2	34.2	77%	35	100.00%	4657
LR778263.1 1/2 2/2 3/3 8/1	Coregonus sp. 'balchen' genome assembly, chromosome: 11	Coregonus sp. 'balchen'	34.2	34.2	77%	35	100.00%	63177489
XM_032678829.1	PREDICTED: Chiroxiphia lanceolata ubiquitin specific peptidase 25 (USP25), transcript variant X3, mRNA	lance-tailed manakin	34.2	34.2	77%	35	100.00%	4730
XM_032678827.1	PREDICTED: Chiroxiphia lanceolata ubiquitin specific peptidase 25 (USP25), transcript variant X2, mRNA	lance-tailed manakin	34.2	34.2	77%	35	100.00%	4941
XM_032678826.1	PREDICTED: Chiroxiphia lanceolata ubiquitin specific peptidase 25 (USP25), transcript variant X1, mRNA	lance-tailed manakin	34.2	34.2	77%	35	100.00%	5240
AP022705.1	Plectropomus leopardus DNA, chromosome 6, nearly complete sequence	leopard coralgrouper	34.2	34.2	77%	35	100.00%	36725210
AP022683.1	Epinephelus fuscoguttatus DNA, LG9, complete sequence	brown-marbled grouper	34.2	64.4	86%	35	100.00%	46821238
XM_032448506.1	PREDICTED: Coturnix japonica ubiquitin specific peptidase 25 (USP25), transcript variant X5, mRNA	Japanese quail	34.2	34.2	77%	35	100.00%	4360
XM_032448505.1	PREDICTED: Coturnix japonica ubiquitin specific peptidase 25 (USP25), transcript variant X4, mRNA	Japanese quail	34.2	34.2	77%	35	100.00%	4612
XM_015880491.2	PREDICTED: Coturnix japonica ubiquitin specific peptidase 25 (USP25), transcript variant X3, mRNA	Japanese quail	34.2	34.2	77%	35	100.00%	4515
XM_015880498.2	PREDICTED: Coturnix japonica ubiquitin specific peptidase 25 (USP25), transcript variant X2, mRNA	Japanese quail	34.2	34.2	77%	35	100.00%	4854
XM_015880475.2	PREDICTED: Coturnix japonica ubiquitin specific peptidase 25 (USP25), transcript variant X1, mRNA	Japanese quail	34.2	34.2	77%	35	100.00%	4741
LR761639.1	Crassostrea gigas genome assembly, linkage group: LG6	Pacific oyster	34.2	34.2	77%	35	100.00%	60151564
XM_023227484.3 1/2	PREDICTED: Piliocolobus tephrosceles ubiquitin specific peptidase 25 (USP25), transcript variant X5, mRNA	Ugandan red Colobus	34.2	34.2	77%	35	100.00%	3371
XM_023227483.3 1/2	PREDICTED: Piliocolobus tephrosceles ubiquitin specific peptidase 25 (USP25), transcript variant X4, mRNA	Ugandan red Colobus	34.2	34.2	77%	35	100.00%	4933
XM_023227482.3 1/2	PREDICTED: Piliocolobus tephrosceles ubiquitin specific peptidase 25 (USP25), transcript variant X3, mRNA	Ugandan red Colobus	34.2	34.2	77%	35	100.00%	5124
XM_023227481.3 1/2	PREDICTED: Piliocolobus tephrosceles ubiquitin specific peptidase 25 (USP25), transcript variant X2, mRNA	Ugandan red Colobus	34.2	34.2	77%	35	100.00%	5133
XM_031061930.1	PREDICTED: Geospiza fortis ubiquitin specific peptidase 25 (USP25), transcript variant X2, mRNA	medium ground-finch	34.2	34.2	77%	35	100.00%	3732
XM_031061929.1	PREDICTED: Geospiza fortis ubiquitin specific peptidase 25 (USP25), transcript variant X1, mRNA	medium ground-finch	34.2	34.2	77%	35	100.00%	3635
XM_019017843.2	PREDICTED: Gorilla gorilla gorilla ubiquitin specific peptidase 25 (USP25), transcript variant X5, mRNA	western lowland gorilla	34.2	34.2	77%	35	100.00%	5059
XM_031005211.1	PREDICTED: Gorilla gorilla gorilla ubiquitin specific peptidase 25 (USP25), transcript variant X4, mRNA	western lowland gorilla	34.2	34.2	77%	35	100.00%	3803
XM_031005210.1	PREDICTED: Gorilla gorilla gorilla ubiquitin specific peptidase 25 (USP25), transcript variant X3, mRNA	western lowland gorilla	34.2	34.2	77%	35	100.00%	4430
XM_031005209.1	PREDICTED: Gorilla gorilla gorilla ubiquitin specific peptidase 25 (USP25), transcript variant X2, mRNA	western lowland gorilla	34.2	34.2	77%	35	100.00%	2845
XM_019017842.2	PREDICTED: Gorilla gorilla gorilla ubiquitin specific peptidase 25 (USP25), transcript variant X1, mRNA	western lowland gorilla	34.2	34.2	77%	35	100.00%	5094
CP007142.1	Gynuella sunshinyii YC6258, complete genome	Gynuella sunshinyii YC6258	34.2	34.2	77%	35	100.00%	6475959
XM_037444052.1	PREDICTED: Manduca sexta charged multivesicular body protein 4b (LOC115452512), mRNA	tobacco hornworm	34.2	34.2	77%	35	100.00%	3050

XM_037377838.1	PREDICTED: Falco rusticolus ubiquitin specific peptidase 25 (USP25), transcript variant X2, mRNA	gyrfalcon	34.2	34.2	77%	35	100.00%	4487
XM_037377837.1	PREDICTED: Falco rusticolus ubiquitin specific peptidase 25 (USP25), transcript variant X1, mRNA	gyrfalcon	34.2	34.2	77%	35	100.00%	4697
CP030038.1	Bradyrhizobium sp. CCBAU 51765 chromosome, complete genome	Bradyrhizobium sp. CCBAU 51765	34.2	34.2	77%	35	100.00%	8416216
LR884466.1	Acanthopagrus latus isolate v.2019 genome assembly, chromosome: 7	yellowfin seabream	34.2	34.2	77%	35	100.00%	31194941
XM_036381721.1	PREDICTED: Molothrus ater ubiquitin specific peptidase 25 (USP25), transcript variant X2, mRNA	Molothrus ater	34.2	34.2	77%	35	100.00%	3202
XM_036381714.1	PREDICTED: Molothrus ater ubiquitin specific peptidase 25 (USP25), transcript variant X1, mRNA	Molothrus ater	34.2	34.2	77%	35	100.00%	4646
XR_004971434.1	PREDICTED: Myotis myotis U6 spliceosomal RNA (LOC118667858), ncRNA	Myotis myotis	34.2	34.2	77%	35	100.00%	103
LR862367.1 1/2 2/2 3/2 3/3 3/5 6/1 7/2 8/1 10/1	***Pipistrellus pipistrellus genome assembly, chromosome: 11***	common pipistrelle	34.2	66.4	77%	35	100.00%	69386144
CP056560.1	Enterobacter cloacae strain RHBSTW-00399 chromosome, complete genome	Enterobacter cloacae	34.2	34.2	77%	35	100.00%	5120370
LR812059.1	Danio rerio genome assembly, chromosome: 22	zebrafish	34.2	64.4	95%	35	95.24%	39020267
LR812084.1	Danio rerio genome assembly, chromosome: 22	zebrafish	34.2	64.4	95%	35	95.24%	38893734
XM_033584384.1	Daldinia childiae Nonribosomal peptide synthetase 13 (GL218_07620), partial mRNA	Daldinia childiae	34.2	34.2	77%	35	100.00%	12396
CP048840.1	Crassostrea gigas strain QD chromosome 2	Pacific oyster	34.2	34.2	77%	35	100.00%	61781146
AP022713.1	Plectropomus leopardus DNA, chromosome 14, nearly complete sequence	leopard coralgrouper	34.2	64.4	86%	35	100.00%	33174000
CP047249.1	Aspergillus flavus strain SU-16 chromosome 1	Aspergillus flavus	34.2	34.2	77%	35	100.00%	6572574
CP036186.1 1/2 5/1 7/1 7/2	Pseudomonas stutzeri strain ATCC 14405 chromosome, complete genome	Pseudomonas stutzeri	34.2	66.4	77%	35	100.00%	4639098
CP062046.1	Macrobrachium nipponense isolate FS-2020 chromosome 35	Macrobrachium nipponense	34.2	34.2	77%	35	100.00%	67553775
LR884477.1	Acanthopagrus latus isolate v.2019 genome assembly, chromosome: 18	yellowfin seabream	34.2	34.2	77%	35	100.00%	31735906
CP060632.1	Lachnospiraceae bacterium NSJ-4 chromosome, complete genome	Lachnospiraceae bacterium NSJ-4	34.2	34.2	77%	35	100.00%	2942917
LR877218.1	**Acomys russatus genome assembly, chromosome: 7**	golden spiny mouse	34.2	34.2	77%	35	100.00%	69034057
LR877217.1	**Acomys russatus genome assembly, chromosome: 6**	golden spiny mouse	34.2	34.2	77%	35	100.00%	89442464
LS423640.1	Anas platyrhynchos genome assembly, chromosome: Z	mallard	34.2	96.6	77%	35	100.00%	74035964
LS423620.1	Anas platyrhynchos genome assembly, chromosome: 10	mallard	34.2	94.6	77%	35	100.00%	18704856
LR812615.1	Danio rerio genome assembly, chromosome: 22	zebrafish	34.2	34.2	95%	35	95.24%	34647438
LR812590.1	Danio rerio strain Nadia (NA) genome assembly, chromosome: 22	zebrafish	34.2	34.2	95%	35	95.24%	37117486
LR812565.1	Danio rerio strain Cooch Behar (CB) genome assembly, chromosome: 22	zebrafish	34.2	34.2	95%	35	95.24%	32929056
LR812510.1	Danio aesculapii genome assembly, chromosome: 9	Danio aesculapii	34.2	94.6	86%	35	100.00%	58433418

Pipistrellus pipistrellus genome assembly, chromosome: 11

GenBank: LR862367.1

FASTA Graphics

Go to:

LOCUS LR862367 17 bp DNA linear MAM **03-AUG-2020**

DEFINITION Pipistrellus pipistrellus genome assembly, chromosome: 11.

ACCESSION LR862367 REGION: 58336268..58336284

VERSION LR862367.1

DBLINK BioProject: PRJEB39564

BioSample: SAMEA994724

KEYWORDS .

SOURCE Pipistrellus pipistrellus (common pipistrelle)

ORGANISM Pipistrellus pipistrellus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Laurasiatheria; Chiroptera; Microchiroptera; Vespertilionidae; Pipistrellus.

REFERENCE 1

CONSRTM Wellcome Sanger Institute Data Sharing

TITLE Direct Submission

JOURNAL **Submitted (27-JUL-2020) SC, Wellcome Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, United Kingdom**

COMMENT The assembly mPipPip1.1 is based on 62x PacBio data, 34x 10X Genomics Chromium data, and 58x Arima Hi-C data generated at the Wellcome Sanger Institute. The Hi-C data was derived from a separate sample, mPipPip2. The assembly process included the following sequence of steps: initial PacBio assembly generation with Falcon-unzip, retained haplotig separation with purge_dups, 10X based scaffolding with scaff10x, Hi-C based scaffolding with SALSA2, Arrow polishing, two rounds of FreeBayes polishing. The mitochondrial assembly was produced at The Rockefeller University using mitoVGP. Finally, the assembly was analysed and manually improved using gEVAL. Chromosome-scale scaffolds confirmed by the Hi-C data have been named in order of size.

FEATURES Location/Qualifiers

source 1..17

/organism="Pipistrellus pipistrellus"

/mol_type="genomic DNA"

/db_xref="taxon:59474"

/chromosome="11"

ORIGIN

1 **gctttgccag tgcatat**

//

Pipistrellus pipistrellus genome assembly, chromosome: 11
Sequence ID: LR862367.1 **Length: 69386144 Number of Matches: 2**
Range 1: 58336268 to 58336284 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
34.2 bits(17)	35	17/17(100%)	0/17(0%)	Plus/Plus

```
Query  2         GCTTTGCCAGTGCATAT  18
                 ||||||||||||||||
Sbjct  58336268  GCTTTGCCAGTGCATAT  58336284
```

Range 2: 68321361 to 68321376 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
32.2 bits(16)	138	16/16(100%)	0/16(0%)	Plus/Plus

```
Query  3         CTTTGCCAGTGCATAT  18
                 ||||||||||||||||
Sbjct  68321361  CTTTGCCAGTGCATAT  68321376
```

ALIGNMENT VIEW FLAT QUERY-ANCHORED WITH LETTERS FOR IDENTITIES

Query range 1: 1 to 22

Query	1	CGCTTTGCCAGTGCATATTGTT	22
LR812067.1	33973414	CTTTGCCAGTGCATATTGTT	33973433
LR812067.1	28297982	CCAGTGCATATTGTT	28297968
NC_041308.1	12984	CTTTGCCAGTGCATATTGTT	13003
LR812598.1	31030181	CTTTGCCAGTGCATATTGTT	31030200
LR812598.1	25437388	CCAGTGCATATTGTT	25437374
BX005265.6	53448	CTTTGCCAGTGCATATTGTT	53467
LR877239.1	28766702	GCTTTGCCAGTGCATATTGT	28766683
XM_027028742.2	3437	GCTTTGCCAGTGCATATTGT	3418
LR606196.1	27753606	TTTGCCAGTGCATATTGTT	27753624
LR812527.1	34230208	TTTGCCAGTGCATATTGTT	34230226
LR812527.1	23818089	TTTGCTAGTGCATATTGTT	23818071
LR812527.1	70066863	TGCCAGTGCATATTGTT	70066847
LR535835.1	2054345	TTTGCCAGTGCATATTGTT	2054327
AP018449.1	3406253	GCTTTGCCAGTGCATATTG	3406235
CP062041.1	7234981	CGCTTTGCCAGTGCATATT	7234963
CP062041.1	9288585	TTTGCCAGTGCATAT	9288599
LR812525.1	30108445	TTTGCCAGTGCATATTGTT	30108427
LR812525.1	32027227	TTTGCCAGTGCATAT	32027241
AP022687.1	22176813	TTGCCAGTGCATATTGTT	22176830
AP022676.1	35934191	TTGCCAGTGCATATTGTT	35934208
AP022676.1	27074964	CGCTTTGCCAGTGCA	27074978
AP019298.1	42166	GCTTTGCCAGTGCATATT	42183
XM_002182111.1	1185	CTTTGCCAGTGCATATTG	1202
XM_001387892.1	1119	CGCTTTGCCAGTGCATAT	1136
AC110241.17	114257	CTTTGCCAGTGCATATTG	114274
AC113180.5	16294	CTTTGCCAGTGCATATTG	16311
XM_034347292.1	3973	GCTTTGCCAGTGCATATT	3956
XM_034347291.1	4056	GCTTTGCCAGTGCATATT	4039
XM_034347290.1	4101	GCTTTGCCAGTGCATATT	4084
LR792557.1	6627703	GCTTTGCCAGTGCATATT	6627686
CP029289.2	145139	GCTTTGCCAGTGCATATT	145122
CP027284.1	10571716	TTGCCAGTGCATATTGTT	10571699
CP027284.1	22398000	TTGCCAGTGCATATT	22397986
CP027284.1	24911851	GCTTTGCCAGTGCAT	24911837
XM_022004535.1	1131	TTTGCCAGTGCATATTGT	1114
XM_022004534.1	1251	TTTGCCAGTGCATATTGT	1234
XM_022004533.1	1238	TTTGCCAGTGCATATTGT	1221
XM_022004532.1	1650	TTTGCCAGTGCATATTGT	1633
XM_022004531.1	1658	TTTGCCAGTGCATATTGT	1641
XM_022004530.1	1749	TTTGCCAGTGCATATTGT	1732
XM_021963057.1	4114	GCTTTGCCAGTGCATATT	4097
XM_020557865.1	4331	GCTTTGCCAGTGCATATT	4314
XM_020557864.1	4254	GCTTTGCCAGTGCATATT	4237
XM_020557863.1	4296	GCTTTGCCAGTGCATATT	4279
XM_008234624.2	1317	GCTTTGCCAGTGCATATT	1300
XM_033550415.1	1296	CTTTGCCAGTGCATATT	1312
XR_004496365.1	1142	TGCCAGTGCATATTGTT	1158
XR_004496364.1	1425	TGCCAGTGCATATTGTT	1441
XM_033518452.1	1063	TGCCAGTGCATATTGTT	1079
XM_033518447.1	1875	TGCCAGTGCATATTGTT	1891

XM_033518444.1	1875	TGCCAGTGCATATTGTT	1891
XM_033052136.1	1961	TGCCAGTGCATATTGTT	1977
XM_033052135.1	1961	TGCCAGTGCATATTGTT	1977
XM_032989801.1	1977	TGCCAGTGCATATTGTT	1993
XM_030262676.2	1803	TGCCAGTGCATATTGTT	1819
XM_030262598.2	1994	TGCCAGTGCATATTGTT	2010
XM_030262523.2	1996	TGCCAGTGCATATTGTT	2012
LR778263.1	41253956	TTTGCCAGTGCATATTG	41253972
XM_032678829.1	1454	TGCCAGTGCATATTGTT	1470
XM_032678827.1	1875	TGCCAGTGCATATTGTT	1891
XM_032678826.1	1964	TGCCAGTGCATATTGTT	1980
AP022705.1	8528482	CTTTGCCAGTGCATATT	8528498
AP022683.1	11380536	CTTTGCCAGTGCATATT	11380552
AP022683.1	18222333	GCCAGTGCATATTGT	18222319
XM_032448506.1	1643	TGCCAGTGCATATTGTT	1659
XM_032448505.1	1895	TGCCAGTGCATATTGTT	1911
XM_015880491.2	2008	TGCCAGTGCATATTGTT	2024
XM_015880498.2	2137	TGCCAGTGCATATTGTT	2153
XM_015880475.2	2024	TGCCAGTGCATATTGTT	2040
LR761639.1	42529134	TTTGCCAGTGCATATTG	42529150
XM_023227484.3	2374	TGCCAGTGCATATTGTT	2390
XM_023227483.3	2087	TGCCAGTGCATATTGTT	2103
XM_023227482.3	2374	TGCCAGTGCATATTGTT	2390
XM_023227481.3	2287	TGCCAGTGCATATTGTT	2303
XM_031061930.1	1874	TGCCAGTGCATATTGTT	1890
XM_031061929.1	1987	TGCCAGTGCATATTGTT	2003
XM_019017843.2	2093	TGCCAGTGCATATTGTT	2109
XM_031005211.1	984	TGCCAGTGCATATTGTT	1000
XM_031005210.1	1611	TGCCAGTGCATATTGTT	1627
XM_031005209.1	2093	TGCCAGTGCATATTGTT	2109
XM_019017842.2	2338	TGCCAGTGCATATTGTT	2354
CP007142.1	3765470	TTTGCCAGTGCATATTG	3765486
XM_037444052.1	866	TTTGCCAGTGCATATTG	882
XM_037377838.1	2033	TGCCAGTGCATATTGTT	2049
XM_037377837.1	2033	TGCCAGTGCATATTGTT	2049
CP030038.1	5744144	TTTGCCAGTGCATATTG	5744160
LR884466.1	11115786	TGCCAGTGCATATTGTT	11115802
XM_036381721.1	1978	TGCCAGTGCATATTGTT	1994
XM_036381714.1	1980	TGCCAGTGCATATTGTT	1996
XR_004971434.1	5	GCTTTGCCAGTGCATAT	21
LR862367.1	58336268	GCTTTGCCAGTGCATAT	58336284
LR862367.1	68321361	CTTTGCCAGTGCATAT	68321376
CP056560.1	1890157	CTTTGCCAGTGCATATT	1890173
LR812059.1	37906794	GCTTTGCCAGTGCACATTGTT	37906814
LR812059.1	19399601	CCAGTGCATATTGTT	19399615
LR812084.1	38471293	GCTTTGCCAGTGCACATTGTT	38471273
LR812084.1	19094255	CCAGTGCATATTGTT	19094269
XM_033584384.1	3806	TTGCCAGTGCATATTGT	3790
CP048840.1	19217456	TTTGCCAGTGCATATTG	19217440
AP022713.1	18841814	TGCCAGTGCATATTGTT	18841798
AP022713.1	14182221	TTTGCCAGTGCATAT	14182207
CP047249.1	1107958	TTTGCCAGTGCATATTG	1107942
CP036186.1	2121509	CGCTTTGCCAGTGCATA	2121493
CP036186.1	2117588	CGCTTTGCCAGTGCAT	2117573
CP062046.1	16142310	TTTGCCAGTGCATATTG	16142294
LR884477.1	29489729	CTTTGCCAGTGCATATT	29489713
CP060632.1	1354514	CTTTGCCAGTGCATATT	1354498
LR877218.1	20569113	TGCCAGTGCATATTGTT	20569097
LR877217.1	14277986	CTTTGCCAGTGCATATT	14277970
LS423640.1	5542886	CTTTGCCAGTGCATATT	5542870

LS423640.1	21642490	TTTGCCAGTGCATATT	21642475
LS423640.1	71064731	TTGCCAGTGCATATT	71064745
LS423620.1	7862054	GCTTTGCCAGTGCATAT	7862038
LS423620.1	6985380	TTTGCCAGTGCATAT	6985394
LS423620.1	9009233	TTTGCCAGTGCATAT	9009219
LR812615.1	33837175	GCTTTGCCAGTGCACATTGTT	33837155
LR812590.1	36119604	GCTTTGCCAGTGCACATTGTT	36119584
LR812565.1	31103347	GCTTTGCCAGTGCACATTGTT	31103327
LR812510.1	3424092	TTTGCCAGTGCACATTGTT	3424110
LR812510.1	15241280	TTTGCCAGTGCATATTG	15241264
LR812510.1	28804007	CCAGTGCATATTGTT	28804021

PRODUCTS OF SIGNIFICANT ALIGNMENTS (CDS FEATURE)

- **NADH dehydrogenase subunit 4** GenBank
- **Putative beta-lactamase HcpC precursor** GenBank
- **Predicted protein** GenBank
- **Bud Emergence Mediator** GenBank
- **Myosin-12** GenBank GenBank GenBank GenBank GenBank GenBank GenBank GenBank
- **D-fructose 1,6-bisphosphatase** GenBank
- **Calcium-transporting ATPase 8, plasma membrane-type-like** GenBank GenBank GenBank GenBank GenBank GenBank
- **Phosphoserine aminotransferase** GenBank
- **Ubiquitin carboxyl-terminal hydrolase 25** GenBank
- **Peptidase** GenBank
- **DUF4406 domain-containing protein** GenBank
- **Nonribosomal peptide synthetase 13** GenBank
- **NAD-dependent DNA ligase LigA** GenBank GenBank
- **Dipicolinate synthase subunit B** GenBank

SEQUENCE N° 10/1

(T=2; C=1; G=3; A=0)

SEQUENCE N° 10/1 (T=2; C=1; G=3; A=0)

Original Sequence	**G**	**T**	**G**	**A**	**A**	**A**	**T**	**G**	**G**	**T**	**C**	**A**	**T**	**G**	**T**	**G**	**T**	**G**	**G**	**C**	**G**	**G**
Trend n° 10 (T=2; C=1; G=3; A=0)	**3**	**2**	**3**	**0**	**0**	**0**	**2**	**3**	**3**	**2**	**1**	**0**	**2**	**3**	**2**	**3**	**2**	**3**	**3**	**1**	**3**	**3**
Trend n° 10 (I i D d)	**-**	**D**	**I**	**D**	**d**	**d**	**I**	**I**	**i**	**D**	**D**	**D**	**I**	**I**	**D**	**I**	**D**	**I**	**i**	**D**	**I**	**i**
Trend n° 10/1	**2**	**1**	**2**	**1**	**1**	**1**	**2**	**3**	**3**	**2**	**1**	**0**	**1**	**2**	**1**	**2**	**1**	**2**	**2**	**0**	**2**	**2**
Sequence n° 10/1 (72,73%)	**T**	**C**	**T**	**C**	**C**	**C**	**T**	**G**	**G**	**T**	**C**	**A**	**C**	**T**	**C**	**T**	**C**	**T**	**T**	**A**	**T**	**T**

Query: **TCTCCCTGGTCACTCTCTTATT**

Search Parameters	
Program	blastn
Word size	7
Expect value	1000
Hitlist size	100
Match/Mismatch scores	1,-3
Gapcosts	5,2
Filter string	F
Genetic Code	1

98 sequences selected:

Accession Seq	Description	Comomm Name	Max Score	Total Score	Query C.	E value	Per. Id.	Acc. Len
XM_010904040.3	PREDICTED: Esox lucius SLAM family member 5-like (LOC105030290), mRNA	northern pike	38.2	38.2	86%	2.2	100.00%	909
XM_010904045.3	PREDICTED: Esox lucius CD48 antigen-like (LOC105030295), mRNA	northern pike	38.2	38.2	86%	2.2	100.00%	975
AL929563.15	Mouse DNA sequence from clone RP23-68O14 on chromosome 2, complete sequence	house mouse	38.2	38.2	86%	2.2	100.00%	113198
LR594566.1	Streptopelia turtur genome assembly, chromosome: 20	Streptopelia turtur	36.2	36.2	81%	8.9	100.00%	16376599
LR535827.1	Denticeps clupeoides genome assembly, chromosome: 15	denticle herring	36.2	66.4	81%	8.9	100.00%	21430986
NG_031034.1	Ornithorhynchus anatinus vomeronasal 1 receptor ornAnaV1R-ps3357 pseudogene (ORNANAV1R-PS3357)	platypus	36.2	36.2	81%	8.9	100.00%	1126
NG_031181.1	Ornithorhynchus anatinus vomeronasal 1 receptor ornAnaV1R-ps3504 pseudogene (ORNANAV1R-PS3504)	platypus	36.2	36.2	81%	8.9	100.00%	1038
NG_031677.1	Ornithorhynchus anatinus vomeronasal 1 receptor ornAnaV1R-ps4000 pseudogene (ORNANAV1R-PS4000)	platypus	36.2	36.2	81%	8.9	100.00%	845
NG_031675.1	Ornithorhynchus anatinus vomeronasal 1 receptor ornAnaV1R-ps3998 pseudogene (ORNANAV1R-PS3998)	platypus	36.2	36.2	81%	8.9	100.00%	717
NG_030926.1	Ornithorhynchus anatinus vomeronasal 1 receptor ornAnaV1R-ps3247 pseudogene (ORNANAV1R-PS3247)	platypus	36.2	36.2	81%	8.9	100.00%	1110
NG_030845.1	Ornithorhynchus anatinus vomeronasal 1 receptor ornAnaV1R-ps3166 pseudogene (ORNANAV1R-PS3166)	platypus	36.2	36.2	81%	8.9	100.00%	1123
LR880651.1	Poecilia reticulata genome assembly, chromosome: 7	guppy	36.2	66.4	81%	8.9	100.00%	33249287
AC130281.13	Mus musculus chromosome 3, clone RP23-372I17, complete sequence	house mouse	36.2	36.2	81%	8.9	100.00%	216261
AC101874.8	Mus musculus chromosome 7, clone RP23-366A20, complete sequence	house mouse	36.2	36.2	81%	8.9	100.00%	179708

CP050606.1 1/2 2/2 3/3 3/4 3/5 3/6 5/1	Canis lupus familiaris breed Labrador retriever chromosome 07a	dog	36.2	126	86%	8.9	100.00%	80755643
CP050603.1 1/2 2/2 3/3 3/4 3/5 3/6 5/1	Canis lupus familiaris breed Labrador retriever chromosome 17a	dog	36.2	68.4	86%	8.9	100.00%	64150251
CP050637.1 1/2 2/2 3/3 3/4 3/5 3/6 5/1	Canis lupus familiaris breed Labrador retriever chromosome 07b	dog	36.2	126	86%	8.9	100.00%	80755819
CP050614.1 1/2 2/2 3/3 3/4 3/5 3/6 5/1	Canis lupus familiaris breed Labrador retriever chromosome 17b	dog	36.2	68.4	86%	8.9	100.00%	64169108
XM_032568351.1	PREDICTED: Xiphophorus hellerii titin-like (LOC116723448), mRNA	green swordtail	36.2	36.2	81%	8.9	100.00%	91177
XM_030069654.1	PREDICTED: Myripristis murdjan coiled-coil alpha-helical rod protein 1-like (LOC115371980), transcript variant X2, mRNA	pinecone soldierfish	36.2	36.2	81%	8.9	100.00%	3464
XM_030069653.1	PREDICTED: Myripristis murdjan coiled-coil alpha-helical rod protein 1-like (LOC115371980), transcript variant X1, mRNA	pinecone soldierfish	36.2	36.2	81%	8.9	100.00%	3470
LR594558.1	Streptopelia turtur genome assembly, chromosome: 6	Streptopelia turtur	36.2	36.2	81%	8.9	100.00%	40337708
LR594555.1	Streptopelia turtur genome assembly, chromosome: Z	Streptopelia turtur	36.2	66.4	95%	8.9	100.00%	73908008
XM_028955184.1	PREDICTED: Denticeps clupeoides minichromosome maintenance 10 replication initiation factor (mcm10), transcript variant X3, mRNA	denticle herring	36.2	36.2	81%	8.9	100.00%	2755
XM_028955183.1	PREDICTED: Denticeps clupeoides minichromosome maintenance 10 replication initiation factor (mcm10), transcript variant X2, mRNA	denticle herring	36.2	36.2	81%	8.9	100.00%	2837
XM_028955182.1	PREDICTED: Denticeps clupeoides minichromosome maintenance 10 replication initiation factor (mcm10), transcript variant X1, mRNA	denticle herring	36.2	36.2	81%	8.9	100.00%	2840
XM_028024223.1	PREDICTED: Xiphophorus couchianus titin-like (LOC114148751), mRNA	Monterrey platyfish	36.2	36.2	81%	8.9	100.00%	91226
CP030989.1	Arachis hypogaea cultivar Shitouqi chromosome A07	peanut	36.2	36.2	81%	8.9	100.00%	83117376
XM_023337148.1	PREDICTED: Xiphophorus maculatus titin-like (LOC102219802), mRNA	southern platyfish	36.2	36.2	81%	8.9	100.00%	83999
XM_022234773.1	PREDICTED: Acanthaster planci basement membrane-specific heparan sulfate proteoglycan core protein-like (LOC110979181), transcript variant X6, mRNA	crown-of-thorns starfish	36.2	36.2	81%	8.9	100.00%	11632
XM_022234772.1	PREDICTED: Acanthaster planci basement membrane-specific heparan sulfate proteoglycan core protein-like (LOC110979181), transcript variant X5, mRNA	crown-of-thorns starfish	36.2	36.2	81%	8.9	100.00%	11888
XM_022234771.1	PREDICTED: Acanthaster planci basement membrane-specific heparan sulfate proteoglycan core protein-like (LOC110979181), transcript variant X4, mRNA	crown-of-thorns starfish	36.2	36.2	81%	8.9	100.00%	11891

XM_022234770.1	PREDICTED: Acanthaster planci basement membrane-specific heparan sulfate proteoglycan core protein-like (LOC110979181), transcript variant X3, mRNA	crown-of-thorns starfish	36.2	36.2	81%	8.9	100.00%	12903
XM_022234769.1	PREDICTED: Acanthaster planci basement membrane-specific heparan sulfate proteoglycan core protein-like (LOC110979181), transcript variant X2, mRNA	crown-of-thorns starfish	36.2	36.2	81%	8.9	100.00%	13155
XM_022234768.1	PREDICTED: Acanthaster planci basement membrane-specific heparan sulfate proteoglycan core protein-like (LOC110979181), transcript variant X1, mRNA	crown-of-thorns starfish	36.2	36.2	81%	8.9	100.00%	13158
XM_017309405.1	PREDICTED: Poecilia reticulata titin (LOC103461835), mRNA	guppy	36.2	36.2	81%	8.9	100.00%	93950
XM_016662277.1	PREDICTED: Poecilia formosa titin (ttn), mRNA	Amazon molly	36.2	36.2	81%	8.9	100.00%	103364
XM_015391676.1	PREDICTED: Cyprinodon variegatus titin (ttn), mRNA	sheepshead minnow	36.2	36.2	81%	8.9	100.00%	95566
XM_015043247.1	PREDICTED: Poecilia latipinna titin (ttn), transcript variant X18, mRNA	sailfin molly	36.2	36.2	81%	8.9	100.00%	88892
XM_015043246.1	PREDICTED: Poecilia latipinna titin (ttn), transcript variant X17, mRNA	sailfin molly	36.2	36.2	81%	8.9	100.00%	89735
XM_015043245.1	PREDICTED: Poecilia latipinna titin (ttn), transcript variant X16, mRNA	sailfin molly	36.2	36.2	81%	8.9	100.00%	91521
XM_015043244.1	PREDICTED: Poecilia latipinna titin (ttn), transcript variant X15, mRNA	sailfin molly	36.2	36.2	81%	8.9	100.00%	91535
XM_015043243.1	PREDICTED: Poecilia latipinna titin (ttn), transcript variant X14, mRNA	sailfin molly	36.2	36.2	81%	8.9	100.00%	91797
XM_015043241.1	PREDICTED: Poecilia latipinna titin (ttn), transcript variant X13, mRNA	sailfin molly	36.2	36.2	81%	8.9	100.00%	92154
XM_015043240.1	PREDICTED: Poecilia latipinna titin (ttn), transcript variant X12, mRNA	sailfin molly	36.2	36.2	81%	8.9	100.00%	92691
XM_015043239.1	PREDICTED: Poecilia latipinna titin (ttn), transcript variant X11, mRNA	sailfin molly	36.2	36.2	81%	8.9	100.00%	93168
XM_015043238.1	PREDICTED: Poecilia latipinna titin (ttn), transcript variant X10, mRNA	sailfin molly	36.2	36.2	81%	8.9	100.00%	93444
XM_015043237.1	PREDICTED: Poecilia latipinna titin (ttn), transcript variant X9, mRNA	sailfin molly	36.2	36.2	81%	8.9	100.00%	93714
XM_015043236.1	PREDICTED: Poecilia latipinna titin (ttn), transcript variant X8, mRNA	sailfin molly	36.2	36.2	81%	8.9	100.00%	93801
XM_015043235.1	PREDICTED: Poecilia latipinna titin (ttn), transcript variant X7, mRNA	sailfin molly	36.2	36.2	81%	8.9	100.00%	94494
XM_015043234.1	PREDICTED: Poecilia latipinna titin (ttn), transcript variant X6, mRNA	sailfin molly	36.2	36.2	81%	8.9	100.00%	94614
XM_015043233.1	PREDICTED: Poecilia latipinna titin (ttn), transcript variant X5, mRNA	sailfin molly	36.2	36.2	81%	8.9	100.00%	94758
XM_015043232.1	PREDICTED: Poecilia latipinna titin (ttn), transcript variant X4, mRNA	sailfin molly	36.2	36.2	81%	8.9	100.00%	94782
XM_015043231.1	PREDICTED: Poecilia latipinna titin (ttn), transcript variant X3, mRNA	sailfin molly	36.2	36.2	81%	8.9	100.00%	94848
XM_015043230.1	PREDICTED: Poecilia latipinna titin (ttn), transcript variant X2, mRNA	sailfin molly	36.2	36.2	81%	8.9	100.00%	95256
XM_015043229.1	PREDICTED: Poecilia latipinna titin (ttn), transcript variant X1, mRNA	sailfin molly	36.2	36.2	81%	8.9	100.00%	95361
XM_014984134.1	PREDICTED: Poecilia mexicana titin (ttn), mRNA	Poecilia mexicana	36.2	36.2	81%	8.9	100.00%	103334
CP011894.1	Ovis canadensis canadensis isolate 43U chromosome 9	Ovis canadensis canadensis	36.2	161	100%	8.9	100.00%	94836416

LR880646.1	sequence Poecilia reticulata genome assembly, chromosome: 2	guppy	36.2	36.2	81%	8.9	100.00%	45373629
XM_036132592.1	PREDICTED: Fundulus heteroclitus titin, tandem duplicate 2 (ttn.2), mRNA	mummichog	36.2	36.2	81%	8.9	100.00%	96129
LS423614.1	Anas platyrhynchos genome assembly, chromosome: 4	mallard	36.2	66.4	86%	8.9	100.00%	74522197
AC158296.6	Mus musculus chromosome 7, clone RP24-78M9, complete sequence	house mouse	36.2	36.2	81%	8.9	100.00%	201897
XR_002941147.1	PREDICTED: Pan troglodytes phosphatidylinositol 4-kinase alpha (PI4KA), transcript variant X7, misc_RNA	chimpanzee	34.2	34.2	77%	35	100.00%	6434
XR_002941146.1	PREDICTED: Pan troglodytes phosphatidylinositol 4-kinase alpha (PI4KA), transcript variant X6, misc_RNA	chimpanzee	34.2	34.2	77%	35	100.00%	6665
XM_024352563.1	PREDICTED: Pan troglodytes phosphatidylinositol 4-kinase alpha (PI4KA), transcript variant X5, mRNA	chimpanzee	34.2	34.2	77%	35	100.00%	6713
XM_024352562.1	PREDICTED: Pan troglodytes phosphatidylinositol 4-kinase alpha (PI4KA), transcript variant X4, mRNA	chimpanzee	34.2	34.2	77%	35	100.00%	6720
XM_024352561.1	PREDICTED: Pan troglodytes phosphatidylinositol 4-kinase alpha (PI4KA), transcript variant X3, mRNA	chimpanzee	34.2	34.2	77%	35	100.00%	6735
XR_002941145.1	PREDICTED: Pan troglodytes phosphatidylinositol 4-kinase alpha (PI4KA), transcript variant X2, misc_RNA	chimpanzee	34.2	34.2	77%	35	100.00%	6873
XM_024352560.1	PREDICTED: Pan troglodytes phosphatidylinositol 4-kinase alpha (PI4KA), transcript variant X1, mRNA	chimpanzee	34.2	34.2	77%	35	100.00%	6786
XM_024353570.1	PREDICTED: Pan troglodytes putative phosphatidylinositol 4-kinase alpha-like protein P2 (LOC458675), mRNA	chimpanzee	34.2	34.2	77%	35	100.00%	3285
CP027083.1 **1/2 3/6 7/2**	Bos mutus isolate yakQH1 chromosome 15	wild yak	34.2	94.6	81%	35	100.00%	79990276
XM_024240188.1	PREDICTED: Pongo abelii phosphatidylinositol 4-kinase alpha (PI4KA), transcript variant X4, mRNA	Sumatran orangutan	34.2	34.2	77%	35	100.00%	6754
XM_024240187.1	PREDICTED: Pongo abelii phosphatidylinositol 4-kinase alpha (PI4KA), transcript variant X3, mRNA	Sumatran orangutan	34.2	34.2	77%	35	100.00%	6800
XR_002912889.1	PREDICTED: Pongo abelii phosphatidylinositol 4-kinase alpha (PI4KA), transcript variant X2, misc_RNA	Sumatran orangutan	34.2	34.2	77%	35	100.00%	7240
XR_002912888.1	PREDICTED: Pongo abelii phosphatidylinositol 4-kinase alpha (PI4KA), transcript variant X1, misc_RNA	Sumatran orangutan	34.2	34.2	77%	35	100.00%	7285
XM_024010326.1	PREDICTED: Salvelinus alpinus **zinc** finger protein 420 (LOC111979690), transcript variant X2, mRNA	Arctic char	34.2	34.2	77%	35	100.00%	3488
XM_024010325.1	PREDICTED: Salvelinus alpinus **zinc** finger protein 420 (LOC111979690), transcript variant X1, mRNA	Arctic char	34.2	34.2	77%	35	100.00%	3491
XM_012466500.1	PREDICTED: Aotus nancymaae phosphatidylinositol 4-kinase alpha (PI4KA), mRNA	Ma's night monkey	34.2	34.2	77%	35	100.00%	6754
AC236497.1	Xenopus laevis clone CH219-212B10, complete sequence	African clawed frog	34.2	34.2	77%	35	100.00%	140175
AP023477.1 **1/2 2/2 3/2 3/3 5/1 6/1**	Homo sapiens DNA, chromosome 17, nearly complete genome	human	34.2	64.4	77%	35	100.00%	80688777
NM_001362862.2 **1/2 2/2 3/2 3/3 5/1**	Homo sapiens phosphatidylinositol 4-kinase	human	34.2	34.2	77%	35	100.00%	6658

6/1	alpha (PI4KA), transcript variant 2, mRNA							
NM_001362863.2 1/2 2/2 3/2 3/3 5/1 6/1	Homo sapiens phosphatidylinositol 4-kinase alpha (PI4KA), transcript variant 3, mRNA	human	34.2	34.2	77%	35	100.00%	6685
FR799580.1	Leishmania mexicana MHOM/GT/2001/U1103 complete genome, chromosome 27	Leishmania mexicana MHOM/GT/2001/U1103	34.2	34.2	77%	35	100.00%	1044075
XM_024464007.1	Pseudogymnoascus destructans hypothetical protein (VC83_00313), mRNA	Pseudogymnoascus destructans	34.2	34.2	95%	35	95.24%	870
CP013203.1	Paenibacillus sp. IHB B 3084, complete genome	Paenibacillus sp. IHB B 3084	34.2	34.2	77%	35	100.00%	5508564
XM_012733809.1	PREDICTED: Condylura cristata DEAH (Asp-Glu-Ala-His) box polypeptide 34 (DHX34), transcript variant X2, mRNA	star-nosed mole	34.2	34.2	77%	35	100.00%	3971
XM_004694238.2	PREDICTED: Condylura cristata DEAH (Asp-Glu-Ala-His) box polypeptide 34 (DHX34), transcript variant X1, mRNA	star-nosed mole	34.2	34.2	77%	35	100.00%	3963
CP055242.1 1/2 2/2 3/2 3/3 3/5 6/1 8/1 9/1	Solanum tuberosum cultivar Solyntus chromosome 9	potato	34.2	34.2	95%	35	95.24%	70978473
AP023482.1 1/2 2/2 3/2 3/3 5/1 6/1	Homo sapiens DNA, chromosome 22, nearly complete genome	human	34.2	126	86%	35	100.00%	46684173
AP023479.1 1/2 2/2 3/2 3/3 5/1 6/1	Homo sapiens DNA, chromosome 19, nearly complete genome	human	34.2	34.2	95%	35	95.24%	59105444
LR877229.1	**Acomys russatus genome assembly, chromosome: 18**	golden spiny mouse	34.2	94.6	90%	35	100.00%	64120505
LR877198.1	Onychomys torridus genome assembly, chromosome: 11	southern grasshopper mouse	34.2	124	86%	35	100.00%	93020901
LR862368.1 1/2 2/2 3/2 3/3 3/5 6/1 7/2 8/1 9/1	***Pipistrellus pipistrellus genome assembly, chromosome: 12***	common pipistrelle	34.2	128	100%	35	100.00%	58291996
LR862361.1 1/2 2/2 3/2 3/3 3/5 6/1 7/2 8/1 9/1	***Pipistrellus pipistrellus genome assembly, chromosome: 5***	common pipistrelle	34.2	189	100%	35	100.00%	94929986
AC279687.1 1/2 2/2 3/2 3/3 5/1 6/1	Homo sapiens chromosome unknown clone VMRC59-495H23, complete sequence	human	34.2	34.2	77%	35	100.00%	169742
CP002352.1	Bacteroides helcogenes P 36-108, complete genome	Bacteroides helcogenes P 36-108	34.2	34.2	77%	35	100.00%	3998906
AP023475.1 1/2 2/2 3/2 3/3 5/1 6/1	Homo sapiens DNA, chromosome 15, nearly complete genome	human	32.2	92.7	90%	139	95.00%	95537968
AP023476.1 1/2 2/2 3/2 3/3 5/1 6/1	Homo sapiens DNA, chromosome 16, nearly complete genome	human	30.2	90.7	90%	547	100.00%	94690957

Pipistrellus pipistrellus genome assembly, chromosome: 12

GenBank: LR862368.1

FASTA Graphics

Go to:

LOCUS LR862368 58291996 bp DNA linear MAM **03-AUG-2020**

DEFINITION Pipistrellus pipistrellus genome assembly, chromosome: 12.

ACCESSION LR862368

VERSION LR862368.1

DBLINK BioProject: PRJEB39564
BioSample: SAMEA994724

KEYWORDS .

SOURCE Pipistrellus pipistrellus (common pipistrelle)

ORGANISM Pipistrellus pipistrellus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Laurasiatheria; Chiroptera; Microchiroptera; Vespertilionidae; Pipistrellus.

REFERENCE 1

CONSRTM Wellcome Sanger Institute Data Sharing

TITLE Direct Submission

JOURNAL **Submitted (27-JUL-2020) SC, Wellcome Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, United Kingdom**

COMMENT The assembly mPipPip1.1 is based on 62x PacBio data, 34x 10X Genomics Chromium data, and 58x Arima Hi-C data generated at the Wellcome Sanger Institute. The Hi-C data was derived from a separate sample, mPipPip2. The assembly process included the following sequence of steps: initial PacBio assembly generation with Falcon-unzip, retained haplotig separation with purge_dups, 10X based scaffolding with scaff10x, Hi-C based scaffolding with SALSA2, Arrow polishing, two rounds of FreeBayes polishing. The mitochondrial assembly was produced at The Rockefeller University using mitoVGP. Finally, the assembly was analysed and manually improved using gEVAL. Chromosome-scale scaffolds confirmed by the Hi-C data have been named in order of size.

Pipistrellus pipistrellus genome assembly, chromosome: 12
Sequence ID: LR862368.1 **Length: 58291996 Number of Matches: 4**

Range 1: 46682084 to 46682100 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
34.2 bits(17)	35	17/17(100%)	0/17(0%)	Plus/Minus

```
Query  1          TCTCCCTGGTCACTCTC  17
                  |||||||||||||||||
Sbjct  46682100   TCTCCCTGGTCACTCTC  46682084
```

Range 2: 30677988 to 30678003 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
32.2 bits(16)	139	16/16(100%)	0/16(0%)	Plus/Minus

```
Query  3          TCCCTGGTCACTCTCT  18
                  ||||||||||||||||
Sbjct  30678003   TCCCTGGTCACTCTCT  30677988
```

Range 3: 30728449 to 30728468 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
32.2 bits(16)	139	19/20(95%)	0/20(0%)	Plus/Minus

```
Query  3          TCCCTGGTCACTCTCTTATT  22
                  |||||||||||||| |||||
Sbjct  30728468   TCCCTGGTCACTCTATTATT  30728449
```

Range 4: 32170018 to 32170032 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
30.2 bits(15)	547	15/15(100%)	0/15(0%)	Plus/Minus

```
Query  5          CCTGGTCACTCTCTT  19
                  |||||||||||||||
Sbjct  32170032   CCTGGTCACTCTCTT  32170018
```

Pipistrellus pipistrellus genome assembly, chromosome: 5

GenBank: LR862361.1

FASTA Graphics

Go to:

LOCUS LR862361 94929986 bp DNA linear MAM **03-AUG-2020**

DEFINITION Pipistrellus pipistrellus genome assembly, chromosome: 5.

ACCESSION LR862361

VERSION LR862361.1

DBLINK BioProject: PRJEB39564
BioSample: SAMEA994724

KEYWORDS .

SOURCE Pipistrellus pipistrellus (common pipistrelle)

ORGANISM Pipistrellus pipistrellus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Laurasiatheria; Chiroptera; Microchiroptera; Vespertilionidae; Pipistrellus.

REFERENCE 1

CONSRTM Wellcome Sanger Institute Data Sharing

TITLE Direct Submission

JOURNAL **Submitted (27-JUL-2020) SC, Wellcome Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, United Kingdom**

COMMENT The assembly mPipPip1.1 is based on 62x PacBio data, 34x 10X Genomics Chromium data, and 58x Arima Hi-C data generated at the Wellcome Sanger Institute. The Hi-C data was derived from a separate sample, mPipPip2. The assembly process included the following sequence of steps: initial PacBio assembly generation with Falcon-unzip, retained haplotig separation with purge_dups, 10X based scaffolding with scaff10x, Hi-C based scaffolding with SALSA2, Arrow polishing, two rounds of FreeBayes polishing. The mitochondrial assembly was produced at The Rockefeller University using mitoVGP. Finally, the assembly was analysed and manually improved using gEVAL. Chromosome-scale scaffolds confirmed by the Hi-C data have been named in order of size.

FEATURES Location/Qualifiers

source 1..94929986
/organism="Pipistrellus pipistrellus"
/mol_type="genomic DNA"
/db_xref="taxon:59474"
/chromosome="5"

ORIGIN

//

Pipistrellus pipistrellus genome assembly, chromosome: 5
Sequence ID: LR862361.1 **Length: 94929986 Number of Matches: 6**
Range 1: 6633772 to 6633788 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
34.2 bits(17)	35	17/17(100%)	0/17(0%)	Plus/Minus

```
Query  6         CTGGTCACTCTCTTATT  22
                 |||||||||||||||||
Sbjct  6633788   CTGGTCACTCTCTTATT  6633772
```

Range 2: 60936027 to 60936043 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
34.2 bits(17)	35	17/17(100%)	0/17(0%)	Plus/Minus

```
Query  1          TCTCCCTGGTCACTCTC  17
                  |||||||||||||||||
Sbjct  60936043   TCTCCCTGGTCACTCTC  60936027
```

Range 3: 22931834 to 22931848 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
30.2 bits(15)	547	15/15(100%)	0/15(0%)	Plus/Plus

```
Query  2          CTCCCTGGTCACTCT  16
                  |||||||||||||||
Sbjct  22931834   CTCCCTGGTCACTCT  22931848
```

Range 4: 23032086 to 23032100 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
30.2 bits(15)	547	15/15(100%)	0/15(0%)	Plus/Plus

```
Query  2          CTCCCTGGTCACTCT  16
                  |||||||||||||||
Sbjct  23032086   CTCCCTGGTCACTCT  23032100
```

Range 5: 13011174 to 13011192 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
30.2 bits(15)	547	18/19(95%)	0/19(0%)	Plus/Minus

```
Query  1         TCTCCCTGGTCACTCTCTT  19
                 ||||||||||| |||||||
Sbjct  13011192  TCTCCCTGGTCTCTCTCTT  13011174
```

Range 6: 50687507 to 50687521 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
30.2 bits(15)	547	15/15(100%)	0/15(0%)	Plus/Minus

```
Query  1         TCTCCCTGGTCACTC  15
                 |||||||||||||||
Sbjct  50687521  TCTCCCTGGTCACTC  50687507
```

ALIGNMENT VIEW FLAT QUERY-ANCHORED WITH LETTERS FOR IDENTITIES

Query range 1: 1 to 22

Query	**1**	**TCTCCCTGGTCACTCTCTTATT**	**22**
XM_010904040.3	747	TCTCCCTGGTCACTCTCTT	729
XM_010904045.3	788	TCTCCCTGGTCACTCTCTT	770
AL929563.15	92822	CCCTGGTCACTCTCTTATT	92804
LR594566.1	345847	CCTGGTCACTCTCTTATT	345864
LR535827.1	9524453	TCTCCCTGGTCACTCTCT	9524470
LR535827.1	403307	TCTCCCTGGTCACTC	403321
NG_031034.1	785	TCTCCCTGGTCACTCTCT	802
NG_031181.1	697	TCTCCCTGGTCACTCTCT	714
NG_031677.1	637	TCTCCCTGGTCACTCTCT	654
NG_031675.1	637	TCTCCCTGGTCACTCTCT	654
NG_030926.1	769	TCTCCCTGGTCACTCTCT	786
NG_030845.1	782	TCTCCCTGGTCACTCTCT	799
LR880651.1	18341798	TCTCCCTGGTCACTCTCT	18341815
LR880651.1	26543910	TCTCCCTGGTCACTC	26543896
AC130281.13	83242	CCCTGGTCACTCTCTTAT	83259
AC101874.8	21441	TCCCTGGTCACTCTCTTA	21458
CP050606.1	67402414	CTCCCTGGTCACTCTCTT	67402397
CP050606.1	12929274	CCCTGGTCACTCTCT	12929288
CP050606.1	26349503	CTCCCTGGTCACTCT	26349517
CP050606.1	57094204	CTGGTCACTCTCTTA	57094190
CP050603.1	58938169	CCTGGTCACTCTCTTATT	58938152
CP050603.1	8696110	CCCTGGTCACTCTCTT	8696095
CP050637.1	67398811	CTCCCTGGTCACTCTCTT	67398794
CP050637.1	12932850	CCCTGGTCACTCTCT	12932864

CP050637.1	26351875	CTCCCTGGTCACTCT	26351889
CP050637.1	57087982	CTGGTCACTCTCTTA	57087968
CP050614.1	58962803	CCTGGTCACTCTCTTATT	58962786
CP050614.1	8705602	CCCTGGTCACTCTCTT	8705587
XM_032568351.1	89446	TCTCCCTGGTCACTCTCT	89429
XM_030069654.1	1793	TCTCCCTGGTCACTCTCT	1776
XM_030069653.1	1799	TCTCCCTGGTCACTCTCT	1782
LR594558.1	40292066	CCTGGTCACTCTCTTATT	40292049
LR594555.1	52712341	CCTGGTCACTCTCTTATT	52712324
LR594555.1	54643947	CTCCCTGGTCACTCT	54643961
XM_028955184.1	926	TCTCCCTGGTCACTCTCT	909
XM_028955183.1	923	TCTCCCTGGTCACTCTCT	906
XM_028955182.1	926	TCTCCCTGGTCACTCTCT	909
XM_028024223.1	89495	TCTCCCTGGTCACTCTCT	89478
CP030989.1	57446801	TCCCTGGTCACTCTCTTA	57446784
XM_023337148.1	82268	TCTCCCTGGTCACTCTCT	82251
XM_022234773.1	2170	CTCCCTGGTCACTCTCTT	2153
XM_022234772.1	2426	CTCCCTGGTCACTCTCTT	2409
XM_022234771.1	2429	CTCCCTGGTCACTCTCTT	2412
XM_022234770.1	3441	CTCCCTGGTCACTCTCTT	3424
XM_022234769.1	3693	CTCCCTGGTCACTCTCTT	3676
XM_022234768.1	3696	CTCCCTGGTCACTCTCTT	3679
XM_017309405.1	92231	TCTCCCTGGTCACTCTCT	92214
XM_016662277.1	101644	TCTCCCTGGTCACTCTCT	101627
XM_015391676.1	93812	TCTCCCTGGTCACTCTCT	93795
XM_015043247.1	87178	TCTCCCTGGTCACTCTCT	87161
XM_015043246.1	88021	TCTCCCTGGTCACTCTCT	88004
XM_015043245.1	89807	TCTCCCTGGTCACTCTCT	89790
XM_015043244.1	89821	TCTCCCTGGTCACTCTCT	89804
XM_015043243.1	90083	TCTCCCTGGTCACTCTCT	90066
XM_015043241.1	90440	TCTCCCTGGTCACTCTCT	90423
XM_015043240.1	90977	TCTCCCTGGTCACTCTCT	90960
XM_015043239.1	91454	TCTCCCTGGTCACTCTCT	91437
XM_015043238.1	91730	TCTCCCTGGTCACTCTCT	91713
XM_015043237.1	92000	TCTCCCTGGTCACTCTCT	91983
XM_015043236.1	92087	TCTCCCTGGTCACTCTCT	92070
XM_015043235.1	92780	TCTCCCTGGTCACTCTCT	92763
XM_015043234.1	92900	TCTCCCTGGTCACTCTCT	92883
XM_015043233.1	93044	TCTCCCTGGTCACTCTCT	93027
XM_015043232.1	93068	TCTCCCTGGTCACTCTCT	93051
XM_015043231.1	93134	TCTCCCTGGTCACTCTCT	93117
XM_015043230.1	93542	TCTCCCTGGTCACTCTCT	93525
XM_015043229.1	93647	TCTCCCTGGTCACTCTCT	93630
XM_014984134.1	101620	TCTCCCTGGTCACTCTCT	101603
CP011894.1	26322777	TCCCTGGTCACTCACTTATT	26322796
CP011894.1	73003361	TCTCCCTGGTCACTCTCT	73003344
CP011894.1	74063983	CCCTGGTCACTCTCTT	74063968
CP011894.1	73665434	CTCCCTGGTCACTCT	73665448
CP011894.1	13298756	CCCTGGTCACTCTCT	13298742
LR880646.1	41540390	TCTCCCTGGTCACTCTCT	41540373
XM_036132592.1	94364	TCTCCCTGGTCACTCTCT	94347
LS423614.1	55414446	CTCCCTGGTCACTCTCTT	55414429
LS423614.1	7212898	TCTCCCTGGTCACTC	7212884
AC158296.6	181325	TCCCTGGTCACTCTCTTA	181308
XR_002941147.1	5873	TCTCCCTGGTCACTCTC	5889
XR_002941146.1	6104	TCTCCCTGGTCACTCTC	6120
XM_024352563.1	6152	TCTCCCTGGTCACTCTC	6168
XM_024352562.1	6159	TCTCCCTGGTCACTCTC	6175
XM_024352561.1	6174	TCTCCCTGGTCACTCTC	6190
XR_002941145.1	6225	TCTCCCTGGTCACTCTC	6241

XM_024352560.1	6225	TCTCCCTGGTCACTCTC	6241
XM_024353570.1	2728	TCTCCCTGGTCACTCTC	2744
CP027083.1	64175254	CTCCCTGGTCACTCTCT	64175270
CP027083.1	33522521	TCCCTGGTCACTCTC	33522535
CP027083.1	74038647	CCTGGTCACTCTCTT	74038633
XM_024240188.1	6191	TCTCCCTGGTCACTCTC	6207
XM_024240187.1	6237	TCTCCCTGGTCACTCTC	6253
XR_002912889.1	6677	TCTCCCTGGTCACTCTC	6693
XR_002912888.1	6722	TCTCCCTGGTCACTCTC	6738
XM_024010326.1	738	TCTCCCTGGTCACTCTC	754
XM_024010325.1	738	TCTCCCTGGTCACTCTC	754
XM_012466500.1	6185	TCTCCCTGGTCACTCTC	6201
AC236497.1	14566	TCTCCCTGGTCACTCTC	14582
AP023477.1	70373543	CTCCCTGGTCACTCTCT	70373559
AP023477.1	34546277	CTCCCTGGTCACTCT	34546291
NM_001362862.2	6097	TCTCCCTGGTCACTCTC	6113
NM_001362863.2	6124	TCTCCCTGGTCACTCTC	6140
FR799580.1	185176	CTCCCTGGTCACTCTCT	185192
XM_024464007.1	472	CTCCCTCGTCACTCTCTTATT	452
CP013203.1	5136801	CCTGGTCACTCTCTTAT	5136785
XM_012733809.1	2275	TCTCCCTGGTCACTCTC	2259
XM_004694238.2	2267	TCTCCCTGGTCACTCTC	2251
CP055242.1	6706558	TCTCGCTGGTCACTCTCTTAT	6706538
AP023482.1	22644416	TCTCCCAGGTCACTCTCTT	22644434
AP023482.1	16683251	TCTCCCTGGTCACTCTC	16683235
AP023482.1	44075993	CCCTGGTCACTCTCTT	44075978
AP023482.1	15944981	CTCCCTGGTCACTCT	15944967
AP023479.1	48016079	CTCCCTGTTCACTCTCTTATT	48016059
LR877229.1	10280687	CTCCCTGGTCACTCTCT	10280671
LR877229.1	12079709	TGGTCACTCTCTTAT	12079695
LR877229.1	47555154	CTCCCTGGTCACTCT	47555140
LR877198.1	67955308	TCCCTGGTCACTCTCTT	67955292
LR877198.1	13949408	TCTCCCTGGTCACTC	13949394
LR877198.1	39660837	CCCTGGTCACTCTCT	39660823
LR877198.1	87550219	CCTGGTCACTCTCTT	87550205
LR862368.1	30728468	TCCCTGGTCACTCTATTATT	30728449
LR862368.1	46682100	TCTCCCTGGTCACTCTC	46682084
LR862368.1	30678003	TCCCTGGTCACTCTCT	30677988
LR862368.1	32170032	CCTGGTCACTCTCTT	32170018
LR862361.1	13011192	TCTCCCTGGTCTCTCTCTT	13011174
LR862361.1	6633788	CTGGTCACTCTCTTATT	6633772
LR862361.1	60936043	TCTCCCTGGTCACTCTC	60936027
LR862361.1	22931834	CTCCCTGGTCACTCT	22931848
LR862361.1	23032086	CTCCCTGGTCACTCT	23032100
LR862361.1	50687521	TCTCCCTGGTCACTC	50687507
AC279687.1	88398	CTCCCTGGTCACTCTCT	88382
CP002352.1	2034124	TCTCCCTGGTCACTCTC	2034108
AP023475.1	53280683	TCCCTGGTCTCTCTCTTATT	53280702
AP023475.1	53740717	TCCCTGGTCACTCTC	53740731
AP023475.1	81848063	CCCTGGTCACTCTCT	81848077
AP023476.1	71003346	TCTCCCTGGTCACTC	71003360
AP023476.1	5618759	CTGGTCACTCTCTTA	5618745
AP023476.1	53953669	CCCTGGTCACTCTCT	53953655

PRODUCTS OF SIGNIFICANT ALIGNMENTS (CDS FEATURE)

- **SLAM family member 5-like** GenBank
- **CD48 antigen-like** GenBank
- **Titin-like** GenBank
- **Coiled-coil alpha-helical rod protein 1-like** GenBank GenBank
- **Protein MCM10 homolog** GenBank GenBank GenBank
- **Basement membrane-specific heparan sulfate proteoglycan core protein-like** GenBank GenBank GenBank GenBank GenBank GenBank GenBank
- **LOW QUALITY PROTEIN: putative phosphatidylinositol 4-kinase alpha-like protein P2** GenBank
- **Phosphatidylinositol 4-kinase alpha** GenBank GenBank GenBank
- **WAT1-related protein At4g28040-like** GenBank
- **Integrator complex subunit 5** GenBank
- **PIK4CA variant protein** GenBank

CONCLUSIONS

FINAL CONSIDERATIONS

In this work, the sequence of 22 *RNA* bases of the *SARS-Cov-2* Rdrp gene used as a positive control of the rhinopharyngeal swab was modified using the Systems Equilibrium Trichromatic Theory and its computer program. The links between the **RNA** of ***SARS-Cov-2*** and the genome of ***Bat*** and ***Pangolin***, both considered implicated in the Covid-19 infection, have been identified in part and clarified better. These links will be further explored in the **Second Part** of this document. Ample space has been dedicated to the analysis of a sequence with new characteristics (*misc_feature*) of the ***Human Immunodeficiency Virus*** (***HIV-1***) identified by the modification of the sequence of 22 bases of *RNA* of the *Rdrp* gene of *SARS-Cov-2*. The same treatment has been given to several sequences of ***Influenza A virus*** and, above all, to the numerous sequences of ***Bacilli Paranthracis*** and ***Anthracis***. With regard to the latter bacterium, being considered in the highest priority category of infectious agents associated with a potential risk of use as a biological weapon, further investigation by the international scientific community is desirable in order to neutralise the common genomic aspects identified (or further identifiable) at *SARS-Cov-2*. To make replication of the results shown possible and to allow all scientists to make further independent investigations, the procedure used for modifying *DNA* or *RNA* sequences was taught. The modifications of the RNA sequence of the *Rdrp gene* of the *Sars-Cov-2* and the methodology proposed in this document open different perspectives in order to start studies finalized and innovative treatments of the pandemic from Covid 19 and for a more functional reorganization of the databases of the DNA and RNA.

END

OF THE

FIRST PART

www.ttesystems.eu

Corrispondence: ***nunzio.bonaventura@libero.it***

13 April 2021