

**Analisi e Modificazioni di
Sequenze di DNA o di RNA
con la**

**TEORIA TRICROMATICA
DELL'EQUILIBRIO
DEI SISTEMI**



**I MISTERI DEL DNA
(Parte Prima)**

***Relazioni tra SARS-CoV-2,
PIPISTRELLO, PANGOLINO,
INFLUENZAVIRUS A, HIV-1
e i BACILLI PARANTHRACIS
e ANTHRACIS***

**Autore: *Nunzio Bonaventura*
Tecnico informatico: *Vincenzo Viggiano***

13 Aprile 2021

INDICE

Riassunto	pag. 5
Premessa	pag. 6
Guida alla lettura	pag. 8
Introduzione	pag. 12
CAPITOLO I°: <i>Concetti Base della Biologia Molecolare</i>	
Obiettivi del Capitolo I°	pag. 17
1.1 La Biologia molecolare, gli aminoacidi e le proteine	pag. 18
CAPITOLO II°: <i>Procedura di Base per l'Esecuzione di una Ricerca Blast e suoi Obiettivi Principali</i>	
Obiettivi del Capitolo II°	pag. 26
2.1 Come si effettua una ricerca Blast di base	pag. 27
2.2 Come utilizzare i risultati di una ricerca Blast di base	pag. 40
CAPITOLO III°: <i>Dai Pipistrelli agli Esseri Umani attraverso i Pangolini: il Gene RdRp e la Sequenza di Rna "GTGAAATGGTCATGTGTGGCGG"</i>	
Obiettivi del Capitolo III°	pag. 46
3.1 L'ipotesi più accreditata: Dai pipistrelli agli esseri umani attraverso i pangolini	pag. 47
3.2 Il Gene RdRp e la sequenza di RNA "GTGAAATGGTCATGTGTGGCGG" del SARS-Cov-2	pag. 49
CAPITOLO IV°: <i>Ricerca BLAST sulla SEQUENZA di RNA "GTGAAATGGTCATGTGTGGCGG" del Gene RdRp (RdRp_SARSr-F) (WH-Human_1/China/2019-Dec)</i>	
Obiettivi del Capitolo IV°	pag. 56
4.1 Ricerca BLAST sulla sequenza di RNA "GTGAAATGGTCATGTGTGGCGG" del Gene RdRp (RdRp_SARSr-F) - (Hitlist size: 100)	pag. 57
CAPITOLO V°: <i>Ricerche BLAST su due specifiche sequenze di RNA del Gene RdRp</i>	
Obiettivi del Capitolo V°	pag. 75
5.1 Ricerca BLAST sulla sequenza di RNA	

“GTGAAATGGT T ATGTGTGGCGG” del RdRp Gene del “Bat coronavirus BM48-31/BGR/2008, complete genome” – (Hitlist size: 100)	pag. 76
5.2 Ricerca BLAST sulla sequenza di RNA	
“GTGAGATGGTCATGTGTGGCGG” del Gene RdRp del “Bat SARS-like coronavirus isolate bat-SL-CoVZC45, complete genome” – (Hitlist size: 100)	pag. 95
5.3 Ricerca BLAST sulla sequenza di RNA	
“GTGAGATGGTCATGTGTGGCGG” del Gene RdRp del “Bat SARS-like coronavirus isolate bat-SL-CoVZC45, complete genome” – (Hitlist size: 1000)	pag. 122

CAPITOLO VI°: Sostituzioni di BASI nella SEQUENZA di RNA
“GTGAAATGGTCATGTGTGGCGG” del GENE
RdRp (RdRp_SARSr-F) e successive ricerche BLAST

Obiettivi del Capitolo VI°	pag. 126
6.1 Ricerca BLAST sulla sequenza modificata di RNA	
“GTGAAATGGT A ATGTGTGGCGG” del RdRp Gene (Hitlist size 1000)	pag. 127
6.2 Ricerca BLAST sulla sequenza modificata di RNA	
“GTGAAATGGT G ATGTGTGGCGG” del RdRp Gene (Hitlist size 1000)	pag. 129
6.3 Ricerca BLAST sulla sequenza modificata di RNA	
“GTGAT T ATGGTGATGTGTGGCGG” del RdRp Gene (Hitlist size 1000)	pag. 131
6.4 Ricerca BLAST sulla sequenza modificata di RNA	
“GTGAC C ATGGTGATGTGTGGCGG” del RdRp Gene (Hitlist size 100)	pag. 132

CAPITOLO VII°: Procedura per la Modificazione della Sequenza
Originaria e la Generazione di Nuove Sequenze

Obiettivi del Capitolo VII°	pag. 135
7.1 Modificazione parziale di una sequenza originaria e generazione di nuove sequenze	pag. 136
7.2 I ventiquattro “Trend Non Manifesti”	pag. 138
7.3 Procedura per la modificazione parziale della sequenza originale	pag. 139

CAPITOLO VIII°: Modificazioni della Sequenza di RNA
“GTGAAATGGTCATGTGTGGCGG” del Gene RdRp
(RdRp_SARSr-F) e Generazioni di Nuove Sequenze

Obiettivi del Capitolo VIII°	pag. 149
8.1 Modificazioni della sequenza di RNA	
“GTGAAATGGTCATGTGTGGCGG” del Gene RdRp (RdRp_SARSr-F) e generazioni di nuove sequenze	pag. 150

CAPITOLO IX°: Ricerche Blast e Discussione dei Risultati Ottenuti
dagli Allineamenti Significativi della Sequenza n° 1/3

Obiettivi del Capitolo IX°	pag. 158
9.1 Sequenza n° 1/3 (T=0; C=1; G=2; A=3) – (Hitlist size: 100)	pag. 160

9.2 Sequenza n° 1/3 (T=0; C=1; G=2; A=3) – (Hitlist size: 1000)	pag. 167
9.3 Il segmento della Sequenza n° 1/3 identificato come corrispondente al Virus dell’HIV-1	pag. 176
9.4 Predizione dei codoni costitutivi del segmento della sequenza n°1/3 e discussione sui loro ipotetici prodotti	pag. 183
9.5 Sequenza di aminoacidi “ M A M T H ”	pag. 184
9.6 La modificazione della sequenza di aminoacidi “ M A M T H ” nella sequenza di aminoacidi “ M V M T H ”	pag. 190
9.7 “Audentes fortuna iuvat”: la sequenza “ ATGGTAATGACACATT ”	pag. 198
9.8 Dalla sequenza di aminoacidi “ M V M T H ” alla sequenza di aminoacidi “ M V M T H F ”	pag. 220
9.9 Sequenza di aminoacidi “ M A M T H K ”	pag. 227
9.10 Sequenza di aminoacidi “ M A M T H K S Q ”	pag. 240
9.11 Sequenza di aminoacidi “ M A M T H K S Q M K ”	pag. 246
9.12 Sequenza di aminoacidi “ M A M T H K S Q M K P R Q M A P ”	pag. 257

CAPITOLO X°: *Ricerche Blast e Discussione dei Risultati Ottenuti dagli Allineamenti Significativi della Sequenza n° 3/1*

Obiettivi del Capitolo X°	pag. 267
10.1 Sequenza n° 3/1 (T=2; C=0; G=3; A=1) – (Hitlist size: 100)	pag. 270
10.2 I principali “prodotti” identificati attraverso ricerche BLAST eseguite sulla Sequenza n° 3/1 – (Hitlist size: 100) e il ruolo fondamentale svolto dalle Idrolasi di Serina	pag. 396
10.3 Sequenza n° 3/1 (T=2; C=0; G=3; A=1) – (Hitlist size: 5000)	pag. 399
10.4 I principali “prodotti” identificati dalle ricerche BLAST effettuate sulla Sequenza n° 3/1 (Hitlist size: 5000)	pag. 433

CAPITOLO XI°: *Ricerche Blast e Discussione dei Risultati Ottenuti dagli Allineamenti Significativi della Sequenza n° 8/1*

Obiettivi del Capitolo XI°	pag. 437
11.1 Sequenza n° 8/1 (T=1; C=0; G=3; A=2). – (Hitlist size: 100)	pag. 439
11.2 Sequenza n° 8/1 (T=1; C=0; G=3; A=2) – (Hitlist size: 1000)	pag. 446

CAPITOLO XII°: *Ricerche BLAST di altre Sequenze Modificate della Sequenza di RNA “GTGAAATGGTCATGTGTGGCGG” del Gene RdRp (RdRp_SARSr-F)*

Obiettivi del Capitolo XII°	pag. 465
12.1 Sequenze 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	pag. 467

CONCLUSIONI

Considerazioni finali	pag. 633
-----------------------	----------

RIASSUNTO

In questo lavoro, avvalendoci della Teoria Tricromatica dell'Equilibrio dei Sistemi e del suo programma computerizzato, è stata modificata una sequenza di 22 basi di *RNA* del *Gene RdRp* del *SARS-CoV-2*. Questa sequenza è una delle sequenze di *RNA* utilizzate come controllo positivo del tampone rinofaringeo, insieme ad altre sequenze dei geni del *SARS-CoV-2*. Sono stati individuati in parte e chiariti meglio, tra tutti i possibili legami identificati dalle ricerche effettuate nella banca dati del DNA e dell'*RNA*, i legami tra l'*RNA* del *SARS-CoV-2* e il genoma di due animali, il *Pipistrello* e il *Pangolino*, entrambi considerati implicati nell'infezione da Covid-19. Dalle ricerche BLAST emergono però anche inaspettati allineamenti significativi (in cui sono coinvolti non solo delle basi, ma in certi casi interi aminoacidi) corrispondenti rispettivamente ad una sequenza dalle caratteristiche nuove (*misc_feature*) del *Virus dell'Immunodeficienza Umana (HIV-1)*, a diverse sequenze dell'*Influenzavirus A* e, soprattutto, a numerose sequenze dei *Bacilli del Paratrace (Bacillus Paranthracis)* e dell'*Antrace (Bacillus Anthracis)*, quest'ultimo un batterio gram-positivo infettivo, agente eziologico del *carbonchio ematico* nell'animale e dell'*antrace* nell'uomo (batterio quest'ultimo considerato nella categoria di massima priorità degli agenti infettivi associati a rischio potenziale d'uso come arma biologica). Per rendere possibile la replicazione dei risultati ottenuti, una parte di questo documento è stata dedicata all'insegnamento della procedura utilizzata per la modificazione di sequenze di DNA o RNA. Le modificazioni della sequenza di *RNA* del *Gene RdRp* del *SARS-CoV-2* e la metodologia proposta in questo documento aprono prospettive diverse per avviare studi finalizzati e trattamenti innovativi della pandemia da Covid 19 e per una riorganizzazione più funzionale delle banche dati del DNA e dell'*RNA*.

PREMESSA

Circa cinque mesi fa ho iniziato a scrivere questo documento e mai avrei pensato di imbarcarmi in un'impresa di tale portata e di doverla concludere in così poco tempo. Purtroppo la pandemia mi costringe a pubblicare uno scritto, diversamente da come sono abituato a fare, in parte incompleto, nonostante la sua corposità, giustificata dalla necessità di copiare nel documento tutti i dati originali più importanti e renderli immediatamente disponibili. Mi scuso per queste limitazioni, ma sono certo che ne comprenderete le motivazioni. Comunque, per rendere la lettura più agevole e comprensibile, all'inizio dei capitoli ho riassunto gli argomenti trattati al loro interno e quale sia lo scopo della loro trattazione. Questo lavoro, pur utilizzando la stessa metodologia dei miei precedenti studi sull'analisi e le modificazioni di sequenze di DNA (studi pubblicati nel sito web www.ttesystems.eu/applicazioni.php, in <https://www.linkedin.com/in/dr-nunzio-bonaventura-41723660/> e in <https://www.facebook.com/Dott.NunzioBonaventura>), assume un carattere totalmente autonomo. Il mio intento iniziale era quello di modificare, avvalendomi della mia Teoria dei Sistemi (www.ttesystems.eu) e del suo programma computerizzato, alcune sequenze di RNA del SARS-CoV-2. L'obiettivo finale sarebbe dovuto essere quello di individuare e chiarire meglio, tra tutti i possibili legami identificati dalle ricerche effettuate nella banca dati del DNA e dell'RNA, i legami tra l'RNA del *SARS-CoV-2* e il genoma di due animali, il *Pipistrello* e il *Pangolino*, entrambi considerati implicati nell'infezione da Covid-19. Questo obiettivo è stato in buona parte raggiunto, ma con mia estrema sorpresa, durante le modificazioni di una sequenza di RNA di un Gene del SARS-CoV-2 mi sono imbattuto in un allineamento significativo corrispondente ad una sequenza dalle caratteristiche nuove (*misc feature*) del *Virus dell'Immunodeficienza Umana (HIV-1)* e in altri corrispondenti a diverse sequenze dell'*Influenzavirus A* e, soprattutto, dei *Bacilli del Paratrace (Bacillus Paranthracis)* e dell'*Antrace (Bacillus Anthracis)*, quest'ultimo un batterio gram-positivo infettivo, agente eziologico del *carbonchio ematico* nell'animale e dell'*antrace* nell'uomo (https://it.wikipedia.org/wiki/Bacillus_anthraxis).

Per evitare subito qualsiasi tipo di fraintendimento, **in questo lavoro non si afferma che il *Virus dell'HIV-1* o i *Bacilli del Paratrace* e dell'*Antrace* sono presenti totalmente nella sequenza di RNA oggetto di studio del SARS-CoV-2. Piuttosto si constata che le modificazioni di questa sequenza di RNA del SARS-CoV-2, attraverso la metodologia che qui verrà proposta e insegnata, permettono di generare (rispettando rigorosamente i "Trend Non Manifesti" della sequenza di RNA del SARS-CoV-2) delle nuove sequenze che (nelle ricerche effettuate nel database BLAST) identificano, per elevatissima somiglianza (somiglianza non solo delle singole basi, ma in certi casi anche di alcuni aminoacidi), una sequenza del *Virus dell'HIV-1* (per approfondimenti si veda il Capitolo IX) e diverse sequenze dei *Bacilli del Paratrace* e dell'*Antrace* (per approfondimenti si veda il Capitolo X). Naturalmente, è possibile eseguire anche la **procedura inversa**, cioè modificando le **nuove sequenze generate**, identificare, per elevatissima somiglianza, la **sequenza originaria**, la **sequenza di RNA del SARS-CoV-2** precedentemente menzionata. Essendo il *Virus dell'HIV-1* già conosciuto per la sua letalità e l'*antrace* considerato nella categoria A (cioè di massima priorità) degli agenti**

infettivi associati a rischio potenziale d'uso come arma biologica, è comprensibile che sia rimasto molto colpito dal risultato della ricerca nella banca dati. In verità, nonostante la sorpresa avuta alla scoperta dei numerosi allineamenti significativi tra la sequenza di **RNA** del **SARS-CoV-2** e il *Bacillo dell'Antrace*, recentemente mi era già capitato di leggere riguardo studi ed esperimenti inerenti l'associazione tra **antrace** e **coronavirus**. In particolare, mi riferisco al protocollo di ricerca fatto dai ricercatori dell'**Università di Pittsburgh** per identificare un potenziale candidato al vaccino contro il coronavirus Sars-Cov-2 che causa la malattia Covid-19 (si veda il documento presentato qui di seguito).

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-anthraxis-sars-cov-2-research?fbclid=IwAR0KfslAd-xM2TEAx7C6xqtUHDmtdnxUPVD8pBRLn9xgbhy_wXSAAC3YOM

Come è riportato nello stesso sito web dell'Università, "...la ricerca prevede l'espressione di una singola proteina SARS-CoV-2 in un **ceppo attenuato (indebolito)** del batterio *Bacillus anthracis*.....questo ceppo attenuato non produce tossine che possono causare danni, ma è utile per fornire proteine virali alle cellule nei topi di laboratorio per produrre immunità e per proteggersi dal Covid-19....". Nella stessa pagina del sito web è scritto che "**Questo studio fornirà approfondimenti su come produrre un vaccino Covid-19 efficace per l'uso nell'uomo, ma non è destinato a produrre o testare direttamente un vaccino....La stessa proteina SARS-CoV-2 verrà introdotta nelle cellule che più laboratori**

in tutto il mondo stanno utilizzando per saperne di più sul Covid-19 e sviluppare candidati al vaccino e altri potenziali trattamenti...”.

L'infezione da **antrace**, si presenta in *tre forme*, ma la **forma polmonare** (rarissima e presente soltanto in categorie a rischio), contratta per inalazione, è **considerevolmente più grave** ed è stata associata alla paura diffusa di un attacco bioterroristico

(https://it.wikipedia.org/wiki/Bacillus_anthraxis).

La *forma polmonare* dell'infezione dell'**antrace** **inizia con i sintomi tipici di un'influenza (mal di gola, febbre, mialgia); a seguire tosse, dolore al petto e difficoltà di respirazione. Questa forma è, dal punto di vista anatomo-patologico, una linfadenite-mediastinite emorragica, piuttosto che una polmonite vera e propria, e non comporta presenza di bacilli nell'espettorato**

(https://db0nus869y26v.cloudfront.net/it/Bacillus_anthraxis).

La mia scoperta di una relazione tra l'RNA del *Gene RdRp (RdRp_SARSr-F)* del **SARS-CoV-2** (gene attualmente utilizzato, insieme ad altri, per fare diagnosi da Covid-19) e il DNA del *Virus dell'Immunodeficienza Umana (HIV-1)* e di quello del *Bacillus Anthracis*, attualmente per me resta inspiegabile, ma meritevole di approfondimenti e chiarimenti da parte di tutta la comunità scientifica. In ogni caso, questa scoperta ha completamente stravolto la tipologia del documento che avevo previsto di scrivere e pubblicare.

Nel mio intento iniziale, questo sarebbe dovuto essere un lavoro iperspecialistico, meno esteso e destinato solo agli studiosi del settore (biologi molecolari, biotecnologi, genetisti, ecc.). Diversamente da quanto previsto, **adesso i miei principali intenti sono due**. Il **primo** è quello di contribuire attivamente nella lotta alla pandemia fornendo, agli **specialisti del settore**, informazioni utili e inedite per svolgere ulteriori approfondimenti scientifici e proseguire lo studio finalizzato alla creazione di *terapie anticorpali monoclonali, antivirali, antibiotici e protocolli di intervento specifici anti Covid-19*.

Il **secondo** intento è quello di svelare per la prima volta e a tutti i **misteri inerenti i “trend non manifesti” presenti in tutte le sequenze di DNA o RNA**, di diffonderne la loro rivoluzionaria conoscenza e di insegnare, con un linguaggio quanto più possibile semplice e accessibile (per quello che la complessità della materia può permettere), la metodologia scientifica con la quale sono giunto alle conclusioni sintetizzate nel riassunto di questo studio. In questo modo **permetterò a tutti gli studiosi**, non soltanto agli specialisti del settore, ma a chiunque abbia una capacità media di manipolare dati, sia interessato e abbia la volontà di farlo, **di replicare autonomamente i miei risultati**. A tale scopo, sarà necessario acquisire i concetti base della biologia molecolare e la prassi necessaria per eseguire le ricerche sugli *allineamenti significativi* di una sequenza di DNA o RNA in una banca dati del Dna e dell'Rna (questi insegnamenti saranno chiaramente evidenziati come riservati al lettore principiante), nonché una metodologia specifica per la modificazione delle sequenze geniche, quest'ultima di interesse anche e soprattutto al lettore esperto. Naturalmente, l'opportunità offerta di replicare i risultati implica la **pubblicazione integrale degli algoritmi utilizzati in questo lavoro e in quelli precedentemente già pubblicati**. Convinto che i ricercatori sapranno fare tesoro di queste informazioni, mi rendo fin da adesso disponibile per eventuali chiarimenti o spiegazioni e auguro a tutti una buona lettura e un proficuo lavoro di approfondimento.

GUIDA ALLA LETTURA

INTRODUZIONE

L'Introduzione è consigliata a tutti i lettori. In essa è spiegata molto sinteticamente la *procedura completa per analizzare e modificare una sequenza di DNA (o di RNA)*. Gli approfondimenti dei contenuti trattati nell'Introduzione sono accessibili attraverso i collegamenti in questo stesso documento o in altri documenti già precedentemente pubblicati.

CAPITOLO I

Il Capitolo I è riservato ai principianti e ha come obiettivo generale quello di permettere al lettore di acquisire alcuni principi fondamentali di *Biologia Molecolare* necessari per la comprensione dei contenuti trattati nei capitoli successivi. In particolare, sono definiti i concetti di *Biologia Molecolare*, di *DNA* (Codificante e Non Codificante), di *RNA* (RNA messaggero, Transfer, Ribosomiale e Non Codificante), di *Basi Azotate*, *Codoni*, *Aminoacidi* e *Sintesi Proteica*.

CAPITOLO II

Il paragrafo 2.1 del Capitolo II è riservato ai principianti. In esso è spiegata la procedura di base per effettuare una ricerca nel **BLAST** database e comparare le *sequenze nucleotidiche* o le *sequenze proteiche* delle molecole di DNA o di RNA. Invece, **il paragrafo 2.2 è consigliato a tutti i lettori**, perché la **Ricerca BLAST** è spiegata e integrata alle altre *Fasi* della procedura completa proposta nel mio modello di studio del Dna (o RNA) e ne costituisce una fase propedeutica. Inoltre, l'esempio utilizzato per spiegare dei concetti fondamentali si riferisce ai risultati di alcune ricerche *BLAST* relative alla sequenza di **RNA** del **SARS-COV-2** oggetto di questo studio.

CAPITOLO III

Il Capitolo III è consigliato a tutti i lettori. Nel paragrafo 3.1 è presentata una ricerca bibliografica che indaga sull'ipotesi che ci possa essere una relazione di trasmissione del SARS-CoV dai pipistrelli ai pangolini e da questi ultimi agli esseri umani. Nel paragrafo successivo, invece, è descritta l'importanza del **gene RdRp del SARS-CoV-2**, chiarito il motivo per il quale è stata selezionata la sequenza di **22 basi di RNA** **"GTGAAATGGTCATGTGTGGCGG"** e spiegati i motivi che giustificano le sue molteplici modificazioni.

CAPITOLO IV

Questo capitolo è consigliato a tutti i lettori che abbiano acquisito dai capitoli precedenti le competenze necessarie per eseguire una Ricerca Blast completa e per interpretarne i

risultati. In esso sono presentati i risultati della *Ricerca Blast* completa eseguita sulla sequenza di **22 basi di RNA** “GTGAAATGGTCATGTGTGGCGG”, sequenza tratta dal **Gene RdRp** (*pSARS-F*). Essendo la **RNA polimerasi RNA dipendente (RdRp)** il motore del processo di **trascrizione** del *SARS-CoV-2*, la comprensione dei risultati di questo capitolo è fondamentale per la interpretazione dei risultati che saranno esposti nei capitoli successivi.

CAPITOLO V

In questo capitolo (la sezione “Obiettivi del Capitolo V” è consigliata a tutti i lettori) sono state effettuate le ricerche BLAST sulle sequenze “GTGAAATGGT**T**ATGTGTGGCGG” e “GTG**A**GATGGTCATGTGTGGCGG”. Queste due sequenze sono importanti perché sono le due sequenze del **gene RdRp** (*RNA polimerasi RNA dipendente*) i cui allineamenti significativi delle ricerche BLAST hanno identificato il **Pipistrello** (l’unico “organismo” che le due sequenze hanno in comune) e il **Pangolino**, animali entrambi attualmente considerati implicati nella trasmissione del *SARS-CoV*.

CAPITOLO VI

Nel Capitolo VI (la sezione “Obiettivi del Capitolo VI” è consigliata a tutti i lettori), per non trascurare l’impatto di eventuali altre variazioni di basi azotate nelle stesse posizioni evidenziate in blu nella sequenza “GTG**A**AATGGT**C**ATGTGTGGCGG” ed identificare eventuali altri “organismi” rilevanti per le finalità di questo studio, sono state eseguite *tutte le sostituzioni di basi possibili*, ottenendo come risultato altre *4 sequenze*. Queste sequenze sono state oggetto di ricerche BLAST.

CAPITOLO VII

Questo capitolo è consigliato a tutti i lettori, ma è **indispensabile** per coloro tra essi che abbiano una forte motivazione a comprendere l’importante e rivoluzionario argomento delle *analisi e delle modificazioni di sequenze di DNA o RNA* basato sui principi della **TEORIA TRICROMATICA DELL’EQUILIBRIO DEI SISTEMI (T.T.E.S.)**: la **procedura** attraverso la quale **modificare la Sequenza Originaria e generare nuove sequenze di DNA o RNA**.

CAPITOLO VIII

Questo capitolo (la sezione “Obiettivi del Capitolo VIII” è consigliata a tutti i lettori) è dedicato alle **modificazioni** della sequenza di RNA “GTGAAATGGTCATGTGTGGCGG” del *Gene RdRp* (*RdRp_SARSr-F*). Le modificazioni di questa **sequenza** e la **metodologia proposta in questo documento aprono prospettive diverse per avviare studi finalizzati e trattamenti innovativi della pandemia da Covid 19**.

CAPITOLO IX

In questo capitolo (la sezione “Obiettivi del Capitolo IX” è consigliata a tutti i lettori) sono riportati i risultati della ricerca BLAST effettuata sulla **Sequenza n° 1/3**. Questa sequenza è una delle tante generate dalla **Sequenza Originaria**, la sequenza “GTGAAATGGTCATGTGTGGCGG” del Gene RdRp. Sono stati indentificati moltissimi “organismi” e, tra essi, due sequenze del ***Pipistrellus pipistrellus*** e una sequenza del ***Virus dell’HIV-1***. La sequenza identificata del ***Virus dell’HIV-1*** è stato oggetto di approfondimento e i risultati sono stati sorprendenti. Sono state rintracciate relazioni con **tre** specie di **pipistrelli**, due specie di **pangolini** e diversi “organismi” implicati in sintomatologie prevalentemente **infettive**, **febbrili** e **polmonari**. Da evidenziare anche le relazioni con il ***Virus dell’Influenza A***, una sequenza che si riferisce al **dominio N-terminale della RNA polimerasi diretta dal DNA** e, soprattutto, un numero notevole di sequenze del ***Bacillus anthracis***, batterio letale e che causa **danni polmonari**.

CAPITOLO X

In questo capitolo (la sezione “Obiettivi del Capitolo X” è consigliata a tutti i lettori) sono riportati i risultati della ricerca BLAST effettuata sulla **Sequenza n° 3/1**. Questa sequenza è una delle tante generate dalla **Sequenza Originaria**, la sequenza “GTGAAATGGTCATGTGTGGCGG” del Gene RdRp. Sono stati indentificati molti “organismi” e, tra essi, ***Manis pentadactyla*** (***Pangolino Cinese***), sei sequenze di ***Pipistrellus pipistrellus***, quattro sequenze del ***Virus dell’Influenza A*** e, soprattutto, un numero considerevole di sequenze dei ***Bacilli Paranthracis***, ***Anthraxis*** e ***Bacillo Cereus***. Il ***Bacillus Anthracis***, per la sua **letalità** e per il suo massiccio coinvolgimento nel **danno polmonare**, sarà oggetto di ampia trattazione.

CAPITOLO XI

In questo capitolo (la sezione “Obiettivi del Capitolo XI” è consigliata a tutti i lettori) sono riportati i risultati della ricerca BLAST effettuata sulla **Sequenza n° 8/1**. Questa sequenza è una delle tante generate dalla **Sequenza Originaria**, la sequenza “GTGAAATGGTCATGTGTGGCGG” del Gene RdRp. Sono stati indentificati molti “organismi” e, tra essi, anche due specie di pipistrelli (***Rhinolophus ferrumequinum*** e ***Pipistrellus pipistrellus***) e un numero considerevole di sequenze dei ***Bacilli Paranthracis*** e ***Anthraxis***.

CAPITOLO XII

In questo capitolo (la sezione “Obiettivi del Capitolo XII” è consigliata a tutti i lettori) sono riportati i risultati delle Ricerche BLAST di altre Sequenze Modificate della Sequenza di RNA “GTGAAATGGTCATGTGTGGCGG” del Gene RdRp (RdRp_SARSr-F). Le sequenze modificate sono le sequenze 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**.

INTRODUZIONE

Molti aspetti metodologici di *analisi e di modificazioni di sequenze di DNA o RNA* proposti in questo scritto si basano sui principi della **TEORIA TRICROMATICA DELL'EQUILIBRIO DEI SISTEMI (T.T.E.S.)**, una teoria dei sistemi con la quale è possibile osservare, analizzare, controllare e modificare lo stato di qualsiasi sistema (<http://www.ttesystems.eu/>; <http://www.ttesystems.eu/applicazioni.php>).

Con opportuni adattamenti del software di base della è possibile analizzare e modificare una sequenza di DNA (o di RNA) in *maniera innovativa*.

La **procedura completa** per analizzare e modificare una sequenza di DNA (o di RNA) prevede **NOVE FASI**

(<http://www.ttesystems.eu/Cms/files/Pdf/Introduzione%20DNA%20ITA.pdf>).

A causa del poco tempo avuto a disposizione per la stesura di questo documento e della enorme quantità di dati presentati, le **Fasi** considerate in questo scritto sono **solo** quelle qui di seguito evidenziate in **rosso** (e non tutte sono state completate come previsto dalla procedura):

Prima Fase: Analisi della Sequenza Originaria

Seconda Fase: Modificazione della Sequenza Originaria

Terza Fase: Confronto delle Rappresentazioni Grafiche

Quarta Fase: Ricerche Blast

Quinta Fase: Scoperta ed Evidenziazione degli Organismi in Comune

Sesta Fase: Specificazione del Prodotto dell'Allineamento Significativo

Settima Fase: Ricerca Bibliografica Mirata su un solo Organismo

Ottava Fase: Ricerca Bibliografica Mirata su due o più Organismi

Nona Fase: Considerazioni Conclusive e Scelte Pragmatiche

Prima Fase: Analisi della Sequenza Originaria

L'acquisizione della sequenza e le sue **4 principali rappresentazioni grafiche** costituiscono la PRIMA FASE di studio, quella dell'**ANALISI DELLA SEQUENZA ORIGINARIA** (si vedano

[http://www.ttesystems.eu/Cms/files/Pdf/Analisi%20e%20modificazioni%20di%20Sequenze%20di%20DNA%20o%20di%20RNA%20con%20la%20T%20T%20E%20S%20\(Capitolo%20II%20C2%B0\).pdf](http://www.ttesystems.eu/Cms/files/Pdf/Analisi%20e%20modificazioni%20di%20Sequenze%20di%20DNA%20o%20di%20RNA%20con%20la%20T%20T%20E%20S%20(Capitolo%20II%20C2%B0).pdf) e

<http://www.ttesystems.eu/Cms/files/Pdf/Introduzione%20DNA%20ITA.pdf>).

L'acquisizione della sequenza di DNA (o di RNA) da analizzare e modificare può essere effettuata direttamente dall'*organismo* oggetto di studio, dal sito del **NCBI** [National Center for Biotechnology Information (1)] o da qualsiasi altra fonte.

(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/>

Seconda Fase: Modificazione della Sequenza Originaria

Successivamente (e a discrezione dello studioso), inizia la SECONDA FASE (ed eventualmente anche le successive), la **MODIFICAZIONE DELLA SEQUENZA ORIGINARIA**.

Il software della **T.T.E.S.** offre la possibilità di modificare la **sequenza originaria** e di generare numerose e diverse **nuove sequenze** di DNA o di RNA che rispettano fedelmente i numerosi e diversi “*trend non manifesti*” della sequenza originaria.

I “**trend non manifesti**” sono *andamenti*, non chiaramente identificabili, presenti in tutte le sequenze di DNA e di RNA (per approfondimenti si vedano i Capitoli VIII e IX in questo documento e <http://www.ttesystems.eu/Cms/files/Pdf/Introduzione%20DNA%20ITA.pdf>).

Terza Fase: Confronto delle Rappresentazioni Grafiche

Segue la TERZA FASE, quella in cui è possibile **CONFRONTARE LE RAPPRESENTAZIONI GRAFICHE** della **sequenza originaria** con quelle delle **nuove sequenze**.

Dal *confronto delle quattro principali rappresentazioni grafiche* della **sequenza originaria** con quelle delle **nuove sequenze** possono emergere importanti *somiglianze e differenze*.

Quarta Fase: Ricerche Blast

La fase successiva, la QUARTA FASE, è quella delle **RICERCHE BLAST** (*Basic Local Alignment Search Tool* (2)), rispettivamente sulla **sequenza originaria** e su tutte le **nuove sequenze** [per approfondimenti si veda il Capitolo II in questo documento e tutti i risultati presentati in [Analisi e modificazioni di Sequenze di DNA o di RNA con la T T E S \(Capitolo I - Parte Seconda\)](#)].

Lo scopo di questa fase è quello di identificare, per la **sequenza originaria** e per le **nuove sequenze**, tutti gli *allineamenti significativi* con le sequenze e con gli *organismi* memorizzati nella banca dati.

Quinta Fase: Scoperta ed Evidenziazione degli Organismi in Comune

La QUINTA FASE è quella della **SCOPERTA** ed **EVIDENZIAZIONE** degli “**organismi**” (e, conseguentemente, delle *basi di Dna o Rna*) che sono *in comune* tra la **sequenza originaria** e le **nuove sequenze** (per approfondimenti si vedano il paragrafo 2.2 del Capitolo II e i diversi esempi presentati nei capitoli successivi in questo documento).

- (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%20ITA.pdf> sono presentati i **criteri** suggeriti per eseguire, in maniera relativamente semplice, la complessa procedura di *evidenziazione*.

Sesta Fase: Specificazione del Prodotto dell'Allineamento Significativo

Nella SESTA FASE è necessario **SPECIFICARE**, per tutti gli “organismi” che sono *in comune*, il “**Prodotto**” dell'*allineamento significativo* e quali *basi modificate di Dna (o Rna)* hanno in comune. Un esempio di questa ultima procedura è quello presentato a pagina 33 del Capitolo I - *Parte Prima* di [Analisi e modificazioni di Sequenze di DNA o di RNA con la T T E S \(Capitolo I° - Parte Prima\)](#).

Complessivamente, i risultati di questa fase sono molto importanti perché (insieme ai risultati delle FASI SUCCESSIVE) da essi dipenderà l'eventuale scelta d'intervenire concretamente nella modificazione del **materiale biologico** delle *sequenze analizzate per la realizzazione delle diverse e possibili finalità pratiche*.

Ricerca Bibliografica Mirata su un solo Organismo

Nella SETTIMA FASE viene effettuata una **RICERCA BIBLIOGRAFICA MIRATA** allo scopo di confermare l'esistenza d'importanti relazioni tra le caratteristiche (per es., i “prodotti” evidenziati nella fase precedente) di uno degli *organismi* [per es., *Pseudomonas*. Per approfondimenti si vedano le pagine 37 - 46 di [Analisi e modificazioni di Sequenze di DNA o di RNA con la T T E S \(Capitolo I° - Parte Prima\)](#)] identificato con la ricerca Blast eseguita sulla **nuova sequenza** e alcune caratteristiche funzionali della **sequenza originaria**.

Ricerca Bibliografica Mirata su due o più Organismi

Nell'OTTAVA FASE, a discrezione del ricercatore e per le sue specifiche finalità, può essere effettuata una **RICERCA BIBLIOGRAFICA MIRATA** allo scopo di confermare l'esistenza d'importanti relazioni tra le caratteristiche (per es., i “prodotti” evidenziati nella fase precedente) di **due o più organismi** [per es., *Pseudomonas* e *Heligmosomoides polygyrus*. Per approfondimenti si veda [Analisi e modificazioni di Sequenze di DNA o di RNA con la T T E S \(Capitolo I° - Parte Prima\)](#)], identificati con la ricerca Blast eseguita sulla **nuova sequenza** e alcune caratteristiche funzionali della **sequenza originaria**.

Nona Fase: Considerazioni Conclusive e Scelte Pragmatiche

In questa ULTIMA FASE, si traggono le **CONCLUSIONI** del lavoro eseguito nelle fasi precedenti e si fanno le **SCELTE PRAGMATICHE** necessarie, come ad esempio, quella di utilizzare le *conoscenze acquisite* e/o le *nuove sequenze di DNA o RNA* (parziali o integrali) per scopi di ricerca scientifica, industriale, alimentare, ecc., oppure quella d'intervenire concretamente nella modificazione del **materiale biologico** (il DNA o l'RNA) di una o di più *sequenze analizzate* per la realizzazione delle diverse e possibili finalità pratiche.

CAPITOLO I

**CONCETTI BASE
DELLA BIOLOGIA
MOLECOLARE**

OBIETTIVI DEL CAPITOLO I

Questo capitolo è riservato ai principianti e ha come obiettivo generale quello di permettere ai lettori di acquisire alcuni principi fondamentali di *Biologia Molecolare* necessari per la comprensione dei contenuti trattati nei capitoli successivi.

In particolare, sono definiti i concetti di *Biologia Molecolare*, di *DNA* e *RNA* (entrambi Codificanti e Non Codificanti), di *Aminoacidi* e di *Sintesi Proteica*.

Le *Basi Azotate* sono considerate con particolare attenzione per il ruolo che svolgono nella codifica degli aminoacidi; infatti, l'**ordine nella disposizione sequenziale dei nucleotidi** (di cui le basi azotate sono i principali costituenti) costituisce il cuore dell'*informazione genetica*.

Importanti anche i concetti di *RNA messaggero*, *Transfert* e *Ribosomiale*, implicati nei processi di *traduzione* e *trascrizione* e indispensabili alla comprensione della *sintesi proteica*.

Alla fine del capitolo sono allegate due **Tabelle dei Codoni** (o *triplette*) che è importante imparare ad utilizzare per tradurre un codice genetico in una sequenza di aminoacidi.

1.1 LA BIOLOGIA MOLECOLARE, GLI AMINOACIDI E LE PROTEINE

La **biologia molecolare** è la branca della *biologia* (la scienza che studia i processi fisici e chimici dei sistemi viventi) che si occupa degli esseri viventi a livello dei meccanismi *molecolari* (cioè, delle *molecole*, unità fondamentali delle *sostanze*, composte da due o più *atomi*, di *elementi* simili o diversi, uniti insieme) alla base della loro *fisiologia* (del loro funzionamento come esseri viventi), concentrandosi in particolare sulle interazioni tra le *macromolecole* (molecole di dimensioni molto grandi e di peso molecolare molto elevato), ovvero *proteine* e *acidi nucleici* (DNA e RNA) (https://it.wikipedia.org/wiki/Biologia_molecolare).

Negli organismi viventi sono riscontrabili **due tipi di acidi nucleici** (macromolecole deputate alla conservazione e al trasporto dell'informazione genetica): **DNA** e **RNA**.

Il **DNA** (o *Acido Desossiribonucleico*) è un acido nucleico che contiene le informazioni genetiche necessarie alla *biosintesi* (alla sintesi chimica che avviene all'interno di un organismo vivente) di **RNA** e **proteine**, molecole indispensabili per lo sviluppo ed il corretto funzionamento della maggior parte degli organismi viventi (<https://it.wikipedia.org/wiki/DNA>).

L'**RNA** (o *Acido Ribonucleico*) è implicato in vari ruoli biologici di codifica, decodifica, regolazione e espressione dei geni [unità ereditarie fondamentali degli organismi viventi, ovvero sequenze *nucleotidiche* (vedi sotto) di DNA o più raramente, in certi **virus**, di RNA] (<https://it.wikipedia.org/wiki/RNA>).

Dal punto di vista chimico, il **DNA** ha una struttura a **doppia catena** (nota come doppia elica) i cui elementi costituenti sono chiamati **nucleotidi**. I *nucleotidi* sono costituiti da tre componenti fondamentali: un gruppo fosfato, uno zucchero pentoso e una base azotata.

Uno dei **concetti fondamentali** che è necessario memorizzare per poter comprendere i contenuti di questo documento è il seguente:

Le **basi azotate** che entrano nella formazione dei *nucleotidi* sono **quattro**:

adenina (A);
timina (T);
citosina (C);
guanina (G).

Nell'**RNA** al posto della **timina** è presente l'uracile, U.

Attenzione:

Di prassi, nelle banche dati del DNA e dell'RNA, la base azotata *uracile* è definita sempre con la lettera **T** (è bene ricordarlo perché questo può generare confusione).

Come mostrato in Fig. 1.1, il **DNA** può essere globalmente definito come una *doppia catena polinucleotidica* (in cui sono presenti numerose basi azotate **A**, **T**, **C** e **G**), cioè formata da più nucleotidi, *antiparallela* (perché le due catene di nucleotidi scorrono in posizione inversa l'una rispetto all'altra), *orientata* (perché i nucleotidi si organizzano secondo leggi chimiche precise che ne specificano l'orientamento nello spazio), *complementare* (perché una catena è speculare all'altra), *spiralizzata* (perché la catena si attorciglia a doppia elica su se stessa), *informazionale* (perché porta in sé l'informazione genetica) (<https://it.wikipedia.org/wiki/DNA>).

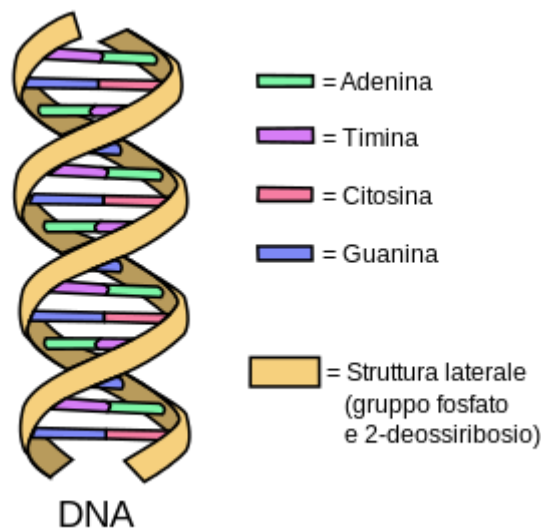


Fig. 1.1 [Struttura a doppia elica del DNA. Sono messi in evidenza gli accoppiamenti tra le quattro basi azotate. **E' fondamentale ricordare che l'Adenina si accoppia sempre con la Timina e la Citosina sempre con la Guanina** (immagine tratta da <https://it.wikipedia.org/wiki/DNA>)]

Nella **struttura laterale del DNA**, ogni molecola di fosfato forma un ponte molecolare collegando il **carbonio in posizione 3'** di una molecola di deossiribosio **con quello in posizione 5'** dello zucchero successivo (si veda la Fig. 1.2 nella pagina successiva). In conseguenza di questi legami asimmetrici **ogni filamento di DNA ha un senso, determinato dalla direzione dei ponti molecolari** (*legami fosfodiesterici*).

In una doppia elica, il senso di un filamento è opposto a quello del filamento complementare. Per tale motivo, i due filamenti che costituiscono una doppia elica sono detti **antiparalleli**. Le estremità asimmetriche di un filamento di DNA sono definite *estremità 5'* ed *estremità 3'* e, *in vivo*, l'unica modalità di sintesi di un *acido nucleico* segue il senso 5' → 3', motivo per cui le sequenze di DNA ed RNA sono scritte in senso 5' → 3' (<https://it.wikipedia.org/wiki/DNA>).

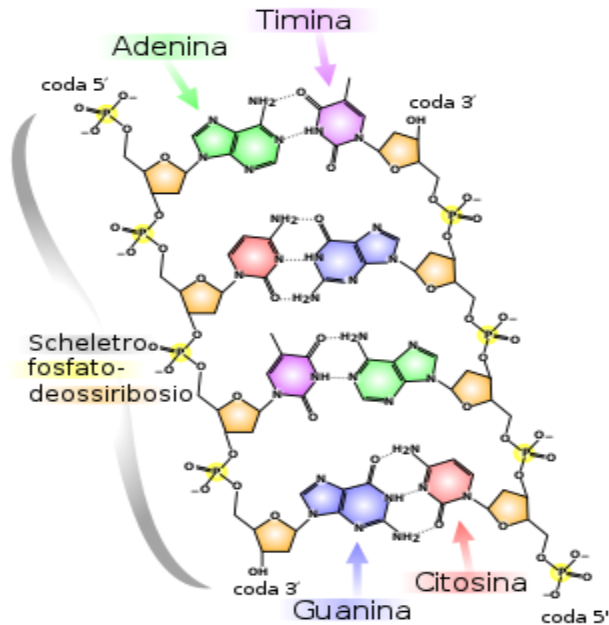


Fig. 1.2 [Struttura chimica del DNA. **E' fondamentale ricordare che l'unica modalità di sintesi di un acido nucleico segue il senso 5' → 3'** (immagine tratta da <https://it.wikipedia.org/wiki/DNA>)].

Come il DNA, l'RNA è assemblato come una *catena di nucleotidi*, ma a differenza del DNA è più frequente in natura come un **singolo filamento ripiegato su se stesso**, piuttosto che un doppio filamento accoppiato (<https://it.wikipedia.org/wiki/RNA>).

Molti **virus** [*entità biologiche* con caratteristiche di *parassiti obbligati*, in quanto si replicano esclusivamente all'interno delle cellule degli organismi [https://it.wikipedia.org/wiki/Virus_\(biologia\)](https://it.wikipedia.org/wiki/Virus_(biologia))] codificano le loro informazioni genetiche utilizzando un genoma a RNA.

Un altro **concetto fondamentale** che è necessario memorizzare per poter comprendere i contenuti di questo documento è il seguente: l'**ordine nella disposizione sequenziale dei nucleotidi** costituisce l'*informazione genetica*.

Essendo i nucleotidi sostanzialmente costituiti e rappresentati dalle numerose basi azotate A, T, C e G, ne segue che l'**ordine sequenziale assunto dalle basi azotate A, T, C e G** è il cuore dell'informazione genetica.

Facciamo un esempio.

La seguente sequenza di 22 basi azotate

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

costituirà un'informazione genetica diversa dalla seguente sequenza di 22 basi azotate

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

tranne che nell'area evidenziata in grigio, dove l'informazione è la stessa perché le basi sono le stesse e sono poste nello stesso ordine.

In **bioinformatica** questa informazione genetica è trattata in **maniera binaria**, cioè come **una serie di 0 e 1**, e trasformata in **algoritmi**.

L'*informazione genetica* è tradotta tramite il **codice genetico**. Esso è l'insieme di regole con le quali viene tradotta l'informazione codificata nei *nucleotidi* costituenti i *geni* per la sintesi di *proteine* nelle cellule attraverso gli *aminoacidi* (molecole organiche costitutive delle *proteine*) corrispondenti.

La sequenza **amminoacidica** prodotta forma le **proteine**, macromolecole biologiche costituite da catene di *aminoacidi* che svolgono una vasta gamma di funzioni all'interno degli organismi viventi, tra cui funzioni *metaboliche* (le trasformazioni chimiche delle macromolecole che avvengono nelle cellule), funzione di sintesi come *replicazione del DNA* (il processo biologico di produzione di due repliche identiche di DNA da una molecola di DNA originale), funzione di *segnalazione cellulare* (un processo che rende le cellule capaci di percepire e rispondere correttamente al loro microambiente) e il *trasporto* di molecole da un luogo ad un altro dell'organismo (<https://it.wikipedia.org/wiki/Proteine>).

Quelli che seguono sono **concetti fondamentali** che è necessario memorizzare per poter comprendere i contenuti di questo documento. Esistono **tre tipi diversi di RNA**:

- **RNA messaggero (mRNA)**, necessario per *trascrivere* la sequenza di aminoacidi per la sintesi delle proteine (https://it.wikipedia.org/wiki/RNA_messaggero);
- **RNA di trasporto** o **RNA Transfer (tRNA)**, costituito dalla tripletta complementare o *anticodone* che identifica l'aminoacido nel mRNA. Esso trasferisce un amminoacido specifico di una catena polipeptidica in crescita al *sito ribosomiale* della sintesi proteica durante la *traduzione* (https://it.wikipedia.org/wiki/RNA_transfer);
- **RNA ribosomiale** o **rRNA**, aggancia l'*RNA messaggero* e scorrendolo legge (traduce) le triplette (o *Codoni*. Per approfondimenti si veda la pagina successiva) in interazione con il *tRNA* (che trasporta l'aminoacido). L'rRNA non codifica direttamente le proteine, ma è il componente essenziale dei ribosomi (che provvedono all'assemblaggio delle proteine) presenti in tutte le cellule viventi (https://it.wikipedia.org/wiki/RNA_ribosomiale).

Gli organismi cellulari utilizzano l'RNA messaggero (**mRNA**) per trasmettere le informazioni genetiche mediante le basi azotate *guanina (G)*, *uracile (U)*, *adenina (A)* e *citosina (C)*, che dirigono la sintesi di proteine specifiche (<https://it.wikipedia.org/wiki/RNA>).

Il processo di **traduzione genetica** è possibile solo in presenza dell'RNA, che è generato per *complementarità con le quattro basi dei nucleotidi del DNA* in un processo noto come **trascrizione** ([https://it.wikipedia.org/wiki/Trascrizione_\(biologia\)](https://it.wikipedia.org/wiki/Trascrizione_(biologia))). Tale processo non genera solo filamenti di RNA destinati alla traduzione, ma anche frammenti già in grado di svolgere svariate funzioni biologiche (ad esempio all'interno dei ribosomi, dove l'RNA ha una funzione strutturale).

Durante la **trascrizione**, il *DNA a doppio filamento* produce *RNA* a partire dal cosiddetto *filamento antisenso*; l'altro filamento, complementare, è detto *senso*. Con il termine **RNA antisenso** (https://it.wikipedia.org/wiki/RNA_antisenso), si intende un *RNA complementare* in sequenza con uno o più mRNA. In alcuni organismi, la presenza di un **RNA anti-senso può inibire l'espressione genetica attraverso l'appaiamento con gli specifici RNA**.

La **sintesi proteica** (il processo di *traduzione genetica* che porta alla formazione delle proteine) è possibile grazie all'esistenza di uno specifico **linguaggio a tre lettere** che associa i *codoni* [**3 nucleotidi sull'RNA** o **triplette**, per esempio **AUG** (che nel database di Blast è scritto **ATG**, perché **U = T**)] agli *amminoacidi*, i costituenti le proteine [per esempio, il codone **AUG** (o **ATG**) codifica l'aminoacido **Metionina (Met/M)**]. Le due **Tabelle dei Codoni** mostrate nella pagina successiva (Fig. 1.3 e Fig. 1.4) sono utilizzate per tradurre un codice genetico in una sequenza di aminoacidi.

Il **codoni** sono presenti lungo l'mRNA (*RNA messaggero*), un tipo di RNA che **codifica** l'informazione per l'inserimento di uno specifico *amminoacido* durante la *sintesi proteica* o per la **fine** della stessa [definito *codone di stop* (https://it.wikipedia.org/wiki/Codone_di_stop) o codone di *terminazione* o codone *non senso*, il quale è una tripletta di basi che **non codifica** per nessun amminoacido (https://it.wikipedia.org/wiki/DNA_non_codificanteDNA) e blocca quindi la traduzione del filamento di mRNA in proteina. Esistono **tre codoni di stop: UAA, UGA e UAG**]. Esistono anche **tre codoni di inizio**, il più comune è **AUG** [che viene letto come l'aminoacido *Metionina (Met/M)* o, nei batteri, come *Formilmetionina*], mentre gli altri sono due sono **GUG** e **UUG**, i quali normalmente codificano rispettivamente gli aminoacidi **Valina (Val/V)** e **Leucina (Leu/L)**, ma come codoni iniziali sono tradotti come *Metionina* o *Formilmetionina*.

Il **codone** è alla base del codice genetico (<https://it.wikipedia.org/wiki/Codone>).

Utilizzando **4 lettere** (**A, C, G e T** o **U**) a gruppi di **tre lettere** (per esempio, **AUG**) si possono avere fino a **64 combinazioni diverse** ($4 \times 4 \times 4$), in grado di *codificare* i **venti diversi amminoacidi** esistenti. Poiché esistono **64 triplette possibili** e solo **20 amminoacidi**, il *codice genetico* è detto *ridondante* perché alcuni amminoacidi possono essere codificati da **più triplette diverse**. Per esempio, l'aminoacido **Valina**, in inglese *Valine (Val/V)*, può essere codificato indifferentemente dai codoni **GUU, GUC, GUA e GUG**. Non è invece vero il contrario: ad ogni tripletta corrisponderà un solo amminoacido.

Prima di concludere, è importante sapere che, solo una piccola frazione della sequenza totale di un genoma può essere *trascritta e tradotta* per la *sintesi proteica*, mentre la maggior parte costituisce sequenze ripetute di **DNA non codificante** (https://it.wikipedia.org/wiki/DNA_non_codificante). Anche se le ragioni di questo non sono ancora del tutto chiare, le sequenze di DNA che non codificano una proteina possono essere trascritte **in RNA non codificante** (https://it.wikipedia.org/wiki/RNA_non_codificante), coinvolto nella regolazione dell'espressione genica.

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		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	(Leu/L) Leucine ↑	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		TCA		TAA	Stop (Ochre) ^[note 2]	TGA	Stop (Opal) ^[note 2]	A
	TTG		TCT		TAG	Stop (Amber) ^[note 2]	TGG	(Trp/W) Tryptophan ↑	G
C	CTT	(Leu/L) Leucine ↑	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)

CAPITOLO II

**PROCEDURA DI BASE
PER L'ESECUZIONE DI UNA
RICERCA BLAST
E SUOI OBIETTIVI
PRINCIPALI**

OBIETTIVI DEL CAPITOLO II

L'obiettivo di questo capitolo è quello di insegnare ai principianti motivati la procedura di base per effettuare una ricerca nel **BLAST** database (*Basic Local Alignment Search Tool* o *Strumento di Ricerca di Allineamento Locale* (1)) e il modo più semplice di utilizzare i risultati ottenuti da questa ricerca. Le informazioni sulle caratteristiche delle sequenze sono acquisite direttamente dal sito del **NCBI** [*National Center for Biotechnology Information* (2)].

Usando **BLAST** si possono comparare le *sequenze nucleotidiche* o le *sequenze proteiche* delle molecole di DNA o di RNA. Una *ricerca BLAST* permette al ricercatore di confrontare una *sequenza di interesse* con un database di *sequenze già conosciute* (nucleotidiche o proteiche) e di identificare, tra queste ultime, quelle che presentano delle **somiglianze** con la *sequenza di interesse*. Ad esempio, in seguito all'infezione in corso da *Covid-19*, dopo aver **sequenziato** parte del genoma di questo nuovo Coronavirus – cioè aver determinato l'ordine delle **basi azotate** [*Guanina (G)*, *Uracile* o *Timina (U o T)*, *Adenina (A)* e *Citosina (C)*] nel caso di un acido nucleico (DNA o RNA) o degli amminoacidi nel caso di **proteine** – gli scienziati sono stati e sono ancora interessati a compiere delle ricerche BLAST nel genoma (di tutti gli “organismi”) conservato nel database per vedere se contiene **geni con sequenze somiglianti**.

Trovati dei **geni con sequenze somiglianti**, la fase successiva è quella di indagare sperimentalmente la funzione dei geni identificati e ottenere così degli indizi sulle possibili funzioni dei geni del nuovo Coronavirus. La ricerca **BLAST** assegna dei **punteggi** alle *sequenze somiglianti identificate* e *ordina* queste sequenze in funzione di tali punteggi. Le *sequenze somiglianti* considerate dalla ricerca **BLAST** saranno SOLO quelle il cui **punteggio** (score) supererà una certa *soglia critica* e saranno definite **HPS** (**H**igh-scoring **S**egment **P**air o **C**oppia di **S**egmenti con **P**unteggio **E**levato).

In conclusione, un **HSP** è un sottosegmento di una *coppia di sequenze, nucleotidi o amminoacidi*, che condividono un **alto livello di somiglianza**. Il **livello di somiglianza tra le sequenze** dipende dalle *impostazioni dell'algoritmo* di allineamento locale o globale che le ha generate. A seconda delle sequenze, dell'algoritmo di allineamento e delle impostazioni utilizzate, è possibile trovare numerosi HSP all'interno di una data **coppia di sequenze immesse**, cioè la **sequenza conosciuta** (presente nel database) e la **sequenza di interesse** (quella oggetto di studio) (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 COME SI EFFETTUA UNA RICERCA BLAST DI BASE

Per effettuare una ricerca **BLAST** scrivere nel motore di ricerca la parola “BLAST” e cliccare su “BLAST: Basic Local Alignment Search Tool” (freccie rosse in Fig. 2.1).

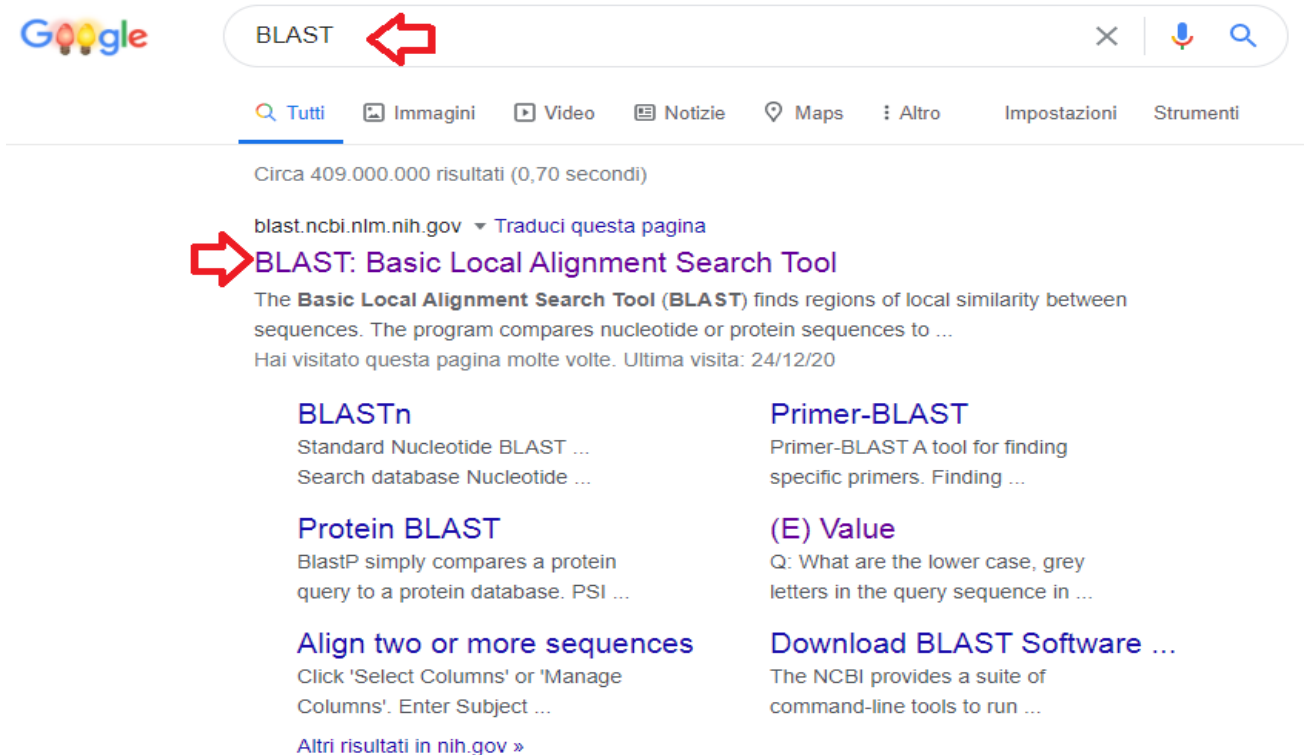


Fig. 2.1

Proseguite cliccando sul tasto previsto per scegliere il tipo di ricerca che desiderate effettuare. Sono previste **diverse possibilità di scelta** (si veda la Fig. 2.2 nella pagina successiva), ma per le finalità di questo studio e insegnamento, ne consideriamo solo **due**:

1) **Nucleotide Blast** (freccia rossa nella Fig. 2.2 nella pagina successiva)

Con questo tipo di ricerca abbiamo la possibilità di inserire una sequenza nucleotidica (una sequenza di T, C, G e A) di nostro *interesse* (come ad esempio, la sequenza *ATTGCGCGGATTATGTCC*) e identificare quelle altamente somiglianti tra le sequenze nucleotidiche conosciute;

2) **Protein Blast** (freccia verde nella Fig. 2.2 nella pagina successiva)

Con questo tipo di ricerca abbiamo la possibilità di inserire una sequenza di aminoacidi [una sequenza di lettere corrispondenti a *triplette* (o codoni) di basi T, C, G e A] di nostro *interesse* (come ad esempio la sequenza *NRPNH*) e identificare quelle altamente somiglianti tra le sequenze di aminoacidi conosciute. Gli aminoacidi sono i 20 costituenti base delle proteine (*20 lettere*). Le proteine sono formate da sequenze di lettere ed ogni lettera si riferisce ad un codone.

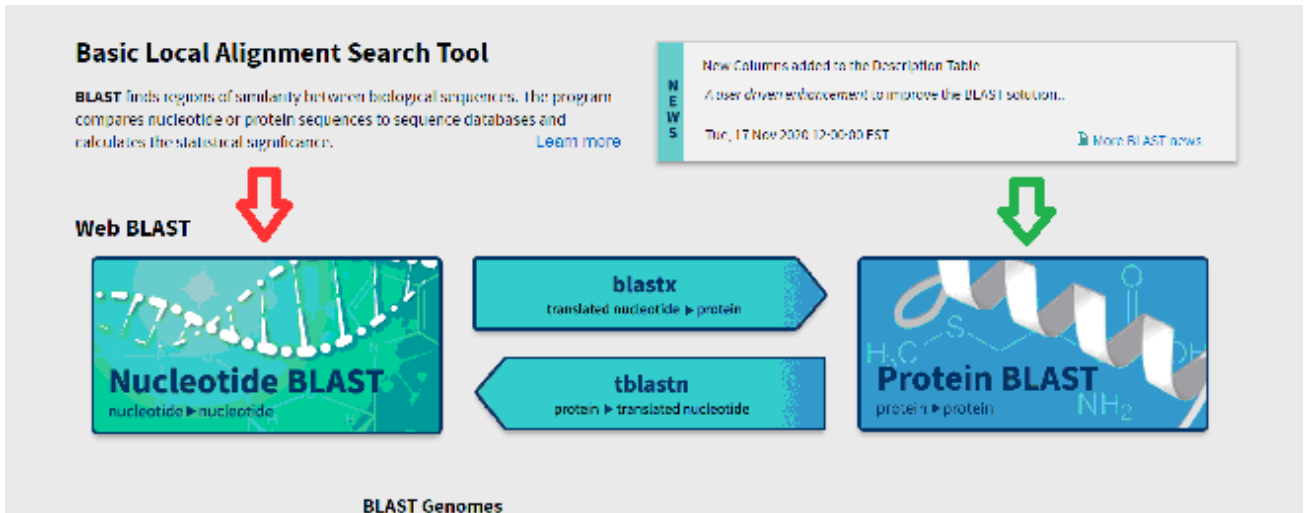


Fig. 2.2

A fini didattici, eseguiamo una **semplice ricerca di base di tipo 1 (Nucleotide Blast, freccia rossa in Fig. 2.2).**

- creiamo una *sequenza di nostro interesse*, una **SEQUENZA NUCLEOTIDICA CASUALE di 18 BASI**:

ATTGCGCGGATTATGTCC

- **copiamo e incolliamo** la sequenza nello spazio previsto in BLAST (freccia rossa in Fig. 2.3).

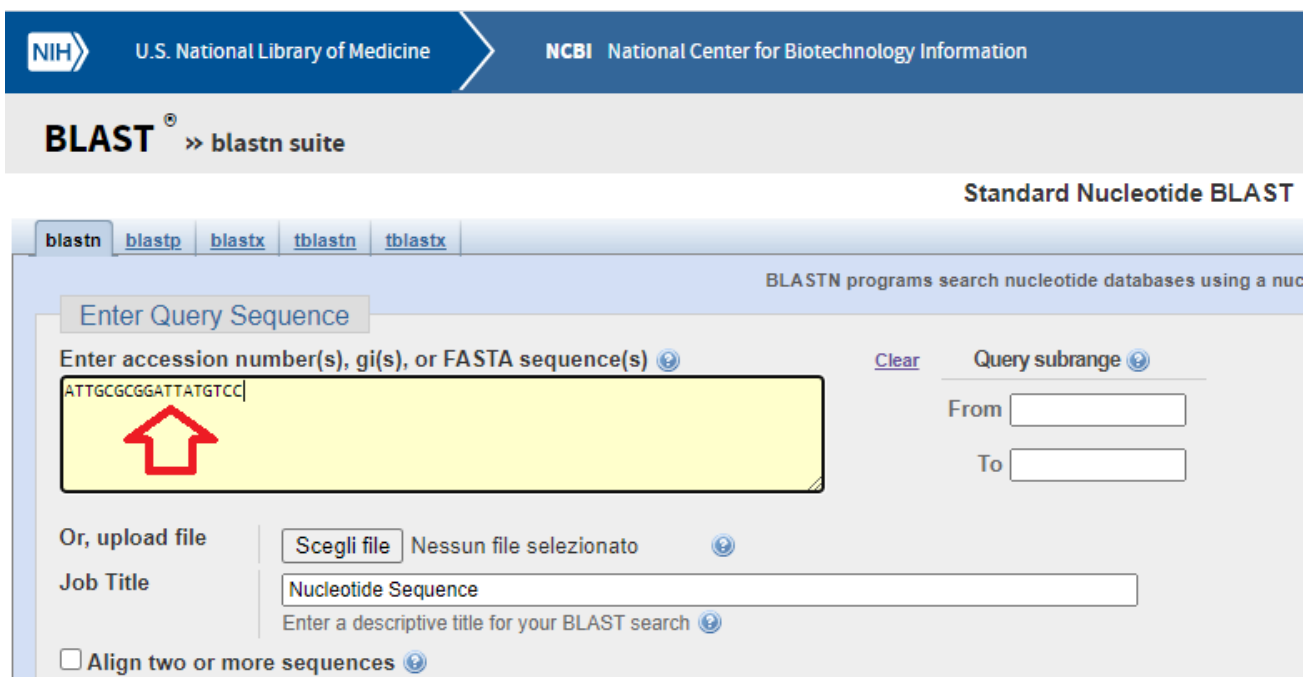


Fig. 2.3

- **clicchiamo** in basso su *Algorithm parameters* (freccia rossa in Fig. 2.4).

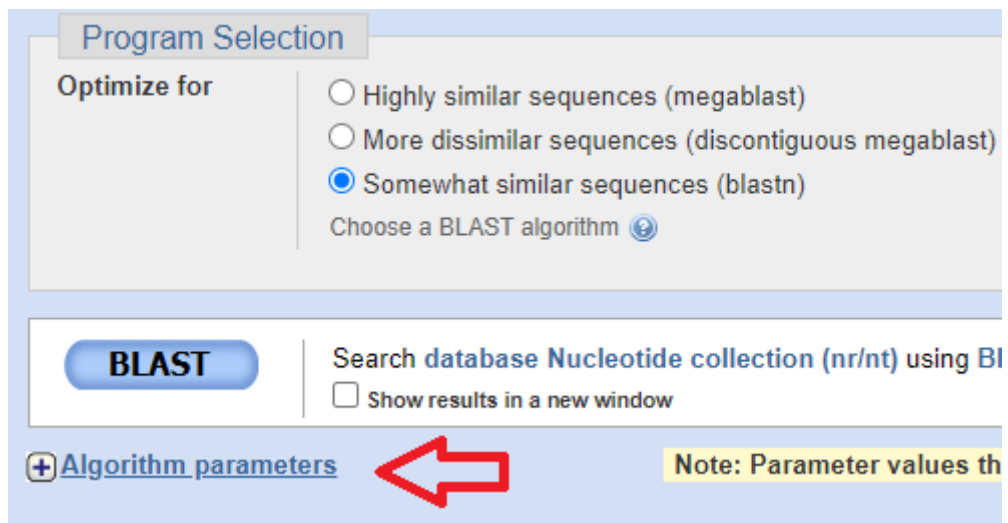


Fig. 2.4

- **stabiliamo il numero massimo di sequenze** che desideriamo ottenere dalla ricerca **BLAST** (*Max target sequences*, freccia rossa in Fig. 2.5). Nell'esempio riportato qui di seguito riportato è stato impostato il numero **10**) e poi **clicchiamo** sul tasto **BLAST** (freccia verde in Fig. 2.5).

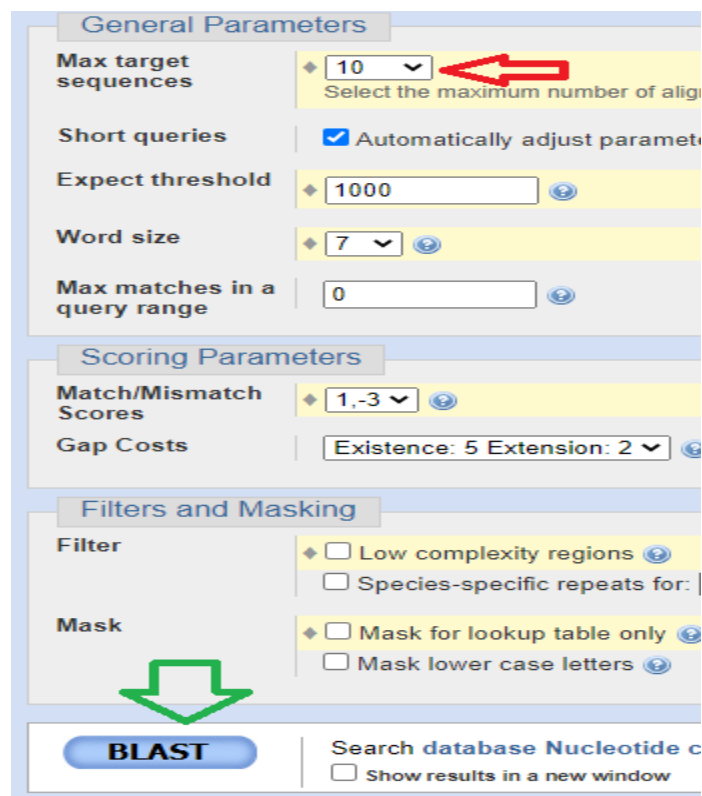


Fig. 2.5

- dopo aver atteso il tempo necessario per l'esecuzione della ricerca **BLAST**, possiamo **salvare** la ricerca (*Save Search*, freccia rossa in Fig. 2.6), ottenere il **riassunto** dei diversi parametri della ricerca (*Search Summary*, freccia verde in Fig. 2.6) e visualizzare la ricerca nei **risultati recenti** (*Recent Results* freccia blu in Fig. 2.6).

Fig. 2.6

- **clickando** sul **riassunto** dei diversi parametri della ricerca [(*Search Summary*, freccia rossa in Fig. 2.7 (a))] sono mostrati i dati generali della ricerca.

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)

- **scorrendo** verso il basso sono mostrate, nelle **Descrizioni** (*Descriptions*), le sequenze che producono **allineamenti significativi** (*Sequences producing significant alignments*) con la **sequenza che abbiamo immesso** (la *sequenza di nostro interesse*). **Cliccando nel piccolo spazio “select all”** [freccia rossa a sinistra dello schermo guardando la Fig. 2.7 (b)] selezioniamo tutte le sequenze che hanno ottenuto allineamenti significativi.

The screenshot shows the 'Descriptions' tab of a BLAST search result. At the top, there are tabs for 'Descriptions', 'Graphic Summary', 'Alignments', and 'Taxonomy'. Below the tabs, there's a header 'Sequences producing significant alignments' and a 'Download' button. A red arrow points to a 'select all' checkbox which is checked, with the text '10 sequences selected' next to it. Below this is a table with columns: Description, Scientific Name, Max Score, Total Score, Query Cover, E value, Per. Ident, Acc. Len, and Accession. The table lists 10 sequences, all with an E value of 36 and 94% query cover.

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/> Digitaria exilis annotation	Digitaria exilis	34.2	34.2	94%	36	100.00%	37200191	LR792820.1
<input checked="" type="checkbox"/> Digitaria exilis genome assembly chromosome_1A	Digitaria exilis	34.2	34.2	94%	36	100.00%	37200191	LR761605.1
<input checked="" type="checkbox"/> Bipolaris sorokiniana ND90Pr uncharacterized protein (COCSADRAFT_316619). mRNA	Bipolaris soroki...	34.2	34.2	94%	36	100.00%	2253	XM_007701312.1
<input checked="" type="checkbox"/> Bathycoccus prasinus genomic . Chromosome_8	Bathycoccus pr...	34.2	34.2	94%	36	100.00%	937610	FO082271.1
<input checked="" type="checkbox"/> Rhodobacteraceae bacterium SC52 plasmid pSC52_1 complete sequence	Rhodobacterac...	34.2	34.2	94%	36	100.00%	391128	CP049029.1
<input checked="" type="checkbox"/> PREDICTED: Scaptodrosophila lebanonensis DNA-directed RNA polymerase II subunit RPB1-like (LO...	Scaptodrosophi...	34.2	34.2	94%	36	100.00%	5811	XM_030530828.1
<input checked="" type="checkbox"/> Scophthalmus maximus chromosome 2	Scophthalmus...	34.2	34.2	94%	36	100.00%	23836697	CP026244.1
<input checked="" type="checkbox"/> Bathycoccus prasinus hypothetical protein (Bathy08g00080). partial mRNA	Bathycoccus pr...	34.2	34.2	94%	36	100.00%	828	XM_007511727.1
<input checked="" type="checkbox"/> Sclerophora farinacea voucher Wedin 6414 (UPS) RNA polymerase II largest subunit (RPB1) gene, par...	Sclerophora fari...	34.2	34.2	94%	36	100.00%	685	JX000144.1
<input checked="" type="checkbox"/> Culex quinquefasciatus lipase 3 (LOC6044763). mRNA	Culex quinquef...	32.2	32.2	88%	143	100.00%	1317	XM_001862192.2

Fig. 2.7 (b)

- **Cliccando su Alignments** [freccia rossa in alto nello schermo guardando la Fig. 2.8 (a)] e poi selezionando, in *Alignments view*, *Flat query-anchored with letters in identities* [freccia rossa leggermente a destra e più in basso nello schermo guardando la Fig. 2.8 (a)] sono mostrati le **basi della sequenza che abbiamo immesso** (la *sequenza di nostro interesse*, definita **Query**), i **codici di accesso** delle **sequenze identificate** dalla ricerca e l'*allineamento* con le **basi di tutte le sequenze identificate**.

The screenshot shows the 'Alignments' tab of a BLAST search result. At the top, there are tabs for 'Descriptions', 'Graphic Summary', 'Alignments', and 'Taxonomy'. Below the tabs, there's a header 'Alignment view' and a dropdown menu set to 'Flat query-anchored with letters for identities'. A red arrow points to this dropdown menu. To the right, there's a 'Line length' dropdown set to '60'. Below this is a 'Download' button and a 'Query range 1: 1 to 18' section. The main content shows a table of sequence alignments with columns for Query, Accession, and sequence coordinates.

Query	Accession	Coordinates
Query		1 18
LR792820.1	22891764	ATTGCGCGGATTATGTCC 18
LR761605.1	22891764	ATTGCGCGGATTATGTCC 22891780
XM_007701312.1	280	ATTGCGCGGATTATGTCC 22891780
FO082271.1	10326	ATTGCGCGGATTATGTCC 296
CP049029.1	121518	ATTGCGCGGATTATGTCC 10342
XM_030530828.1	422	TTGCGCGGATTATGTCC 121502
CP026244.1	10390556	TTGCGCGGATTATGTCC 406
XM_007511727.1	213	TTGCGCGGATTATGTCC 10390540
JX000144.1	642	ATTGCGCGGATTATGTCC 197
AM778878.1	8924	TTGCGCGGATTATGTCC 626
		TGCGCGGATTATGTCC 8939

Fig. 2.8 (a)

- **Cliccando nel piccolo spazio** previsto alla sinistra dei nomi delle varie sequenze identificate [freccia rossa a sinistra dello schermo guardando la Fig. 2.8 (b)] è possibile selezionare una specifica sequenza, mentre cliccando sul link del **Codice di Accesso** [Accession, freccia rossa a destra dello schermo guardando la Fig. 2.8 (b)] è possibile visualizzare la scheda informativa completa della sequenza.

The screenshot shows the 'Sequences producing significant alignments' section of a BLAST search. The table below lists the top results:

Description	Common Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input type="checkbox"/> Digitaria exilis annotation	Digitaria exilis	34.2	34.2	94%	36	100.00%	37200191	LR792820.1
<input type="checkbox"/> Digitaria exilis genome assembly_chromosome_1A	Digitaria exilis	34.2	34.2	94%	36	100.00%	37200191	LR761605.1
<input type="checkbox"/> Bipolaris sorokiniana ND90Pr uncharacterized protein (COCSADRAFT_316619).mRNA	Bipolaris soro...	34.2	34.2	94%	36	100.00%	2253	XM_007701312.1
<input checked="" type="checkbox"/> Bathycoccus prasinos genomic : Chromosome_8	Bathycoccus p...	34.2	34.2	94%	36	100.00%	937610	FO082271.1
<input type="checkbox"/> Rhodobacteraceae bacterium SC52 plasmid pSC52_1_complete sequence	Rhodobactera...	34.2	34.2	94%	36	100.00%	391128	CP049029.1
<input type="checkbox"/> PREDICTED: Scaptodrosophila lebanonensis DNA-directed RNA polymerase II subunit RPB1-like (LOC...	Scaptodrosop...	34.2	34.2	94%	36	100.00%	5811	XM_030530828.1
<input type="checkbox"/> Scophthalmus maximus chromosome 2	turbot	34.2	34.2	94%	36	100.00%	23836697	CP026244.1
<input type="checkbox"/> Bathycoccus prasinos hypothetical protein (Bathy08g00080), partial mRNA	Bathycoccus p...	34.2	34.2	94%	36	100.00%	828	XM_007511727.1
<input type="checkbox"/> Sclerophora farinacea voucher Wedin 6414 (UPS) RNA polymerase II largest subunit (RPB1) gene, parti...	Sclerophora fa...	34.2	34.2	94%	36	100.00%	685	JX000144.1
<input type="checkbox"/> Microcystis aeruginosa PCC 7806 genome sequencing data_contig C248	Microcystis ae...	32.2	32.2	88%	142	100.00%	13822	AM778878.1

Fig. 2.8 (b)

Tornando su Alignments, selezioniamo *Pairwise* (freccia rossa al centro, in alto e leggermente a destra nello schermo guardando la Fig. 2.9)

The screenshot shows the 'Alignments' view for the selected sequence. The 'Alignment view' dropdown is set to 'Pairwise'. Below the table, the details for the selected match are shown:

Bathycoccus prasinos genomic : Chromosome_8
 Sequence ID: [FO082271.1](#) Length: 937610 Number of Matches: 1

Range 1: 10326 to 10342 [GenBank](#) [Graphics](#) [Next Match](#) [Previous Match](#)

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

Query 1 ATTGCGCGGATTATGTC 17
 Sbjct 10326 ATTGCGCGGATTATGTC 10342

Fig. 2.9

- **Cliccando** su *CDS features* o *caratteristiche Codificanti* (freccia rossa al centro, più in basso della precedente freccia nello schermo guardando la Fig. 2.10) sono mostrati i riferimenti più importanti per lo studio della **sequenza che abbiamo immesso** (la *sequenza di nostro interesse*, definita **Query**) e, in particolare, l'allineamento significativo con la sequenza che avevamo già *selezionato* in precedenza **cliccando nel piccolo spazio** previsto [freccia rossa a sinistra dello schermo guardando la Fig. 2.8 (b)] nelle **Descrizioni** (*Descriptions*).

Alignment view: Pairwise

CDS feature [Restore defaults](#)

[Download](#) [GenBank](#) [Graphics](#)

Bathycoccus prasinus genomic : Chromosome_8
 Sequence ID: [FO082271.1](#) Length: 937610 Number of Matches: 1
 Range 1: 10326 to 10342 [GenBank](#) [Graphics](#) [Next Match](#) [Previous Match](#)

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
  
```

Fig. 2.10

Nell'esempio riportato in Fig. 2.10, la sequenza selezionata nelle **Descrizioni** (*Descriptions*) è quella che identifica “*Bathycoccus prasinus genomic : Chromosome_8*”, il cui link del **Codice di Accesso** è [FO082271.1](#) (freccia viola).

- **Cliccando** sul *Codice di Accesso* è possibile visualizzare la **scheda informativa completa** della sequenza selezionata nelle **Descrizioni** (*Descriptions*), sequenza che è stata considerata altamente somigliante alla **sequenza che abbiamo immesso** (la *sequenza di nostro interesse* o **Query**).
- **Cliccando** su *GenBank* (freccia gialla in Fig. 2.10) è possibile visualizzare la scheda informativa **specifico** del **Gene** della sequenza **Query** (si veda la Fig. 2.11 mostrata nella pagina successiva). Tra le altre importanti informazioni, notiamo le date di **pubblicazione** (freccia rossa in alto a destra nello schermo guardando la Fig. 2.11), di **presentazione** per la pubblicazione e i **riferimenti** di chi ha fatto la pubblicazione (freccia rossa più in basso e leggermente a sinistra nello schermo guardando la Fig. 2.11).

Bathycoccus prasinus genomic : Chromosome_8

GenBank: FO082271.1

[FASTA](#) [Graphics](#)Go to:






LOCUS FO082271 17 bp DNA linear PLN 09-NOV-2012
 DEFINITION Bathycoccus prasinus genomic : Chromosome_8. 
 ACCESSION [FO082271](#) REGION: 10326..10342
 VERSION FO082271.1
 KEYWORDS green alga, Mamiellales, picoeukaryote, phytoplankton.
 SOURCE Bathycoccus prasinus
 ORGANISM [Bathycoccus prasinus](#)
 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 prasinus"
 /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="CC017910.1"
 /translation="MAIVDVSPFQKICIGPASHHQTHAKVNTPKTTTRLKNTTSQEEKD
 DDDTAINKTLFRQSLPLSLQPRHNPRNTLFREFYDAGEIPIAVAHGGGSNAVTVRVDP
 ASLDYHRYLPLFFDGIREDHPYRLLSVQGAQQLLAVRPEKVADVLPQVILPLQNALK
 TRDAPIVAVAMDILRHLATCEKSIGQMLVPYYRQLLPILNVFRNCCLKVTLNVLSDG
 SKRDGKELIGKGRSAKSLRVLVKETLRVLERVGGEDALINIKYMVPTYESCVAIAER"
 ORIGIN
 1 attgcgcgga ttatgtc 
 //

Fig. 2.11

La freccia **gialla** nella Fig. 2.11 indica una parte o l'intera **sequenza che abbiamo immesso** (la *sequenza di nostro interesse* o **Query**) e che è stata trovata identica (*allineamento significativo*) alla sequenza selezionata nelle **Descrizioni** (*Descriptions*), quella che identifica “*Bathycoccus prasinus genomic : Chromosome_8*”.

La freccia **blu** nella Fig. 2.11 (mostrata nella pagina precedente) indica un'altra importantissima informazione, cioè il **Prodotto** (*product*) della sequenza selezionata nelle **Descrizioni** (*Descriptions*). Nell'esempio riportato il **Prodotto** è una **Proteina Ipotetica** (*hypothetical protein*) la cui sequenza di **aminoacidi** (che nella scheda è descritta dopo *translation =*) è indicata dalla freccia **verde**. Per inciso ricordo che una **sequenza di aminoacidi**, come quella di nostro *interesse*, la sequenza *NRPNH*, è descritta da una sequenza di lettere ed ogni lettera si riferisce ad un codone (o tripletta di basi T, C, G e A).

Nella Fig. 2.12 presentata qui di seguito, sottolineato in **verde**, è il **Prodotto** (*hypothetical protein*) della Sequenza selezionata nelle **Descrizioni** (*Descriptions*), quella che identifica “*Bathycoccus prasinus genomic : Chromosome_8*”, Nella stessa immagine è mostrata **Query**, la **sequenza che abbiamo immesso** (la *sequenza di nostro interesse*), a cui è allineato **Sbjct** (la sequenza risultante dall'allineamento significativo con *Query*). L'**Identità** (Identities) tra *Query* e *Sbjct* è il 100% (17 lettere su 17, freccia rossa in basso in Fig. 2.12). Attenzione però, l'**Identità** è del 100% perché la **sequenza di basi identica è ininterrotta** (cioè non ha lettere mancanti o diverse); tuttavia, le **basi iniziali immesse** per la ricerca erano **18** (una base in più), quindi **rispetto al numero totali di basi la percentuale d'Identità è minore**.

Alignment view: Pairwise

CDS feature [Restore defaults](#)

[Download](#) [GenBank](#) [Graphics](#)

Bathycoccus prasinus genomic : Chromosome_8
 Sequence ID: [FO082271.1](#) Length: 937610 Number of Matches: 1

Range 1: 10326 to 10342 [GenBank](#) [Graphics](#) [Next Match](#) [Previous Match](#)

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

Fig. 2.12



Fig. 2.14

Nell'allineamento stabilito in BLAST, la **sequenza posta in alto** corrisponde alla **sequenza Query**, la **sequenza che abbiamo immesso** (la *sequenza di nostro interesse*) ([https://it.qaz.wiki/wiki/Sense_\(molecular_biology\)](https://it.qaz.wiki/wiki/Sense_(molecular_biology))), a cui è allineato *Sbjct* (la sequenza risultante dall'allineamento significativo con *Query*). Nel nostro esempio, come è stato già detto, l'**Identità** (Identities) tra *Query* e *Sbjct* è del 100%.

Come si può notare, la freccia **blu** (mostrata in Fig. 2.13) indica la tripletta **ATG** e sotto di essa corrisponde la tripletta speculare **TAC**, quella indicata dalla freccia **gialla**.

Come è stato spiegato nel capitolo II, alla lettera **A** posta su una catena si associa la lettera speculare **T** posta sull'altra. Lo stesso dicasi con le lettere **G** e **C**. Quindi, semplificando, **ATG** e **TAC** sono un modo diverso per riferirsi alla stessa sequenza di DNA, perché corrispondono alle lettere poste sulle due catene speculari. Tuttavia, **TAC** non è il codone che codifica l'**aminoacido H**, codificato invece, come già anticipato a pagina 36, dai codoni **CAT** e **CAC**. Se però leggiamo la **sequenza posta in basso** (*Genes*), la sequenza *Sbjct* in Fig. 2.13, in senso 5' → 3' (il senso indicato dalle frecce nere), cioè dalla destra dello schermo verso la sua sinistra, il codone **TAC** si leggerà **CAT** e corrisponderà ad uno dei due codoni che codificano l'**aminoacido H** (si veda la Fig. 2.15 in basso in questa pagina). **Attenzione:** per decidere il senso di lettura del codone bisogna riferirsi al senso di lettura stabilito dalla direzione indicata dalla **Freccia Viola** mostrata in Fig. 2.13.

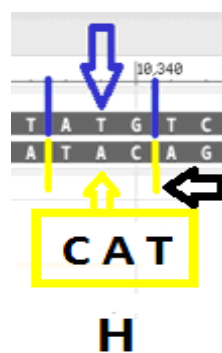


Fig. 2.15

Se il senso di lettura stabilito dalla direzione indicata dalla **Freccia Viola** scorre in direzione inversa a quella mostrata in Fig. 2.13, cioè in senso 5' → 3' parallelamente alla **sequenza posta in alto** [si veda la Fig. 2.16 mostrata nella pagina successiva. La figura si riferisce al grafico della sequenza di **RNA** (GTGAAATGGTC**ATG**TGTGGCGG) che è oggetto di studio di questo lavoro su **SARS-COV-2**, la **sequenza posta in alto**], allora

l'aminoacido codificato non sarà l'*Istidina* (**H**), ma quello indicato dalla lettera **M** (freccia rossa in basso), cioè la *Metionina* (in inglese Methionine, **Met/M**) e il cui codone è costituito dalla tripletta **ATG** corrispondente direttamente con la **sequenza posta in alto** indicata dalla freccia nera diretta verso il basso.

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

GenBank: MW411920.1

[GenBank](#) [FASTA](#)

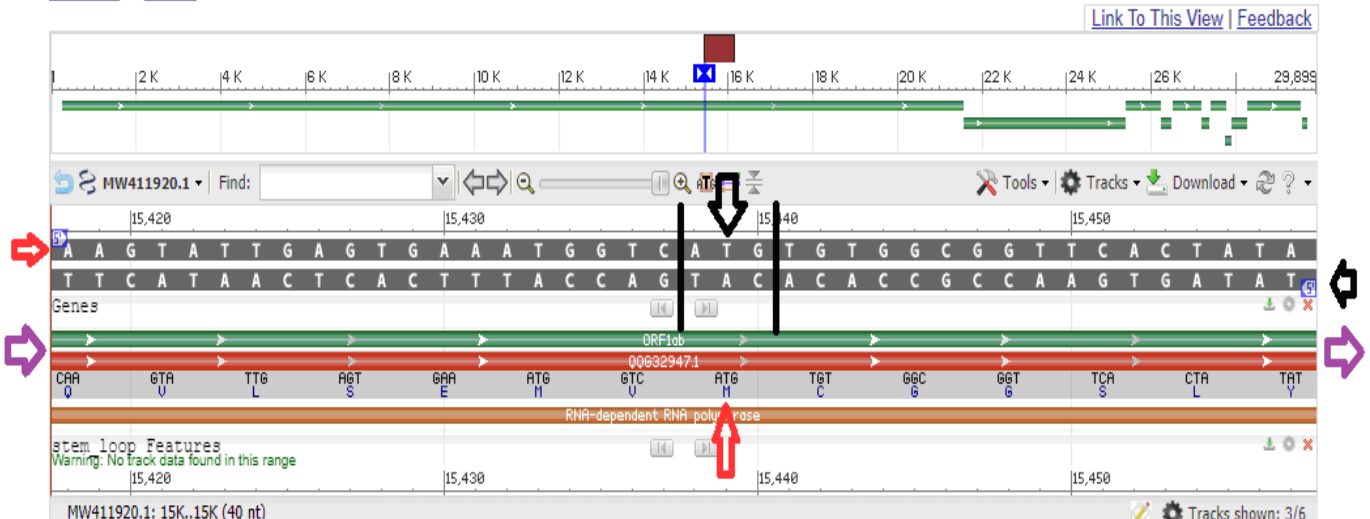


Fig. 2.16

Si confronti la Fig. 2.16 con la Fig. 2.17 mostrata qui di sotto, copia identica della Fig. 2.13 e già discussa nelle pagine precedenti.

Bathycoccus prasinos genomic : Chromosome_8

GenBank: FO082271.1

[GenBank](#) [FASTA](#)

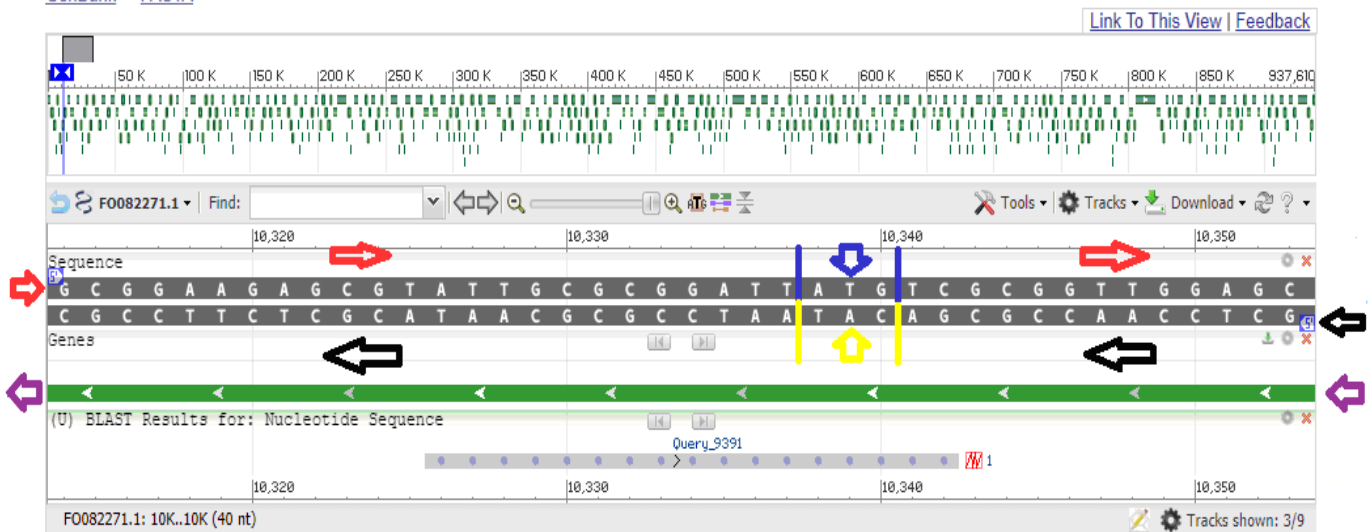


Fig. 2.17

Ancora un particolare da chiarire prima di concludere il paragrafo. In Fig. 2.12 a pagina 35 e nella tabella con i link attivi riportata nella pagina successiva, è mostrata la lettera **R** in **viola**. Questo colore della lettera indica al lettore che nella **sequenza *Sbjct*** (la sequenza risultante dall'allineamento significativo con **Query**) l'**aminoacido** successivo ad **H** (Histidin) è l'**Arginina** (in inglese Arginine, **Arg/R**), ma che l'ultima base del codone (la terza base) dell'**Arginina** (base necessaria per codificare la **R**) non è presente nella **sequenza Query**, perché la **sequenza Query** ha una *base diversa* dalla **Sequenza *Sbjct*** o semplicemente perché è *più corta* e ne è quindi mancante.

Nel caso del nostro esempio, la **sequenza Query** ha una *base diversa* dalla **Sequenza *Sbjct***. Mentre nella **sequenza Query** alle basi "TC" segue la base "C", nella **Sequenza *Sbjct*** segue la base "G" (come potete constatare nella **catena posta in alto** nella Fig. 2.17 dopo il codone **ATG**), base necessaria per codificare l'**aminoacido Arginina** (Codone DNA "TCG" = Codone RNA "CGA" = **R**).

Nella tabella con i **link attivi** riportata qui di seguito è possibile fare un po' di esperienza sui contenuti trattati in questo paragrafo. A tale scopo, consiglio anche di creare delle sequenze casuali, di immetterle in BLAST e di esercitarsi (sia con sequenze di DNA, sia con sequenze di RNA) ad identificare i codoni codificanti seguendo le indicazioni fornite.

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 COME UTILIZZARE I RISULTATI DELLA RICERCA BLAST DI BASE

La **Ricerca BLAST** costituisce la **QUARTA FASE** della **procedura completa** proposta nell'*Introduzione alla Teoria Tricromatica dell'Equilibrio dei Sistemi applicata alle Analisi e Modificazioni di Sequenze di DNA o di RNA* (<http://www.ttesystems.eu/Cms/files/Pdf/Introduzione%20DNA%20ITA.pdf>).

La **Quarta Fase** è preceduta dalle fasi di **Analisi della Sequenza Originaria** (la sequenza che abbiamo intenzione di modificare), **Modificazione della Sequenza Originaria** e **Confronto delle Rappresentazioni Grafiche della Sequenza Originaria** con quelle delle **Nuove Sequenze** ottenute dalle sue modificazioni.

Come accennato nella **Premessa**, purtroppo la pandemia mi costringe a pubblicare uno scritto in parte incompleto e con una ridotta quantità di approfondimenti e collegamenti interni. Esso è infatti mancante, tra le altre cose, del *confronto delle quattro principali rappresentazioni grafiche della sequenza originaria* [la sequenza di **RNA (GTGAAATGGTCATGTGTGGCGG)** che è oggetto di studio di questo lavoro su **SARS-COV-2**] con quelle delle **nuove sequenze** (nuove sequenze che presenterò nei capitoli successivi dopo aver insegnato la procedura di *modificazione della sequenza originaria*) e le analisi delle loro *somiglianze e differenze*.

Solo dopo l'esecuzione delle **Tre Fasi** sopra descritte, si sarebbe dovuto procedere all'esecuzione delle **ricerche BLAST**, rispettivamente sulla **sequenza originaria** e su tutte le **nuove sequenze** [si vedano [Analisi e modificazioni di Sequenze di DNA o di RNA con la T T E S \(Capitolo I° - Parte Prima\)](#); [Analisi e modificazioni di Sequenze di DNA o di RNA con la T T E S \(Capitolo I - Parte Seconda\)](#)].

Come spiegato nel paragrafo precedente, usando **BLAST** si possono comparare le *sequenze nucleotidiche* o le *sequenze proteiche* delle molecole di DNA o di RNA. Una **ricerca BLAST** permette al ricercatore di confrontare una *sequenza di interesse* con un database di *sequenze già conosciute* (nucleotidiche o proteiche) e di identificare, tra queste ultime, quelle che presentano delle **somiglianze** con la *sequenza di interesse*.

In questo lavoro, lo scopo principale della **Ricerca BLAST** è quindi quello di identificare, per la **sequenza originaria** (la sequenza di **RNA del SARS-COV-2**) e per le **nuove sequenze** (le *nuove sequenze* ottenute dalle modificazioni della *sequenza originaria*), tutti gli *allineamenti significativi* con le sequenze e con gli *organismi* memorizzati nel BLAST database.

Come abbiamo visto nel paragrafo precedente, le informazioni che si possono ottenere dagli *allineamenti significativi* sono tante e tutte molto importanti.

Qui di seguito riassumo le principali informazioni, le quali sono oggetto anche di approfondimenti nelle **Fasi** successive della procedura completa proposta nell'*Introduzione alla Teoria Tricromatica dell'Equilibrio dei Sistemi applicata alle Analisi e Modificazioni di Sequenze di DNA o di RNA*:

- possibilità di identificare delle sequenze che producono **allineamenti significativi** (*Sequences producing significant alignments*) con la **sequenza che abbiamo immesso** (la *sequenza di nostro interesse o Query*), come mostrato nella Fig. 2.7 (b) del paragrafo precedente.
- possibilità di visualizzare l'**allineamento di tutte le basi delle sequenze identificate**, come mostrato nella Fig. 2.8 (a) del paragrafo precedente.
- possibilità di selezionare una **specifica sequenza e visualizzazione della scheda informativa completa** della sequenza selezionata, come mostrato nella Fig. 2.8 (b) del paragrafo precedente;
- possibilità di visualizzare la scheda informativa **specifica del Gene** della sequenza **Query** (si vedano le Figure 2.10 e 2.11 mostrate nel paragrafo precedente). Tra le altre importanti informazioni, notiamo le date di **pubblicazione**, di **presentazione** per la pubblicazione e i **riferimenti** di chi ha fatto la pubblicazione. Inoltre, sono mostrate: **1)** una parte o l'intera **sequenza che abbiamo immesso (Query)**; **2)** il **Prodotto** (*product*, ad esempio, una **Proteina Ipotetica**, *hypothetical protein*) della **sequenza specifica** selezionata tra le sequenze identificate dagli *allineamenti significativi*; **3)** la **sequenza completa di aminoacidi** che codifica il *prodotto* ottenuto.
- possibilità di visualizzare la sequenza di **aminoacidi** corrispondenti ad una parte o all'intera **sequenza che abbiamo immesso** (ad esempio, la sequenza *NRPNH*) e l'**Identità** (*Identities*) tra *Query* e *Sbjct* (la sequenza risultante dall'allineamento significativo con *Query*), come mostrato nella Fig. 2.12 nel paragrafo precedente.
- possibilità di visualizzare il **Grafico dinamico** (come mostrato nella Fig. 2.13 nel paragrafo precedente) in cui sono mostrate due sequenze che scorrono in direzione opposta, la **sequenza posta in alto** (*Sequence*, che coincide con *Query*) che segue il senso $5' \rightarrow 3'$ e quella **posta in basso** (*Genes*, che coincide con **sequenza specifica** identificata dagli *allineamenti significativi*) che scorre in direzione contraria o antisenso ($3' \rightarrow 5'$).
- possibilità d'identificare facilmente, nella parte inferiore del grafico mostrato nelle figure 2.15 e 2.16, i *codoni* e gli *aminoacidi* corrispondenti alla sequenza **Query**.

La FASE delle RICERCA *BLAST* è propedeutica alle Fasi successive, prima tra tutte la **QUINTA FASE**, la fase della **SCOPERTA** ed **EVIDENZIAZIONE** degli “**organismi**” che sono *in comune* tra la **sequenza originaria** (nel nostro caso, la sequenza di **RNA** del **SARS-COV-2**) e le **nuove sequenze** (le *nuove sequenze* ottenute dalle modificazioni della *sequenza originaria*).

Facciamo un esempio che riguarda la RICERCA *BLAST* della sequenza di **RNA** del **SARS-COV-2** oggetto di questo studio.

Come vedremo nel capitolo IV, l’“**organismo**” (tra tutti gli “organismi”) evidenziati dalla RICERCA *BLAST* della sequenza di **RNA** (**GTGAAATGGTCATGTGTGGCGG**) del **SARS-COV-2** è solo l'**Homo Sapiens** (l'Essere Umano), questo perché la sequenza oggetto di studio è stata riscontrata solo negli esseri umani.

“**Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2**”, diagnosticata dalla presenza della sequenza “**GTGAAATGGTCATGTGTGGCGG**”, non riguarda organismi diversi dall’**Homo Sapiens**, almeno fino ad adesso.

In conclusione, l’unico “**organismo**” (tra tutti gli “organismi”) che potrà essere **scoperto** ed **evidenziato** *in comune* tra la **sequenza originaria** (la sequenza di **RNA** del **SARS-COV-2**) e le **nuove sequenze** (quelle ottenute dalle modificazioni della *sequenza originaria*) potrà essere, se sarà individuato, solo l’**Homo Sapiens**.

Se invece (come vedremo nel capitolo V) prendiamo in considerazione **due sequenze** quasi totalmente identiche alla sequenza “**GTGAAATGGTCATGTGTGGCGG**”, cioè quelle che ne differiscono **SOLO** per una base, come le sequenze “**GTGAAATGGT**T**ATGTGTGGCGG**” e “**GTGAG**A**TGGTCATGTGTGGCGG**”, allora le cose cambiano radicalmente.

La ricerca BLAST fatta sulla prima sequenza (“**GTGAAATGGT**T**ATGTGTGGCGG**”) **non evidenzia** la presenza dell’**Homo Sapiens** tra gli “organismi” identificati, tranne se la ricerca si estende fino alla mutazione della base evidenziata in rosso, rispettivamente la **T** con la **C**.

Gli “organismi” identificati dalla ricerca BLAST su questa sequenza sono: il **Pipistrello**, il **Pangolino Malese** e quello **Cinese** (specie di formichieri) e la **Piccola Egretta** (una specie di uccello).

Invece, gli “organismi” identificati dalla ricerca BLAST sulla sequenza “**GTGAG**A**TGGTCATGTGTGGCGG**” sono: l’**Homo Sapiens**; **Rhinolophus Sinicus** (chiamato *Ferro di Cavallo*), **Rhinolophus Pusillus** e **Rhinolophus Monoceros** (varianti della specie del **Pipistrello**); **Altererythrobacter namhicola** (un batterio Gram-negativo); **Cebus Imitator** (Cappuccino panamense dalla faccia bianca, una specie di scimmia); **Timema bartmani** (una specie di insetto stecco della famiglia Timematidae) e **Timema Tahoe**.

Come si può notare, il **Pipistrello** è l’unico “organismo” che le **due sequenze** hanno in comune, questo perché questa specie di mammifero può presentare nel suo genoma l’una o l’altra sequenza (“**GTGAAATGGT**T**ATGTGTGGCGG**” o “**GTGAG**A**TGGTCATGTGTGGCGG**”).

Questo risultato della ricerca BLAST indica che la modifica, più o meno rilevante (nel nostro esempio la modifica è stata minima perché ha coinvolto **SOLO** una base per entrambe le **nuove sequenze** generate), di una **sequenza originaria** (nello specifico **GTGAAATGGTCATGTGTGGCGG**) può generare **nuove sequenze** i cui allineamenti significativi individuati dalla ricerca identificano **organismi comuni** non tra la **sequenza originaria** e le **nuove sequenze**, ma SOLO tra le nuove sequenze.

Questo implica la necessità di fare delle *analisi comparate* tra gli allineamenti delle **nuove sequenze** generate e confrontare sia gli “**organismi comuni**” identificati, sia i **PRODOTTI** (per esempio proteine, ma non solo. Per il concetto di “Prodotto” si veda il paragrafo precedente) inerenti le loro sequenze.

Questo aspetto riguarda la SESTA FASE della **procedura completa**, quella della **Specificazione del Prodotto dell'Allineamento Significativo**, in cui è necessario **SPECIFICARE**, per tutti gli “organismi” che sono *in comune* alla **sequenza originaria** (o che potrebbero essere in comune solo tra le **nuove sequenze**), il “**Prodotto**” dell’*allineamento significativo* e quali *basi modificate di Dna (o Rna)* hanno in comune.

Come già detto nell’*Introduzione*, i risultati di questa fase sono molto importanti perché (insieme ai risultati delle FASI SUCCESSIVE, **Ricerca Bibliografica Mirata su un solo Organismo o su due o più Organismi e Considerazioni conclusive e Scelte Pragmatiche**) da essi dipenderà l’eventuale scelta d’intervenire concretamente nella modificazione del materiale biologico delle sequenze analizzate per la realizzazione delle diverse e possibili finalità pratiche.

Prima di concludere questo capitolo, è fondamentale sottolineare che nel tempo, come ci è stato comunicato dagli stessi responsabili del BLAST DATABASE, si possono riscontrare “**eventuali incongruenze e successive variazioni degli allineamenti significativi....**” ed esse “**....devono essere attribuite ai regolari aggiornamenti e/o alle modifiche dovute alle parti mobili di questo complesso DATABASE**”.

CAPITOLO III

**DAI PIPISTRELLI
AGLI ESSERI UMANI
ATTRAVERSO I PANGOLINI:
IL GENE RdRp
E LA SEQUENZA DI RNA
“GTGAAATGGTCATGTGTGGCGG”**

OBIETTIVI DEL CAPITOLO III

Nel paragrafo 3.1 di questo capitolo è presentata una brevissima, ma molto recente, **ricerca bibliografica** che indaga sull'ipotesi che ci possa essere una relazione di trasmissione del SARS-CoV dai pipistrelli ai pangolini e da questi ultimi agli esseri umani.

Nel paragrafo 3.2, invece, è descritta l'importanza del **gene RdRp** o *RNA-dependent RNA polymerase* (in italiano *RNA polimerasi RNA-dipendente*) del SARS-CoV-2, chiarito il motivo per il quale è stata selezionata la sequenza di **22 basi di RNA** “**GTGAAATGGTCATGTGTGGCGG**”, i motivi che giustificano le sue molteplici modificazioni e spiegato a grandi linee il piano di sviluppo dei prossimi capitoli.

3.1 L'IPOTESI PIU' ACCREDITATA:

DAI PIPISTRELLI AGLI ESSERI UMANI ATTRAVERSO I PANGOLINI

L'abstract dell'articolo "**Genomic characterization and infectivity of a novel SARS-like coronavirus in Chinese bats**" (Caratterizzazione genomica e infettività di un nuovo coronavirus simile alla SARS nei pipistrelli cinesi) [(JOURNAL Emerg Microbes Infect 7 (1), 154 (2018) PMID: [PMC6135831](#); DOI: [10.1038/s41426-018-0155-5](#)], così recita: "*Il coronavirus SARS (SARS-CoV), l'agente causale della grande epidemia di SARS nel 2003, ha avuto origine nei pipistrelli. Molti coronavirus simili al SARS (SL-CoV) sono stati rilevati nei pipistrelli, in particolare quelli che risiedono in Cina, Europa e Africa. Per comprendere ulteriormente la relazione evolutiva tra il SARS-CoV e i suoi bacini idrici, tra il 2015 e il 2017 sono stati raccolti 334 pipistrelli dalla città di Zhoushan, nella provincia dello Zhejiang, in Cina. L'amplificazione PCR della proteina coronavirusale conservata RdRp ha rilevato coronavirus nel 26,65% dei pipistrelli appartenenti a questa regione, e questo numero è stato influenzato dai cambiamenti stagionali.....È importante sottolineare che, utilizzando test patogeni, abbiamo scoperto che il virus può riprodursi e causare malattie nei ratti da latte, e ulteriori studi hanno dimostrato che delle particelle simili a virus possono essere osservate nel cervello dei ratti da latte mediante microscopia elettronica. Pertanto, questo studio ha aumentato la nostra comprensione della diversità genetica degli SL-COV trasportati dai pipistrelli e ha anche fornito una nuova prospettiva per studiare la possibilità di trasmissione tra specie di SL-COV utilizzando ratti da latte come modello animale.*".

E' stata avanzata l'ipotesi che ci possa essere una relazione di trasmissione del SARS-CoV dai pipistrelli ai pangolini e da questi ultimi agli esseri umani.

In "**Molecular evolution and phylogenetic analysis of SARS-CoV-2 and hosts ACE2 protein suggest Malayan pangolin as intermediary host**" (L'evoluzione molecolare e l'analisi filogenetica della SARS-CoV-2 e delle proteine ACE2 suggeriscono il pangolino malese come ospite intermedio), Braz J Microbiol. 2020 Dec; 51(4):1593-1599. Published: **26 June 2020**. PMID: [PMC7319214](#) DOI: [10.1007/s42770-020-00321-1](#), gli autori concludono che "*...il pangolino è diventato un ospite opportuno che agisce da mediatore per l'ingresso e il salto di SARS-CoV-2 dal pipistrello all'uomo.*".

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**" [Un aggiornamento sull'origine del SARS-CoV-2: Nonostante l'identità vicina, i coronavirus derivati dal pipistrello (RaTG13) e dal pangolino variavano nel sito critico di legame e nei residui di glicani legati all'O], J Med Virol. **2020 Jul 7**;10.1002/jmv.26261. doi: [10.1002/jmv.26261](#). PMID: [PMC7361880](#) DOI: [10.1002/jmv.26261](#), gli autori avanzano dei dubbi e affermano che "*...le rotte di trasmissione della sorgente iniziale di SARS-CoV-2 rimangono oscure e controverse.....Solo la Serina (Ser) nel pangolino e sia la Treonina (Thr) che la Serina (Ser) glicani legati all'O sono stati osservati in RaTG13, suggerendo che uno studio dettagliato relativo a coV è necessario nel Pangolino (Manis javanica) e nel Pipistrello (Rhinolophus affinis).*".

Al contrario, in **“Conservation analysis of SARS-CoV-2 spike suggests complicated viral adaptation history from bat to human”** (L’analisi di conservazione del SARS-CoV-2 spike suggerisce una complicata storia di adattamento virale dal pipistrello all’uomo), *Evol Med Public Health*. **2020 Nov 5**; 2020(1):290-303. PMID: [PMC7665476](https://pubmed.ncbi.nlm.nih.gov/33396801/) DOI: [10.1093/emph/eoaa041](https://doi.org/10.1093/emph/eoaa041) si conclude *“...che la sequenza spike del SARS-CoV-2 è il risultato di molteplici eventi di ricombinazione durante la sua trasmissione dal pipistrello all’umano, e proponiamo un quadro di storia evolutiva che risolve la relazione dei coronavirus BCoV-RaTG13 e il pangolino con SARS-CoV-2.”*

Per gli autori di **“Computational Analysis of SARS-CoV-2 and SARS-Like Coronavirus Diversity in Human, Bat and Pangolin Populations”** (Analisi computazionale della diversità del coronavirus SARS-CoV-2 e simile alla SARS nelle popolazioni umane, di pipistrelli e pangolini) **2020 Dec 30**;13(1):E49. PMID: 33396801 DOI: [10.3390/v13010049](https://doi.org/10.3390/v13010049), *“...Le origini evolutive del virus rimangono sfuggenti e la comprensione delle sue complesse firme mutazionali potrebbe guidare la progettazione e lo sviluppo del vaccino.....abbiamo caratterizzato diverse varianti ad alto impatto (inserimento/cancellazione in-frame o stop gain) nelle popolazioni bat-CoV e pangolin-CoV, alcune delle quali si trovano nella stessa posizione di amminoacidi e possono evidenziare loci di potenziale rilevanza funzionale.”*

Infine, in **“Novel coronavirus disease (COVID-19) pandemic: A recent mini review”** [Nuova pandemia di coronavirus (COVID-19): una recente mini recensione], *Comput Struct Biotechnol J*. **2020 Dec 31**. PMID: [PMC7773542](https://pubmed.ncbi.nlm.nih.gov/33396801/) DOI: [10.1016/j.csbj.2020.12.033](https://doi.org/10.1016/j.csbj.2020.12.033), per gli autori *“...L’analisi filogenetica dei genomi suggerisce che la possibile comparsa di ricombinazione tra Cov simili alla Sars da Pangolino e Pipistrello potrebbe aver portato all’origine del Sars-Cov-2 e dell’epidemia della Covid-19...”* e completano dicendo che *“...L’infezione da SARS-CoV-2, con una pesante carica virale nel corpo, distrugge i polmoni umani attraverso la tempesta di citochine, specialmente nelle persone anziane e nelle persone con disturbi immunosoppressi. Un certo numero di farmaci sono stati riutilizzati e impiegati, ma ancora, nessun medicinale antivirale specifico è stato approvato dalla FDA per trattare questa malattia. Questa recensione fornisce uno stato attuale del COVID-19, epidemiologia, una panoramica della filogenesi, delle modalità di azione, della diagnosi e dei possibili metodi di trattamento e vaccini.”*

In conclusione, **allo stato attuale, non sembra esistere ancora nessuna certezza sulla veridicità dell’ipotesi più accreditata della relazione di trasmissione del SARS-CoV dai pipistrelli ai pangolini e da questi ultimi agli esseri umani.**

3.2 IL GENE RdRp E LA SEQUENZA DI RNA “GTGAAATGGTCATGTGTGGCGG” DEL SARS-COV-2

Il gene RdRp [o *RNA-dependent RNA polymerase* (in italiano *RNA polimerasi RNA-dipendente*) del SARS-CoV-2, è parte di *ORF1ab* (Open Reading Frame, *quadro* o *cornice di lettura aperto*, fase di lettura che consente di codificare un'intera proteina, senza incontrare codoni di stop prematuri e quindi formare una proteina tronca - https://it.wikipedia.org/wiki/Open_reading_frame) è specifico del SARS-CoV-2 (<https://it.wikipedia.org/wiki/SARS-CoV-2>), mentre il gene "E" (proteina dell'involucro del SARS-CoV-2) appartiene a tutti i *beta coronavirus* (https://it.wikipedia.org/wiki/Test_diagnostici_per_la_COVID-19).

Nelle Figure 3.1, 3.2 e 3.3 (presentate nelle pagine successive) sono mostrate **sei sequenze ottenute dal 2019-nCoV** che vengono utilizzate come controllo positivo del tampone rinofaringeo

(<https://www.eurosurveillance.org/content/table/10.2807/1560-7917.ES.2020.25.3.2000045.t1?fmt=ahah&fullscreen=true>).

L'obiettivo di ogni **virus** è quello di replicarsi il più possibile. I virus realizzano questo scopo mediante l'azione della **RNA polimerasi** (o *enzima che forma lunghe catene di RNA*). Nell'infezione del SARS-CoV-2, durante il processo denominato di *trascrizione*, il genoma del virus è replicato (copiato) attraverso la **RNA polimerasi-RNA dipendente (RdRp)**.

Per i motivi sopra esposti, del Gene RdRp (*pSARS-F*), è stata selezionata la sequenza di **22 basi di RNA** “GTGAAATGGTCATGTGTGGCGG” (indicata dalla freccia verde in Fig. 3.1), attualmente utilizzata, assieme ad altre sequenze, come controllo positivo (nel tampone rinofaringeo) della presenza dell'infezione da SARS-CoV-2.

Tutto il lavoro di approfondimento presentato nei capitoli successivi riguarda SOLO questa sequenza e le sue molteplici modificazioni.

In fase iniziale, allo scopo di identificare i suoi diversi *allineamenti significativi* e “prodotti”, la sequenza è stata oggetto di **una Ricerca BLAST completa** (si veda il Capitolo IV).

Successivamente sono state eseguite **due Ricerche BLAST complete**, rispettivamente sulle sequenze “GTGAAATGGT**T**ATGTGTGGCGG” (sequenza di RNA del Gene RdRp del “*Bat coronavirus BM48-31/BGR/2008, complete genome*”, si veda il Capitolo V) e “GTGA**G**ATGGTCATGTGTGGCGG” (sequenza di RNA del Gene RdRp del “*Bat SARS-like coronavirus isolate bat-SL-CoVZC45, complete genome*”, si veda il Capitolo V), indicate dalle due frecce rosse e da quella blu in Fig. 3.1. Queste due sequenze differiscono dalla sequenza “GTGAAATGGTCATGTGTGGCGG” **SOLO** per una base (rispettivamente, la **T** al posto della **C** e la **G** al posto della **A**). Come accennato nel paragrafo 2.2 del Capitolo II e come vedremo nel Capitolo V, queste sequenze sono importanti perché la semplice variazione di queste basi (indicate in rosso nelle due sequenze) permette di evidenziare *allineamenti significativi* con specifici “organismi” e fare ipotesi sulle loro relazioni, in particolare tra il *Pipistrello*, il *Pangolino* e l'*Essere Umano*.

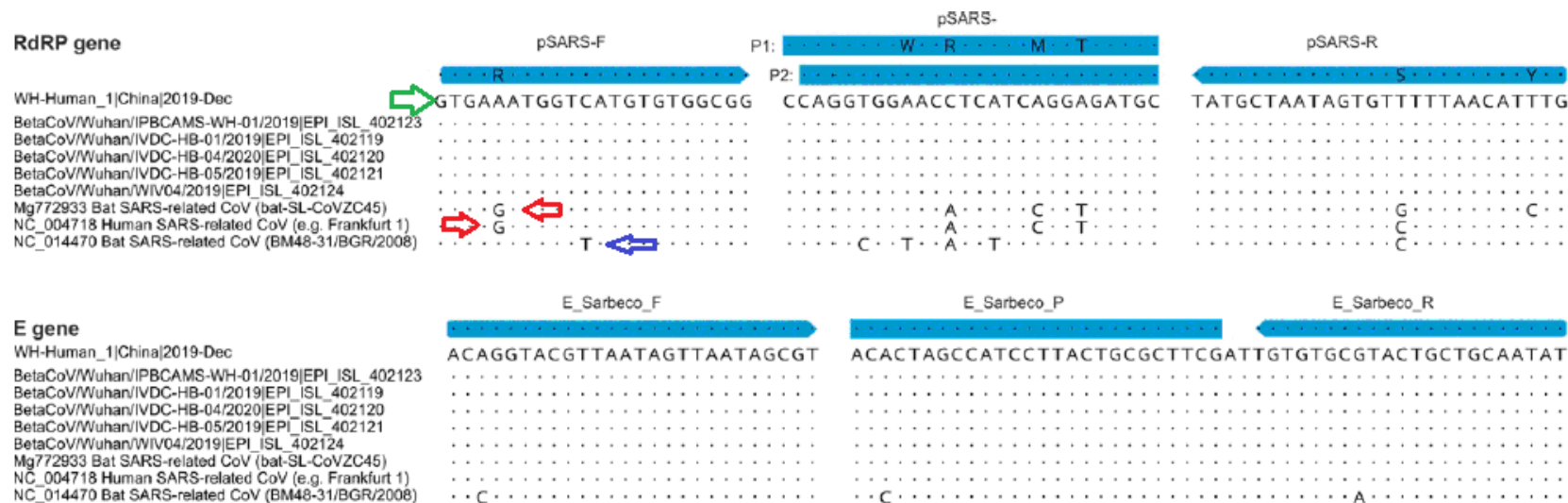


Figure 2 Partial alignments of oligonucleotide binding regions. Panels show six available sequences of 2019-nCoV, aligned to the corresponding partial sequences of SARS-CoV strain Frankfurt 1, which can be used as a positive control for all three RT-PCR assays. The alignment also contains the most closely-related 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:

RdRP_SARSr-P2

RdRP_SARSr_Oligos

WH-Human_1 | China | 2019-Dec
 BetaCoV/Wuhan/IPBCAMS-WH-01/2019 | EPI_ISL_402123
 BetaCoV/Wuhan/IVDC-HB-01/2019 | EPI_ISL_402119
 BetaCoV/Wuhan/IVDC-HB-04/2020 | EPI_ISL_402120
 BetaCoV/Wuhan/IVDC-HB-05/2019 | EPI_ISL_402121
 BetaCoV/Wuhan/WIV04/2019 | EPI_ISL_402124
 NC_004718 (SARS coronavirus, complete genome)
 DQ022305 (Bat SARS coronavirus HKU3-1, complete genome)
 AB889995 (SARS bat coronavirus RdRp gene for RNA dependent RNA polymerase, p...
 AB889998 (SARS bat coronavirus RdRp gene for RNA dependent RNA polymerase, p...
 DQ412043 (Bat SARS coronavirus Rm1, complete genome)
 JX993987 (Bat coronavirus Rp/Shaanxi2011, complete genome)
 KF294442 (SARS-related bat coronavirus isolate Longquan-4 RNA-dependent RNA pol...
 KF294455 (SARS-related bat coronavirus isolate Anlong-111 orf1ab polyprotein and ...
 KF294456 (SARS-related bat coronavirus isolate Jiyuan-331 orf1ab polyprotein gene, ...
 KJ473811 (BtRF-BetaCoV/JL2012, complete genome)
 KJ473813 (BtRF-BetaCoV/SX2013, complete genome)
 KJ473814 (BtRS-BetaCoV/HuB2013, complete genome)
 KU973690 (SARS-related coronavirus isolate F29 RdRP mRNA, partial cds)
 MG772846 (Bat SARS-like coronavirus isolate bat-SL-CoVnCoV80 RNA-dependent RNA...
 MG772849 (Bat SARS-like coronavirus isolate bat-SL-CoVnCoV66 RNA-dependent RNA...
 MG772862 (Bat SARS-like coronavirus isolate bat-SL-CoVZ2_113 RNA-dependent RNA...
 MG772867 (Bat SARS-like coronavirus isolate bat-SL-CoVZ2_99 RNA-dependent RNA...
 MG772870 (Bat SARS-like coronavirus isolate bat-SL-CoVZ2_76 RNA-dependent RNA...
 MG772879 (Bat SARS-like coronavirus isolate bat-SL-CoVZ2_46 RNA-dependent RNA...
 MG772886 (Bat SARS-like coronavirus isolate bat-SL-CoVZ2_8 RNA-dependent RNA p...
 MG772891 (Bat SARS-like coronavirus isolate bat-SL-CoVZ2_1 RNA-dependent RNA p...
 MG772903 (Bat SARS-like coronavirus isolate bat-SL-CoVZXC44 RNA-dependent RNA...
 MG772933 (Bat SARS-like coronavirus isolate bat-SL-CoVZXC45, complete genome)
 MG772934 (Bat SARS-like coronavirus isolate bat-SL-CoVZXC21, complete genome)
 KY352407 (Severe acute respiratory syndrome-related coronavirus strain BtKY72, co...
 NC_014470 (Bat coronavirus BM48-31/BGR/2008, complete genome)
 KC633199 (Betacoronavirus BtCoV/Rhi_fer/It17/ITA/2009 RNA-dependent RNA polym...
 KC633200 (Betacoronavirus BtCoV/Rhi_fer/It2/ITA/2009 RNA-dependent RNA polym...
 KC633202 (Betacoronavirus BtCoV/Rhi_eur/BB99-04/BGR/2009 RNA-dependent RNA...
 KC633203 (Betacoronavirus BtCoV/Rhi_eur/BB98-98/BGR/2008 RNA-dependent RNA...
 KC633206 (Betacoronavirus BtCoV/Rhi_fer/It15/ITA/2009 RNA-dependent RNA polym...
 KC633211 (Betacoronavirus BtCoV/Rhi_hip/R07-09/SPA/2010 RNA-dependent RNA p...
 KC633213 (Betacoronavirus BtCoV/Rhi_hip/R45-03/SPA/2010 RNA-dependent RNA p...
 KC633220 (Betacoronavirus BtCoV/Rhi_hip/Slo52/SLO/2009 RNA-dependent RNA pol...

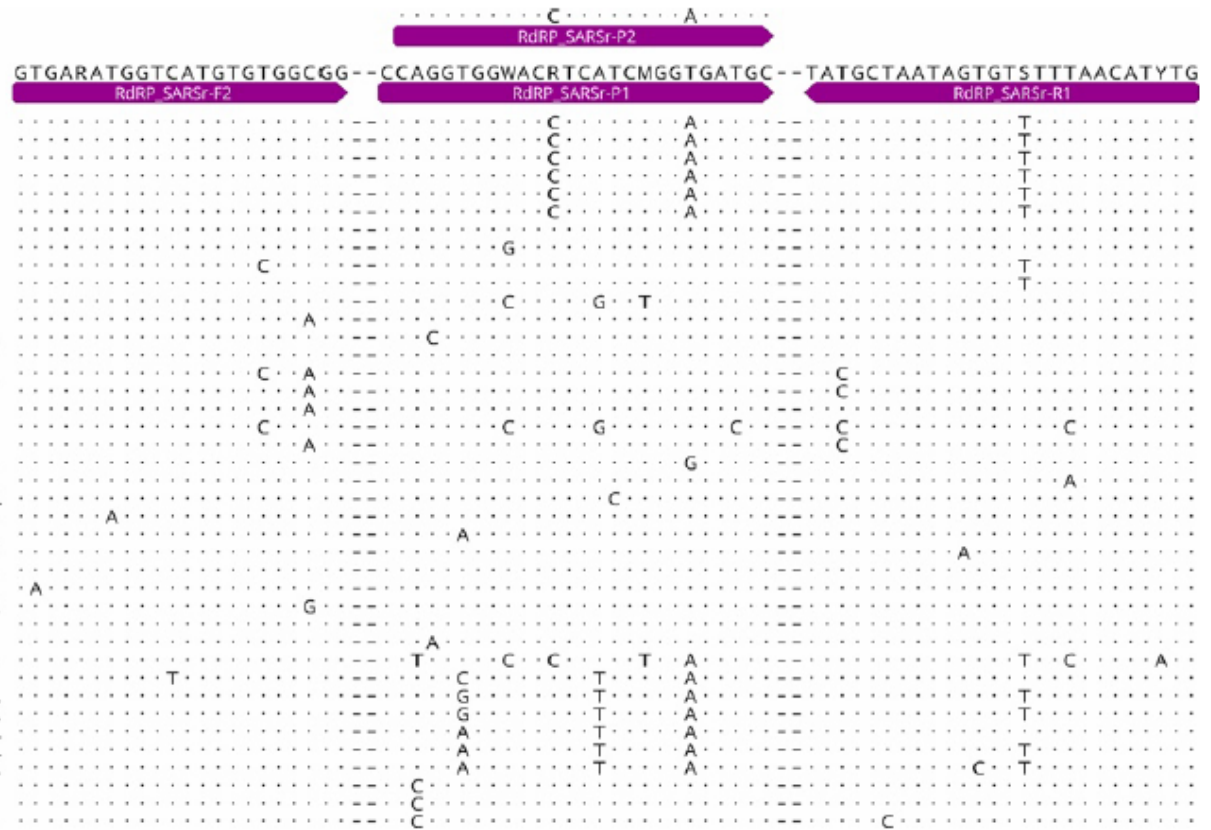


Fig. 3.2

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

Per non trascurare l’impatto di eventuali altre variazioni di basi azotate nelle stesse posizioni evidenziate in blu nella sequenza “GTGAAATGGTCATGTGTGGCGG” ed identificare eventuali altri “organismi” rilevanti per le finalità di questo studio, sono state eseguite **tutte le sostituzioni di basi possibili**, ottenendo come risultato le seguenti altre **4 sequenze**:

“GTGAAATGGTAATGTGTGGCGG”, “GTGAAATGGTGATGTGTGGCGG”

“GTGATATGGTCATGTGTGGCGG”, “GTGACATGGTCATGTGTGGCGG”

Queste sequenze sono state anch’esse oggetto di **quattro Ricerche BLAST complete** (si veda il Capitolo VI).

Effettuato questo laborioso lavoro preliminare di sostituzioni di basi per identificare tutti gli “organismi” e “prodotti” dipendenti da tali modifiche della sequenza “GTGAAATGGTCATGTGTGGCGG”, iniziano le modificazioni di questa stessa sequenza attraverso la nuova metodologia proposta in questo scritto, quella basata sui principi della **TEORIA TRICROMATICA DELL’EQUILIBRIO DEI SISTEMI (T.T.E.S.)** e presentata nei Capitoli VIII e IX.

Questa metodologia prevede la **sostituzione di almeno il 70% delle basi** (ma si può arrivare anche al 100%), ottenendo così delle **sequenze quasi o totalmente nuove**.

Come vedremo nei Capitoli VIII e IX, la sostituzione delle basi deve essere effettuata necessariamente seguendo i “Trend Non Manifesti” della sequenza originaria, la sequenza “GTGAAATGGTCATGTGTGGCGG”.

Il modello teorico qui proposto prevede che **tutte le nuove sequenze** generate dalla **sequenza originaria** avranno una *stretta relazione “funzionale”* con la **sequenza originaria** perché generate nel rigoroso rispetto dei suoi numerosi “Trend Non Manifesti”.

Tutte le **nuove sequenze generate** saranno poi sottoposte a *Ricerche Blast* e utilizzate per tentare di verificare l’ipotesi più accreditata della relazione di trasmissione del SARS-CoV dai *pipistrelli* ai *pangolini* e da questi ultimi agli *esseri umani*.

Tale ipotesi potrebbe essere maggiormente plausibile se tutte o la maggior parte delle **nuove sequenze** generate dalla **sequenza originaria** presentassero *allineamenti significativi* con gli “organismi” attesi, ossia il *Pipistrello* e/o il *Pangolino* o, in alternativa, con altri “organismi intermedi”, come quelli identificabili dalle modificazioni di una SOLA base della sequenza “GTGAAATGGTCATGTGTGGCGG”.

Come già anticipato nel **Riassunto** e nella **Premessa**, questo lavoro di modificazioni della **sequenza originaria** a riservato sorprese inaspettate e clamorose, in particolare la presenza di *allineamenti significativi* corrispondenti ad una sequenza dalle caratteristiche sconosciute del *Virus dell’Immunodeficienza Umana (HIV-1)* e quelli corrispondenti a diverse sequenze del *Bacillo dell’Antrace (Bacillus Anthracis)*, un batterio gram-positivo infettivo, agente eziologico del *carbonchio ematico* nell’animale e dell’*antrace* nell’uomo.

CAPITOLO IV

**Ricerca BLAST sulla
SEQUENZA di RNA
“GTGAAATGGTCATGTGTGGCGG”
del Gene RdRp
(RdRp_SARSr-F)
(WH-Human_1|China|2019-Dec)**

OBIETTIVI DEL CAPITOLO IV

In questo capitolo è presentata *la Ricerca Blast* completa eseguita sulla sequenza di **22 basi di RNA** “GTGAAATGGTCATGTGTGGCGG”, sequenza tratta dal **Gene RdRp** (*pSARS-F*) e attualmente utilizzata, insieme ad altre sequenze, come controllo positivo (nel tampone rinofaringeo) della presenza dell’infezione da *SARS-CoV-2*.

Come già è stato detto in precedenza, il **Gene RdRp** ("RNA-dependent RNA polymerase") è parte del gene “**ORF1ab**”, lettura aperta frame 1ab di *SARS-CoV-2*.

Il **Gene RdRp** è importante perché mediante l’azione della **RNA polimerasi RNA dipendente (RdRp)** il virus (il *SARS-CoV-2*) responsabile della **Covid-19** realizza la sua replicazione (copia di se stesso).

La **RNA polimerasi RNA dipendente (RdRp)** è quindi il motore del processo di **trascrizione** del *SARS-CoV-2*.

Sebbene vi siano delle differenze, “*l’RNA polimerasi svolge essenzialmente la stessa funzione in tutte le cellule dai batteri all’uomo..... la trascrizione inizia quando l’RNA polimerasi si lega ad una regione speciale, il promotore, all’inizio del gene.....fino a raggiungere un terminatore....Il prodotto immediato della trascrizione si chiama trascritto primario e consiste di un RNA che si estende dal promotore al terminatore ma è molto instabile...*” (<https://www.tesionline.it/appunti/Biologia-molecolare/Le-RNA-polimerasi-e-il-ciclo-della-trascrizione/255/75>).

Uno dei modi possibili per impedire al virus di replicarsi è quello di impedire all’**Rna polimerasi Rna dipendente (RdRp)** di svolgere adeguatamente la sua funzione. Tecnicamente questo è possibile introducendo delle **sequenze di Rna modificato**, così da sabotare la fase di **trascrizione** virale. Tuttavia, il *SARS-CoV-2* è capace di correggere gli **errori in fase di trascrizione** determinati dalla rilevazione di **Rna modificato** e proseguire indisturbato la replicazione.

Quindi, per raggiungere l’obiettivo di sabotare la **RdRp** è necessario un sofisticato processo di mascheramento dell’**Rna modificato** o, in alternativa, sviluppare **potenti inibitori chimici** (farmaci) della **RdRp**.

4.1 Ricerca BLAST sulla SEQUENZA DI RNA “GTGAAATGGTCATGTGTGGCGG” del Gene RdRp (RdRp_SARSr-F) – (Hitlist size: 100)

Per i motivi esposti nella pagina precedente e nel paragrafo 3.2 del Capitolo III, del **Gene RdRp** (*pSARS-F*) è stata selezionata la sequenza di **22 basi di RNA** “GTGAAATGGTCATGTGTGGCGG”. Su questa sequenza è stata eseguita una **Ricerca Blast completa** e, per approfondimenti, è stata selezionata la prima sequenza della lista dei primi **100 allineamenti significativi** riscontrati (quella evidenziata in grassetto).

PER LE INFORMAZIONI UTILI AD INTERPRETARE I DATI PRESENTATI NELLE PAGINE SUCCESSIVE SI VEDA IL CAPITOLO 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	Comomn 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

MW411887.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/VA-DCLS-2266/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	44.1	44.1	100%	0.037	100.00%	29869
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	Severe acute	44.1	44.1	100%	0.037	100.00%	29858

MW411877.1	coronavirus 2 isolate SARS-CoV-2/human/USA/CA-LACPHL-AF00014/2020, complete genome Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/VA-DCLS-2253/2020, complete genome	respiratory syndrome coronavirus 2 Severe acute respiratory syndrome coronavirus 2	44.1	44.1	100%	0.037	100.00%	29893
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

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-2/human/USA/VA-DCLS-2325/2020, complete genome	Severe acute respiratory syndrome coronavirus 2	44.1	44.1	100%	0.037	100.00%	29868
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

VISTA DI ALLINEAMENTO VERTICALE AGGANCIATO A QUERY CON LETTERE D'IDENTITA'

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
HLKDGTCGLVEVEKGVLPQLEQPYVFIKRS DARTAPHGVMVELVAELEGIQYGRSGE
TLGVLVPHVGEIPVAYRKVLLRKNKNGKAGGHSYGADLKSFDLGDELGTDPYEDFQEN
WNTKHSSGVTRELMRELNGGAYTRYVDNNFCGPDGYPLECIKDLLARAGKASCTLSEQ
LDFIDTKRGVYCCREHEHEIAWYTERSEKSYELQTPFEIKLAKKFDIFNGECPNFVFP
LNSIIKTIQPRVEKKKLDGFMGRIRSVYPVSPNECNQMCLSTLMKCDHCGETSWQTG
DFVKATCEFCGTENLTKEGATTCGYLPQNAVVKIYCPACHNSEVGPESHSLAEYHNESG
LKTILRKGGRTIAFGGCVFSYVGCNKCAYWVPRASANIGCNHTGVVGESEGLNDNL
LEILQKEKVNINIVGDFKLNEEIAIILASFSASTSAFVETVKGLDYKAFKQIVESCNG
FKVTKGKAKKGAWNIGEQKSI LSPLYAFASEAARVRSIFSR TLETAQNSVRVLQKAA
ITILDGISQYSLRLIDAMMFTSDLATNNLVVMAYITGGVVQLTSQWL TNIFGTVYEKL

KPVLDWLEEKFKEGVEFLRDGWEIVKFI STCACEIVGGQIVTCAKEIKESVQTFKFLV
NKFLALCADS I I IGGAKLKALNLGETFVTHSKGLYRKCVCKSREETGLLMLPKAPKEI I
FLEGETLPTPEVLTEEVVLKTDGLQPLEQPTSEAVEAPLVGTPVCINGLMLEIKDTEK
YCALAPNMMVTNNTFTLKGGAFTKVTFGDDTVIEVQGYKSVNITFELDERIDKVLNEK
CSAYTVELGTEVNEFACVVADAVIKTLQPVSELLTPLGIDLDEWSMATYYLFDSEGEF
KLASHMYCSFYPPDEDEEEGDCEEEEFEPSTQY EYGTEDDYQGKPLEFGATSAAALQPE
EEQEEDWLDDDSQQTVGQQDGEDNQTTTIQTIVEVQPQLEMELTPVVQTI EVNSFSG
YLKLTDNVYIKNADIVEEAKVKPTVVVNAANVYLKHGGGVAGALNKATNNAMQVESD
DYIATNGPLKVGGSVLSGHNLA KHCLHVVGPNVNGEDIQLLKSAYENFNQHEVLLA
PLLSAGIFGADPIHSLRVCVDTVR TNVYLAVFDKNLYDKLVSSFLEMKSEKQVEQKIA
EIPKEEVKPFITESKPSVEQRKQDDKKIKACVEEVTTTLEETKFLTENLLLYIDINGN
LHPDSATLVSDIDITFLKKDAPYIVGDVVQEGVLTAVVIPTKKAGGTTEMLAKALRKV
PTDNYITTYPGQLNGYTV EEAKTVLKCKSAFYIILPSIISNEKQEILGTVSWNLREM
LAHAEE TRKLM PVCVETKAI VSTIQ RKYKGIKI QEGVVDYGARFYFYT SKTTVASLIN
TLNDLNETLVTMPLGYVTHGLNLEEAARYMRS LKVPATVSVSSPDAVTAYNGYLTSSS
KTPEEHFIETISLAGSYKDWSSGQSTQLGIEFLKRGDKSVYYT SNPTTFHLDGEVIT
FDNLKTL LSLREVRTIKVFTTVDNINLHTQVVDMSMTYGGQFGPTYLDGADVTKIKPH
NSHEGKTFYVLPND DTLRVEAFEYHYHTDPSFLGRYMSALNHTKKWKYPQVNGLTSIK
WADNNCYLATAALLTLQQIELKFNPPALQDAYRARAGEAANFCALILAYCNKTVGELG
DVRETMSYLFQHANLDSCKRVLNVVCKTCGQQQTTLKGEAVMYMGTLSYEQFKKG VQ
IPCTCGKQATKYL VQQESPFVMSAPPAQYELKHGTFTCASEYTGNYQCGHYKHITSK
ETLYCIDGALLTKSSEYKGPITDVFYKENSYTTTIKPVTYKLDGVVCTEIDPKLDNYY
KKDNSYFTEQPIDLVPNQPYPNASFDNFK FVCDNIK FADDLNQLTG YKKPASRELKVT
FFPDLNGDVVAIDYKHYTPSFKKGAKLLHKPIVWHVNNATNKATYKPTWCIRCLWST
KPVETSNSFDVLKSEDAQGM DN LACEDLKPVSEEVVENPTIQKDVLECNVKTTEVVD
IILK PANNSLKITEEVGHTDLMAAYVDNSSLTIKKPNELSRVGLKTLATHGLAAVNS
VPWDTIANYAKPFLNKV VSTTNIVTRCLNRVCTNYMPYFFTL LLLQLCTFTRSTNSRI
KASMP TTI AKNTVKS V GKFCLEASFNYLKS PNF SKLINI I IWFLLLSVCLGSLIYSTA
ALGVLM SNLGMPSYCTGYREGYLNSTNV TIATYCTGSI PCSVCLSGLDSDLTYP SLET
IQITISSFKWDLTAFGLVAEWF LAYILFTRFFYV LGLAAIMQLFFSYFAVHFISNSWL
MWLIINLVQMAPISAMVRMYIFFASFYVWKS YVHVVDGCNSSTCMMCYKRNRATRVE
CTTIVNGVRRSFYVYANGGKGFCKLHNWNCV NCDTFCAGSTFISDEVARDLSLQFKRP
INPTDQSSYIVDSVTVKNGSIHLYFDKAGQKTYERHSLSHFVNLDNLRANNTKGS LPI
NVIVFDGKSKCEESSAKSASVYYSQ LMCQPI LLLDQALVSDVGDSA EVAVKMF DAYVN
TFSSTFNVP IEK LKTLVATAEAEALAKNVSLDNV LSTFISAARQGFVDS DVETKDVVEC
LKLSHQSDIEVTGDSCNNYMLTYNKVENMTPRDLGACIDCSARHINAQVAKSHNIALI
WNVKDFMSLSEQLRQIRSA AKNNLPFKLTCATTRQVNVVVTTKIALKGGKIVNNWL
KQLIKVTLVFLFVA AIFYLITPVHVM SKHTDFSSEIIGYKAIDGGVTRDIASDTCF A
NKHADFDTFWSQRGGSYTN DKACPLIAAVITREVG FVVPGLPGTILRTTNGDFLHFLP
RVFSAVGNICYTPSKLIEYTD FATSACVLAAECTIFKDASGKVPYCYDTNVLEGSVA
YESLRDPTRYV LMDGSI IQFPNTYLEG SVRVVTTFDSEYCRHGT CERSEAGVCVSTSG
RWVLNNDY YRSLPGVFCGVDAVNLLTNMFTPLIQPIGALDISASIVAGGIVAI VVTCL
AYYFMRFRRAFGEYSHVAVFNTL LFLMSFTVLCLTPVYSFLPGVYSVIYLYLTFYLTN
DVSFLAHIQWVMFTPLVPFWIT IAYI ICISTKH FYWFFSNYLKRRVVFNGVSFSTFE

EAALCTFLLNKEMYLKLRSVDVLLPLTQYNRYLALYNKYKYFSGAMDTTSYREAACCHL
AKALNDFSNSGSDVLYQPPQTSITSAVLQSGFRKMAFSPGKVEGCMVQVTCGTTTLNG
LWLDVVVYCPRHVICSEDMLNPNYEDLLIRKSNHNFLVQAGNVQLRVIGHSMQNCVL
KFKVDTANPKTPKYKFVRIQPGQTFSVLACYNGSPSGVYQCAMPNFTIKGSFLNGSC
GSVGFNIDYDCVSFCYMHMELPTGVHAGTDLEGNFYGPFVDRQTAQAAGDTTITVN
VLAWLYAAVINGDRWFLNRFTTTLNDFNLVAMKYNIEPLTQDHVDILGPLSAQTGIAV
LDMCASLKELLQNGMNGRTILGSALLEDEFTPFVDRQCSGVTFQSAVKRTIKGTHHW
LLLTILTSLLVLVQSTQWSLFFFLYENAFLPFAMGI IAMSAFAMMFVKHKAFLCLFL
LPSLATVAYFNMVMPASWVMRIMTWLDMVDTSLSGFKLKDCVMYASAVVLLILMTAR
TVYDDGARRVWTLMNVLTLVYKVVYGNALDQAISMWALII SVTSNYSGVVTTVMFLAR
GIVFMCVEYCPIFFITGNTLQCIMLVYCFGLYFCTCYFGLFCLLNRYFRLTLGVYDYL
VSTQEFRYMNSQGLLPPKNSIDAFKLNKLLGVGKPKIKVATVQSKMSDVKCTSVVL
LSVLQQLRVESSSKLWAQCVQLHNDILLAKDTTEAFEKMSVLLSVLLSMQGAVDINKL
CEEMLDNRATLQAIASEFSSLSYAAFATAQEAYEQAVANGDSEVVLKLLKSLNVAK
SEFDRDAAMQRKLEKMAQAMTQMYKQARSEDKRAKVTSAMQTMLFTMLRKLNDALN
NIINNARDGCVPLNI I PLTTAAKLMVVI PDYNTYKNTCDGTTFTTYASALWEIQQVDA
DSKIVQLSEI SMDNSPNLAWPLIVTALRANSVAVKLNELSPVALRQMSCAAGTTQTA
CTDDNALAYYNTTKGGRFVLALLSDLQDLKWARFPKSDGTGTIYTELEPPCFVTDTP
KGPVKYLYFIKGLNLRGMVLSLAATVRLQAGNATEVPANSTVLSFCAFAVDAK
AYKDYLASGGQPI TNCVKMLCTHTGTGQAI TVTPEANMQESFGGASCCLYCRCHIDH
PNPKGFCDLKGYVQI PTTCCANDPVGFTLKNVTCTVCGMWKGYGCSCDQLREPMLQSA
DAQSFLNRVCGVSAARLTPCGTGTS TDVVYRAFDI YNDKVAGFAKFLKTNCCRFQEKD
EDDNLIDSYFVVKRHTFSNYQHEETIYNLLKDCPAVAKHDFKFRIDGMVPHISRQR
LTKYTMADLVYALRHFDEGNCDTLKEILVTYNCCDDYFNKKDWYDFVENPDILRVYA
NLGERVVRQALLKTVQFC DAMRNAGIVGVLTLDNQLNGNWDYDFGDFIQTTPGSGVPVV
DSYYSLLMPI LTLTRALTAESHVDTDLTKPYIKWDLKLYDFTEERLKLFDRYFKYWDQ
TYHPNCVNCDDRCILHCANFNVLVSTVFPLTSFGPLVRKIFVDGVFPVVSTGYHFRE
LGVVHNQDVNLHSSRLSFKELLVYAADPAMHAASGNLLLDKRTTCF SVAALTNVAFQ
TVKPGNFNKDFYDFAVSKGFKEGSSVELKHFFFAQDGNAAISDYDYRYNLPTMCDI
RQLLFVVEVDKYDFCYDGGCINANQVIVNNLDKSAGFPFNKWKARLYYDSMSYEDQ
DA
LFAYTKRNVIPITITQMNLYAISAKNRARTVAGVSI CSTMTRQFHQKLLKSI AAT
RGATVVI GTSKFGGWHNMLKTVYSDVENPHLMGWDY PKCDRAMPNMLRIMASLVLAR
KHTTCCSLSHRFYRLANCAQVL **SEMVMCGG** SLYVKPGGTSSGDATTAYANSVFNICQ
AVTANVNALLSTDGNKIADKYVRNLQHRLYECLYRNRDVTDFVNEFYAYLRKHFSMM
ILSDDAVVCFNSTYASQGLVASIKNFKSVLYYQNNVFMS
EAKCWTETDLTKGPHEFCS
QHTMLVKQGDYVYLPYPDP SRILGAGCFVDDIVKTDGTLMIERFVSLAIDAYPLTKH
ENQEYADVFLYLQYIRKLHDEL TGHMLDMSVMLTNDNTSRWEP EFYEAMYPHTV
LQAVGACVLCNSQTS LRCGACIRRPFLCCKCCYDHVISTSHKLVLSVNPYVCNAPGCD
VTDVTLQYLGLGMSY YCKSHKPPISFPLCANGQVFLYKNTCVGSDNVTDFNAIATCDW
TNAGDYILANTCTERLKLFAAETLKATEETFKLSYGIATVREVLSDRELHLSWEVGKP
RPPLNRNYVFTGYRVTKNSKVQIGEYTFEKGDYGDVVYRGTTTTYKLVNGDYFVLTSH
TVMPLSAPTLPQEHYVRITGLYPTLNISDEFSSNVANYQKVGMMQKYSTLQGGPGTGK

SHFAIGLALYYPSARIVYTACSHAAVDALCEKALKYLPIDKCSRIIPARARVECFDKF
KVNSTLEQYVFCTVNALPETTADIVVFDEISMATNYDLSVVNARLRAKHVYVIGDPAQ
LPAPRTLLTKGTLEPEYFNVCRLMKTIGPDMFLGTCRRCPAEIVDTVSALVYDNKLK
AHKDKSAQCFKMFYKGVITHDVSSAINRPQIGVVREFLTRNPAWRKAVFISPYNSQNA
VASKILGLPTQTVDSSQGSEYDYVIFTQTTEHAHSCNVNRFNVAITRAKVGILCIMS
RDLYDKLQFTSLEIPRRNVATLQAENVNGLFKDCSKVITGLHPTQAPTHLSVDTKFKT
EGLCVDIPGIPKDMTYRRLISMMGFKMNYQVNGYPNMFITREEAIRHVRAWIGFDVEG
CHATREAVGTNLPLQLGFSTGVNVLVAVPTGYVDTPDNTDFSRVSAKPPPGDQFKHLIP
LMYKGLPWNVRIKIQMLSDTLKNLSDGVVFLWAHGFELTSMKYFVKIGPERTCCL
CDRRATCFSTASDTYACWHHSIGFDYVYNPFMIDVQQWGFQGNLQSNHDLYCQVHGNA
HVASCDAIMTRCLAVHECFVVRVDWTIEYPIIGDELKINAACRQVQHMVVKAALLADK
FPVLHDIGNPKAIKQVQADVEWKFYDAQPCSDKAYKIEELFYSYATHSKFTDGVCL
FWNCNVDRYPANSIVCRFDTRVLSNINLPGCDGGSLYVNKHAFHTPAFDKSAFVNLKQ
LPFFYYSDSPCESHGKQVSDIDYVPLKSATCITRCNLGGAVCRHHANEYRLYLDAYN
MMISAGFSLWVYKQFDTYNLWNTFTRLQSLNVAFNVVNKGHFDGQQGEVPSIINNT
VYTKVDGVDVELFENKTTLPVNVAFELWAKRNKIPVEVKILNNLGVDIAANTVIWYD
KRDAPAHISTIGVCSMTDIACKPTETICAPLTVFFDGRVDGQVDFRNARNGVLITEG
SVKGLQPSVGPQASLNGVTLIGEAVKTQFNYYKKVDGVVQQLPETYFTQSRNLQEFK
PRSQMEIDFLELAMDEFIERYKLEGYAFEHIVYGDFFSHSQLGGLHLLIGLAKRFKESF
FELEDFIPMDSTVKNYFITDAQTGSSKCVCSVIDLLDDFVEI IKSQDLSVSVKVVV
TIDYTEISFMLWCKDGHVETFPKLSQAWQPGVAMPNFKMQRMLLEKCDLQNYGD
SATLPGKIMMNVAKYTQLCQYLNTLTLAVPYNMRVIFHGAGSDKGVAPGTAVLRQWLP
TGTLLVDSDLNDFVSDADSTLIGDCATVHTANKWDLIISDMYDPKTKNVTKENDSKEG
FFTYICGFIQQKALALGGSVAIKITEHSWNADLYKLMGHFAWWTAFVTNVNASSEAF
IGCNYLKGPCEQIDGYVMHANYIFWRNTNPIQLSSYSLFDMSKFPLKLRGTAVMSLKE
GQINDMILSLLSKGRLLIRENNRVVISSDVLVNN"

mat_peptide

<1..>22

/gene="ORF1ab"

/product="RNA-dependent RNA polymerase"

ORIGIN

1 **gtgaaatggt catgtgtggc gg**

//

VISTA DELL'ALLINEAMENTO IN COPPIA (FUNZIONE CDS)

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		

NOTA 1: Prestare particolare attenzione agli aminoacidi **M** e **V** (si veda anche la Fig. 4.1) perché saranno oggetto di approfondimenti nei capitoli successivi.

NOTA 2: “**ORF1ab**”, lettura aperta frame 1ab di SARS-CoV-2, include il **Gene RdRp** (“**RNA-dependent RNA polymerase**”).

La sequenza “GTGAAATGGTCATGTGTGGCGG” (oggetto del presente studio) è parte del /product=“**RNA-dependent RNA polymerase**” ed è usata, insieme ad altre sequenze dello stesso Gene, nel controllo positivo (nel tampone rinofaringeo) del SARS-CoV-2.

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

GenBank: MW411920.1

[GenBank](#) [FASTA](#)

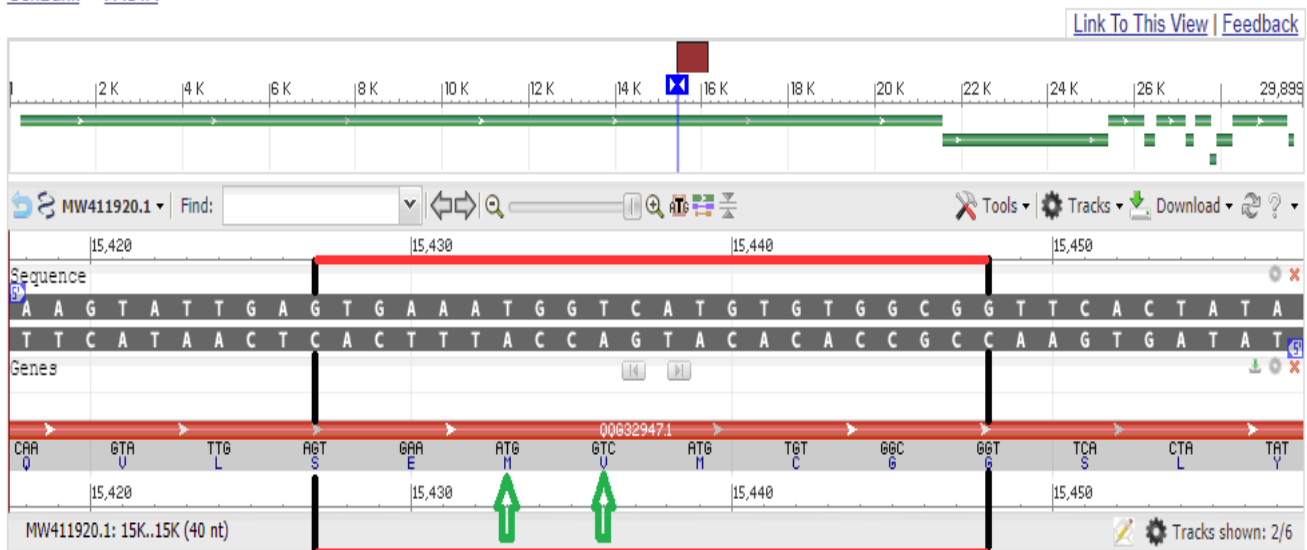


Fig. 4.1

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 orflab 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, Immunology 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="orflab"
                   /codon_start=1
                   /product="orflab polyprotein"
                   /protein_id="QIB84671.1"
                   /translation="LFAYTKRNVIPITITQMNLKYAISAKNRARTVAGVSICSTMTNRQ
FHQKLLKSIAATRGA TVVIGTSKFYGGWHNMLKTVYSDVENPHLMGWDYPKCDRAMPN
MLRIMASLV LARKHTTCCSLSHRFYRLANCAQVLSSEMVMCGGSLYVKPGGTSSGDAT
TAYANSVFNICQAVTANVNALLSTDG NKIADKYVRNLQHRLYECLYRNRD VDTDFVNE
FYAYLRKHFSMMILSDDAVVCFNSTYASQGLVASIKNFKSVLYYQNNV FMS"

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

```

ORIGIN

```

1 agaggagcta ctgtagtaat tggacaagc aaattctatg gtggttgcca caacatgta
61 aaaactgttt atagtgatgt agaaaacct caccttatgg gttgggatta tcctaaatgt
121 gatagagcca tgcctaacat gcttagaatt atggcctcac ttgttcttgc tcgcaaacat
181 acaacgtggt gtagcttgtc acaccgtttc tatagattag ctaatgagtg tgctcaagta
241 ttgagtgaaa tggtcatgtg tggcggttca ctatatgta aaccagggtgg aacctcatca
301 ggagatgcca caactgctta tgctaatagt gtttttaaca tttgtcaagc tgtcacggcc
361 aatgttaatg cacttttatc tactgatggg aacaaaattg ccgataagta tgtccgcaat
421 ttacaacaca gactttatga gtgtctctat agaaatagag atgttgacac agactttgtg
481 aatgagtttt acgcatattt gcgtaaacat ttctcaatga tg

```

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

GenBank: MT072668.1

[GenBank](#) [FASTA](#)

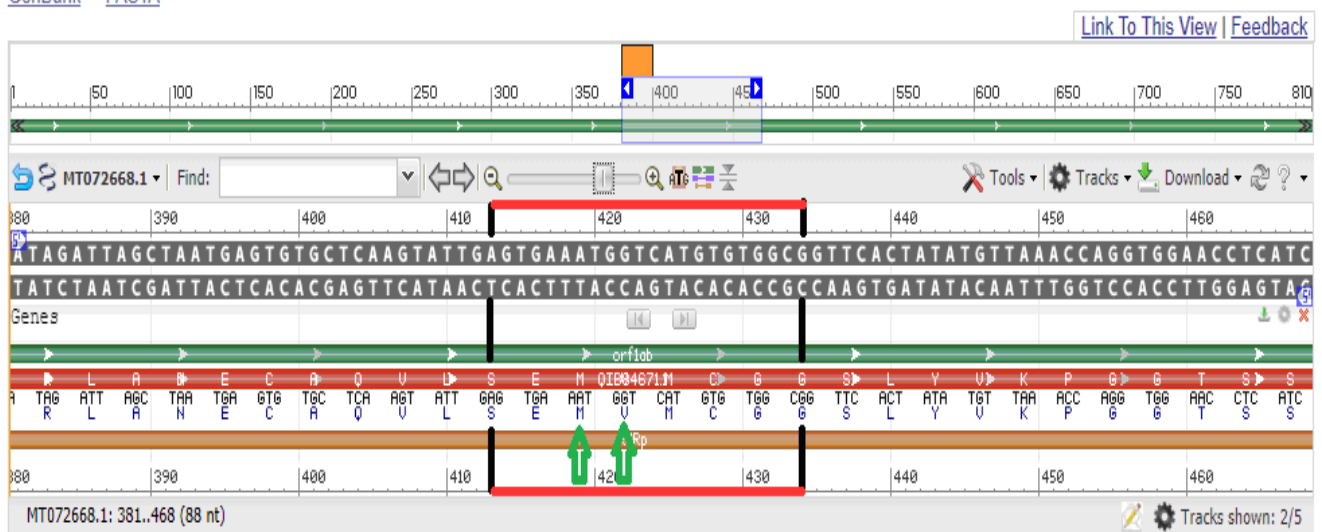


Fig. 4.2 (Attenzione: Nonostante le basi siano diverse da quelle evidenziate in Fig.4.1, gli aminoacidi sono gli stessi, **M** e **V**)

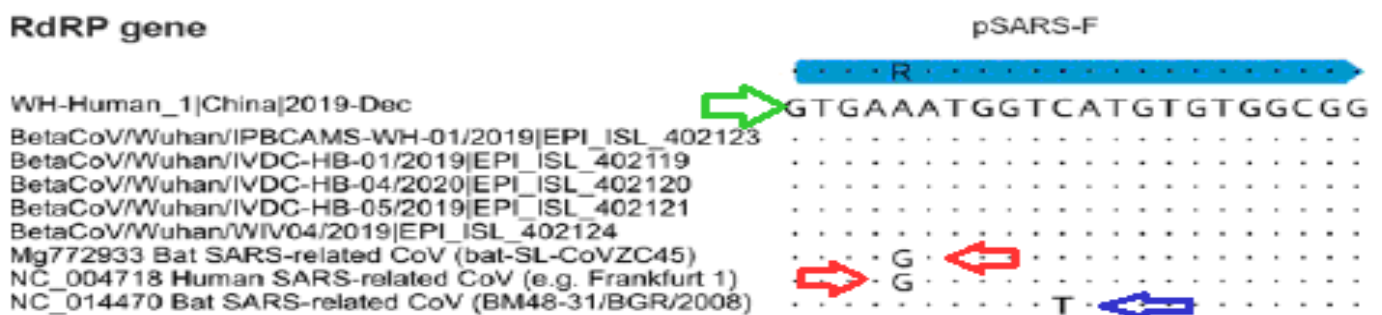
CAPITOLO V

**RICERCHE BLAST SU DUE
SPECIFICHE SEQUENZE DI
RNA DEL GENE RdRp**

OBIETTIVI DEL CAPITOLO V

In questo capitolo sono state effettuate le ricerche BLAST su **due sequenze** quasi totalmente identiche alla sequenza “GTGAAATGGTCATGTGTGGCGG” (la sequenza considerata nel paragrafo precedente e oggetto di studio di questo documento), cioè quelle che ne differiscono SOLO per una base, rispettivamente le sequenze “GTGAAATGGT**T**ATGTGTGGCGG” e “GTGAG**A**TGGTCATGTGTGGCGG”.

Come già visto nel Capitolo III in Fig. 3.1 (della quale qui di seguito riporto il quadrante interessato), queste due sequenze sono importanti perché sono le due sequenze del **gene RdRp** (*RNA polimerasi RNA dipendente*, il motore del processo di **trascrizione** del SARS-CoV-2) i cui allineamenti significativi delle ricerche BLAST identificano il *Pipistrello* e il *Pangolino*.



Più precisamente, gli “organismi” identificati dalla ricerca BLAST sulla sequenza “GTGAAATGGT**T**ATGTGTGGCGG” sono: il *Pipistrello*, il *Pangolino Malese* e quello *Cinese* (specie di formichieri) e la *Piccola Egretta* (una specie di uccello). La ricerca BLAST di questa sequenza **non evidenzia** la presenza dell’*Homo Sapiens* tra gli “organismi” identificati, tranne se la ricerca si estende fino alla mutazione della base (la **T** con la **C**) evidenziata in rosso.

Invece, gli “organismi” identificati dalla ricerca BLAST sulla sequenza “GTGAG**A**TGGTCATGTGTGGCGG” sono: l’*Homo Sapiens*; *Rhinolophus Sinicus* (chiamato *Ferro di Cavallo*), *Rhinolophus Pusillus* e *Rhinolophus Monoceros* (varianti della specie del *Pipistrello*); *Altererythrobacter namhicola* (un batterio Gram-negativo); *Cebus Imitator* (Cappuccino panamense dalla faccia bianca, una specie di scimmia); *Timema bartmani* (una specie di insetto stecco della famiglia Timematidae) e *Timema Tahoe*.

Come si può notare, il *Pipistrello* è l’unico “organismo” che le **due sequenze** hanno in comune, questo perché questa specie di mammifero può presentare nel suo genoma l’una o l’altra sequenza (“GTGAAATGGT**T**ATGTGTGGCGG” o “GTGAG**A**TGGTCATGTGTGGCGG”).

Questo risultato della ricerca BLAST indica che la modifica, più o meno rilevante, di una **sequenza originaria** può generare **nuove sequenze** i cui allineamenti significativi individuati dalla ricerca identificano **organismi comuni** non tra la **sequenza originaria** (“GTGAAATGGTCATGTGTGGCGG”) e le **nuove sequenze**, ma SOLO tra le nuove sequenze (per approfondimenti si veda anche il paragrafo 2.2 del Capitolo II).

5.1 Ricerca BLAST sulla SEQUENZA di RNA

“GTGAAATGGTTATGTGTGGCGG” del RdRp Gene del “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	Comomn 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), 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%	29576
MW290971.1	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
MW290970.1	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
MW290969.1	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
MW290968.1	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
MW290967.1	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
MW290966.1	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
MW290965.1	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
MW290964.1	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

MW290963.1	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 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%	29790
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	Severe acute respiratory syndrome coronavirus 2	36.2	36.2	100%	8.9	95.45%	29777

MW290956.1	<p>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</p> <p>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</p>	Severe acute respiratory syndrome coronavirus 2	36.2	36.2	100%	8.9	95.45%	29860
MW290955.1	<p>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</p>	Severe acute respiratory syndrome coronavirus 2	36.2	36.2	100%	8.9	95.45%	29854
MW290954.1	<p>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</p>	Severe acute respiratory syndrome coronavirus 2	36.2	36.2	100%	8.9	95.45%	29865
MW290953.1	<p>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</p>	Severe acute respiratory syndrome coronavirus 2	36.2	36.2	100%	8.9	95.45%	29865
MW290952.1	<p>Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/UT-UPHL-201116393/2020, complete genome</p>	Severe acute respiratory syndrome coronavirus 2	36.2	36.2	100%	8.9	95.45%	29896
MW290951.1	<p>Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/UT-UPHL-201116223/2020, complete genome</p>	Severe acute respiratory syndrome coronavirus 2	36.2	36.2	100%	8.9	95.45%	29824
MW290950.1	<p>Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/UT-UPHL-201115394/2020 ORF1ab polyprotein</p>	Severe acute respiratory syndrome coronavirus 2	36.2	36.2	100%	8.9	95.45%	29864

	(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
MW290949.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-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	Severe acute respiratory syndrome coronavirus 2	36.2	36.2	100%	8.9	95.45%	29865
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
	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
MW290946.1	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
MW290945.1	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
MW290944.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-	Severe acute	36.2	36.2	100%	8.9	95.45%	29813
MW290943.1								

	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, complete cds	respiratory syndrome coronavirus 2							
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	

VISTA DI ALLINEAMENTO VERTICALE AGGANCIATO A QUERY CON LETTERE D'IDENTITA'

Query range 1: 1 to 22

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

<u>MW292635.1</u>	15406	GTGAAATGGTCATGTGTGGCGG	15427
<u>MW292634.1</u>	15396	GTGAAATGGTCATGTGTGGCGG	15417
<u>MW292633.1</u>	15419	GTGAAATGGTCATGTGTGGCGG	15440
<u>MW292632.1</u>	15430	GTGAAATGGTCATGTGTGGCGG	15451
<u>MW292631.1</u>	15396	GTGAAATGGTCATGTGTGGCGG	15417
<u>MW292630.1</u>	15431	GTGAAATGGTCATGTGTGGCGG	15452
<u>MW292629.1</u>	15394	GTGAAATGGTCATGTGTGGCGG	15415
<u>MW290973.1</u>	15391	GTGAAATGGTCATGTGTGGCGG	15412
<u>MW290972.1</u>	15248	GTGAAATGGTCATGTGTGGCGG	15269
<u>MW290971.1</u>	15393	GTGAAATGGTCATGTGTGGCGG	15414
<u>MW290970.1</u>	15393	GTGAAATGGTCATGTGTGGCGG	15414
<u>MW290969.1</u>	15406	GTGAAATGGTCATGTGTGGCGG	15427
<u>MW290968.1</u>	15392	GTGAAATGGTCATGTGTGGCGG	15413
<u>MW290967.1</u>	15344	GTGAAATGGTCATGTGTGGCGG	15365
<u>MW290966.1</u>	15399	GTGAAATGGTCATGTGTGGCGG	15420
<u>MW290965.1</u>	15415	GTGAAATGGTCATGTGTGGCGG	15436
<u>MW290964.1</u>	15400	GTGAAATGGTCATGTGTGGCGG	15421
<u>MW290963.1</u>	15318	GTGAAATGGTCATGTGTGGCGG	15339
<u>MW290962.1</u>	15393	GTGAAATGGTCATGTGTGGCGG	15414
<u>MW290961.1</u>	15393	GTGAAATGGTCATGTGTGGCGG	15414
<u>MW290960.1</u>	15393	GTGAAATGGTCATGTGTGGCGG	15414
<u>MW290959.1</u>	15392	GTGAAATGGTCATGTGTGGCGG	15413
<u>MW290958.1</u>	15393	GTGAAATGGTCATGTGTGGCGG	15414
<u>MW290957.1</u>	15354	GTGAAATGGTCATGTGTGGCGG	15375
<u>MW290956.1</u>	15388	GTGAAATGGTCATGTGTGGCGG	15409
<u>MW290955.1</u>	15382	GTGAAATGGTCATGTGTGGCGG	15403
<u>MW290954.1</u>	15393	GTGAAATGGTCATGTGTGGCGG	15414
<u>MW290953.1</u>	15393	GTGAAATGGTCATGTGTGGCGG	15414
<u>MW290952.1</u>	15424	GTGAAATGGTCATGTGTGGCGG	15445
<u>MW290951.1</u>	15393	GTGAAATGGTCATGTGTGGCGG	15414
<u>MW290950.1</u>	15392	GTGAAATGGTCATGTGTGGCGG	15413
<u>MW290949.1</u>	15393	GTGAAATGGTCATGTGTGGCGG	15414
<u>MW290948.1</u>	15393	GTGAAATGGTCATGTGTGGCGG	15414
<u>MW290947.1</u>	15404	GTGAAATGGTCATGTGTGGCGG	15425
<u>MW290946.1</u>	15393	GTGAAATGGTCATGTGTGGCGG	15414
<u>MW290945.1</u>	15415	GTGAAATGGTCATGTGTGGCGG	15436
<u>MW290944.1</u>	15382	GTGAAATGGTCATGTGTGGCGG	15403
<u>MW290943.1</u>	15393	GTGAAATGGTCATGTGTGGCGG	15414
<u>MW290942.1</u>	15428	GTGAAATGGTCATGTGTGGCGG	15449
<u>MW290941.1</u>	15393	GTGAAATGGTCATGTGTGGCGG	15414
<u>MW290940.1</u>	15427	GTGAAATGGTCATGTGTGGCGG	15448
<u>MW290939.1</u>	15392	GTGAAATGGTCATGTGTGGCGG	15413
<u>MW290938.1</u>	15393	GTGAAATGGTCATGTGTGGCGG	15414
<u>MW290937.1</u>	15425	GTGAAATGGTCATGTGTGGCGG	15446
<u>MW290936.1</u>	15392	GTGAAATGGTCATGTGTGGCGG	15413
<u>MW290935.1</u>	15427	GTGAAATGGTCATGTGTGGCGG	15448
<u>MW288285.1</u>	15414	GTGAAATGGTCATGTGTGGCGG	15435
<u>MW288284.1</u>	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="MESLVPGFNEKTHVQLSLPVLQVRDVLVRGFGDVEEAVAEARQ
HLIEGTCGIVDLQKGVLPQLEQPYIFLKRC DARTAPHG HVMVELVAELDGVQYGRSGE
SLGVLVPHVGETPIGYRKVLVRKNGKNGAGGHLYGADLRFYDLGDELGTDPLDDFQQD
WNTKHGSGLRDLFRELNGGVYTRYVDNNFCGPDGYPLECIKDLLARAGKSSAPLAEQ
LDFLESKRGVYCCREHEHEIAWYTERSDKSYELQTFPFDITNAKKFDSFKGECPKFVFP
LNSTVKVLQPRVEKKKTEGFLGRIRTVYQVASPGECNSMHLSTYMNCNHCGEKSWQTC
DFLTATCEMCGNQNTVEEGPTTCGYVPSNAVVKMVC PACQNPEIGPDHSVADYHNNSK
IETRLRKGGRIKSFGGCVFSYVGCYNKRAFVWPRAAANIGSNHTGVVGEVETMNE DL
LQILSRERVVINIVGEFCLNEEIAILLASLSASTSAFVETVKNLDFKTFKKIIESCGN
YKVTGKFKPGVWNI GTSKSLTPLHCFSSQAAGVVR SIFSRTLATANHSIVDLHRAA
MIIFSDISDQANRVL DAMVNTSDLVTESVVMAYLTGGLVQQVSTWLSQLLNTSVDKF
SAVLRWLEQKLQGGIDFLRQAWGILKLLVTGAYVVIRGKIQVVNTSLIECVTSFVDVV
NKVFELCTDYITVAGARVRAINFGEVLIAQSRGLYRQCVRARDQLQLLMLPKSPKDVV
FLDGDAYDTLLTSEEVTVKNGTLEALDLELSDVVTGVAEGVPVCVNGLMLELKEKEQ
YCALSPSLLATNNVFTLKG GAPT KGVTFGEDTVVEIQGYKSVKITFELDERVDKVLNE
KCASYTVETGTAEELACVVAESVVKTLQPISELLTPMGIDLDEWSVAKFYLFDESGE
AVLSSHMYCSFYPPDEEEEDLEESDVEYGTEDDYTGAPLEFGASSTVEQDEVHDEE
EDWLAPQEESVLYDQFTDYHKLTDNVFIKCADIVEESLKVNP TVVNAANIHLKHGG
GVARALDKATGGSMQKESNDYISTNGPLRVGGSCLLSGHNLAKHCLHVVGPNKNAGED
IKLLDAAYENFNAYEVVLSPLLSAGIFGVSP IQSLETCKRVVRNTVYIVVND SVFDQ
LLAKTPGKTNERPVESSEICEEVNQKPVVEFSETKELHEETNQKLSSEEPVKTRIE
ELNTTVDEAKFLTTKLLLYADVNGNLSEDSKVLIGNDGASFKKGAPYIVGDI ISEGEL
TCVVLPTKAVGGTTHMLTRALKNVP SDTYLTTYPGQVSGYTLDEAKAALKKSRSV FY
IILPSANVNAKEEVLGTVAWNLREMLAHAEE TRKMPVCMDVRAI IISTIQRKYKIGIGQ
EGLVDYKVRFYFYSSKTP IARVISNLNSLGEPLITMPLGYVTHGLNLEESARYMRSVK
VPVVVSVPDAVTSYNGYVTSASKSAEEHF IETVSLAGSYKDSYSQORTELGEVFL
KRGDKIVYHTVGNVIEFHMEGEVLPLEKLTLLALREVKT IKVFTTVDNINLHTQVID
MSMTYGQQLGPTYMDGADLTKVKPHASHENKTFVLP SDDTLRIEAFEYHYHTVDESFF
GRYMSALNHTKRWKYPQVGGLTSIKWADNNCYLSSVLLSLQQIDIKFNAPALQDAYYR
ARAGDAANFCALVLAYSKKT V GELGDVRETMAHLLQHANLES AKRVLNVVCKHCGQKS
TTLSGVEAVMYMGTLSYDHLKRGVKIPCVCGREATQYLVKQESTFVMSAPPAEYTLQ
TGEFLCANEYTGNYQCGHYTHITNRETIYKIDGALLTKITEYKGPVADV FYKETSYST
DIKPVSYKLDGVTYTEINPDLNGYKKNAYYTEQPIDL VPTQPLPNASFDNFRFVCA
NTKFADDLNQMTGFKKPPSRDLTITFFPDLNGDVVAIDYRHYTPTFKKGAKLVHKPIL
WHVNQTTTKSTFKPNMWCLRCLYSTKPVPTSNSFEVLSSDDAQGMDNLACESQQTVAE
EVVDNPTIQKDIIECDVKTTEVGVNVILKPSADGIKVTSELEHEDLMAAYVNETSITI
KKPNELSIMLGLKTIATHGAAAINSVPWIKICAYVKPFLGYVAEQSKNCIKRCFRRVF
NDYMPFLLLTLLQLCTFTKSTNFRIKAAMP IVIARN SVIGGVRFC LDALTMVYKSPKF
SGILTVVMWLLLLSVCLGCLVYAVASFGA ILSGFGLMSYCDGVRAGYVNSSNVTIPDY
CAGSLPCGVCLGGLDSL DAYPALETIQVTISSYKLDLTFVGMMAEWFLAYMLFTKFFY
LLGLFALMQLFFGLFATHFVNNSWLMWLI INVVQMAPISAMVRMYVFFASFYVWKAY

IHVINGCTSSSTCIMCYKRNRRATRVECTTIVNGMKKSFYVYANGGQGFCKLHNWNC
LNCDFTCSTGSTFISDEVARDLSLQFKRPINPTDQSSYNVDSVTVKDGTLYLYFQKAGKLT
YERHPLSYFVNLNLRANNVKGTLPIINVIVFDGKSKCEEAASASVYYSQLMCQPILL
LDQALISDVGDSTEVAVKMFDAYVNAFSSTFNAPMEKLTFTIATAHAEIAKGVSLDSV
LSTFLSAARQGFVDSVDVTDKDVMECLKLSHHSDEITSDSCNNFMLTYNKVENMTPRD
LGACIDCSARHINAQVAKSHNVSLVWNVKDYMSLSEQLRKQIRSAAKNNIPFKLTCA
TTRQVVNVITTKISLKGKGFVSNWFRLLKMTVLMVLVAFIFYFITPTHTLMGHDFV
SSEIIGYKAIHNGVTRDVLTTDDCFANKHTGFDHWFSSQRGGSYRNDKTCPVIAAVITR
EVGFIVPGLPGTVRRASNGDFLHFLPRVFSAVGNICYTPAKLIEYTDFAVSACVLAAE
CTIFKDAQGKVPYCYDTNLLGESSISYSELRPDTRYVLMDGSI IQFPSTYLEGSVRVV
TTFDSEYCRHGTCERSDAGVCLSTNGRWLNNDDYRSIPGVFCGADASDLLFNIFTPL
VRPVGTLDISASVVAGGLIAILVTCVAYYFMKFRRAFGEYNHVVFANALLFLLSFTIL
CLTPAYTFLPGIYSLLYLYLTFYFTNDVSLAHLQWLAMFSPIVPFWITVTVYVVCISI
KHCHWFFSNYLKRRVVFNGVTFSTFEAALCTFLLNKEMYLKLRSETLLPLTQYNRYL
ALYNKYKYFSGALDTSYREAACCHLAKALNDFNSGADVLYQPPQTSITSAVLQSGF
RKMAFPSPGKVEGCMVQVTCGTTLNGLWLDVVYCPRHVICTELDMLNPYDDLIRK
SNHNFLVQASNVQLRVIGHTMQNCLLKLKVDIANPKTPKYKQFVRIQPGQTFVSLACYN
GAPSGVYQCAMRSNHTIKGSFLNGSCGSGVFNIDYDCVSFCYMHMELPTGVHAGTDL
EGNFYGFVDRQTAQAAGTDTTITLNLVLAWLYAAVINGERWFLNRFTTTLNDFNLVAM
KYNYEPLTQDQVDILGPLSAQTVAVMDMCAALKELLQNGLNRTILGSTILEDEFTP
FDVVRQCSGVTQFGKFKKVVKGTHHWLTLTLLTSLLLILVQSTQWSLFFFVYEHAFLEP
FMGMVVCFAACAMVLVKKHAFLECLFLLPSLITVAYFNMIYMPASWVMRMTWLDLVD
TSLSGYRLKDCVMYALAAFLLIIMTARTVYDDAARRVWTVMNVITLVYKVVYGNLSDQA
LAMWALVISVTSNYSGVVTIMFLARAIVFLCVEYYPILFITGNLQCIMLVYCFGLY
CCCCYFGLFCLLNRYFRLLTLGVYDYFVSTQEFYRMSQGLLPPKTSLDAFKLVKLLG
IGGKPCIKVATVQSKMSDIKCTSVVLLSVLQQLRIESSKLWAQCVQLHNDILLAKDT
TEAFKEMVSVLLSVLLSMQGAVDINKLCDEMLNRRATLQAIASEFSSLSYAAAYATAQE
AYEQAVANGDSEVVLLKLLKSLNVAKSEFDRDAAMQRKLEKQAMTQMYKQARSED
KRAKVTSAMQTMFLTMLRKLNDALNNIINNARDGCVPLNIIPLTAAKLMVVVDPYN
TYKNTCDGNTFTYASALWEIQQVVDADSKVVQLSEINMDNSQNLAWPLIVTALRSNSA
VKLQNNELSPVALRQMSCAAGTTQACTDDNALAYNTSKGGRFVLALLSDHQDLKWA
RFPKSDGTGTIYTELEPPCRFVTDTPKGPKVKYLYFIKGLNNLNRGMVLGSLAATVRL
QAGNATEVPANSTVLSFCFAFAVDPAKAYKDYLASGGQPITNCVKMLCTHTGTGQAITV
IPEANMDQESFGGASCCLYCRCHIDHPNPKGFCDLKGKYVQIPTTCTNDPVGFILRNT
VCTVCGMWKGYGCSCDQLREPVMQAADAPAFLNRCVSAARLTPCGTGTSTDVVYRA
FDIYNEKVAGFAKFLKTNCCRFQEVDEEGLLDSYFVVKRHTMSNYQHEETMYNLVKE
CPAVAVHDFKFRVDGDMVPHISRQRLTKYTMADLVYALRHFDEGNCDTLKEILVTYN
CCDDAYFNKKDWYDFVENPDILRVYACLGERVRQALLKTVQFCDAMRDAGIVGLTLD
NQDLNGNWDYDFGDFVQVAPGAGIPIVDSYYSLLMPLILTLTKALAAESHMDCDTTKPLI
KWDLLKYDFTEERLCLFNRYFKYWDQTYHPNCINCLDDRCILHCANFNVLFSTVFPPT
SFGPLVRKIFVDGVPFVSTGYHFRELGVVHNQDVNLHSSRLSFKELLVYAADPAMHA
ASGNLLLDKRTTCFSVAALTNVAFQTVKPGNFNKFYDFAVSKGFFKEGSSVELKHF
FFAQDGNAAISDYDYRYNLPMTCDIRQLLFVVEVVDKYFDCYDGGCINANQVIVNNL
DKSAGFPFNKWKARLYYDSMSYEDQDALFAYTKRNVIPITITQMNLYAISAKNRART

VAGVSI CSTMTRQFHQKLLKSI AATRGATVVIGTSKFGYGGWHNMLKTVYSDVETPNL
MGWDYPKCDRAMPNMLRIMASLVLARKHSTCCNLSHRFYGLANECAQVLS**SEMVMCGG**s
LYVKPGGTSSGDATTAYANSVFNICQAVTANVNALLSTDGNKIADKYVRNLQHRLYEC
LYRNRDVDHEFVEEFYAYLRKHFSMMILSDDAVVCYNSNYAAQGLVASIKNFKAVLYY
QNNVFMSEAKCWETETDLTKGPHEFCSQHTMLVKQGDDYVYLPYPDPSRILGAGCFVDD
IVKTDGTLMIERFVSLAIDAYPLTKHPNQEYADVFLYLQYIRKLDDELTDGHMLDMYS
VMLTNDNTRSRYWEPEFYEAMYTPHTVLQAVGACVLCNSQTSLRGCSIRPFLCCKCC
YDHVISTSHKLVLSVNPYVCNAPGCDVTDVTDVTLQYLYGGMSYYCKSHKPPISFPLCANGQ
VFGLYKNTCVGSDNVTDFNAIATCDWTNAGDYILANTCTERLKLFAAETLKANEETFK
LSYGIATVREVLSRELHLSWEIGKPRPPLNRNYVFTGYRVTKNSKVQIGEYTFEKGD
YGDVVYRGTTTYKLVNGDYFVLTSHVMPLTAPTLPVQEHYVRITGLYPTLNISDEF
SSNVANYQKVMQKYSTLQGGPGTGKSHFAIGLALYPSARIVYTACSHAAVDALCEK
ALKYLPIDKCSRIIPARARVECFDKFKVNSTLEQYVFCTVNALPETTADIVVFEVSM
ATNYDLSVVNARLRAKHVYIGDPAQLPAPRTLTKGTLEPEYFNSVCRLMKTIGPDM
FLGTCRRCPAEIVDTSALVYDNKLRHKGKSSQCFKMFYKGVITHDVSSAINRPQIG
VVREFLTRNPAWRKAVFISPYNSQNAVASKILGLPTQTVDSSQGSEYDYVIFAQTET
AHSCNVNRFNVAITRAKVGILCIMSDDKLDYDKLQFTSLEVPRRSVAVLQSENVTLGFK
DCSKLITGLHPTQAPTYLSVDTKFKTEGLCVDIPGIPKDMTYRRLISMMGFKMNYQVN
GYPNMFITRDEAIKHVRAWIGFDVEGCHATRDAVGTNLPLQLGFSTGVNLVAVPTGYV
DTSAAATEFSRVNAKPPPGDQFKHLIPLMYKGLPWNIVRVKIVQMLSDTLKDLSDRVVF
VLWAHGFELTSMKYFVKIGPERTCCLCDKRATCFSTSSDITYACWHHSVGFYVYNPFM
IDVQQWGFNGNLQSNHDQHCQVHGNAHVASCDAIMTRCLAIHECFVKRVDWSVEYPII
GDELRINVACRKVQHMVKSALLADKFPVLHDIGNPKAIKVPQADVEWKFYDVQPCS
DKAYKIEELFYSYATHHDKFTDGVCLFWNCNVDRYPSNAIVCRFDTRVLSNLNLPGCD
GGSLYVNKHAFHTPAFDKGAFANLQQLPFFYYSDSPCESHGKQVVSDIDYVPLKSATC
ITRCNLGGAVCRHHASEYRQYLDAYNMMISAGFSLWIYKQFDTYNLWNTFTRLQSLN
VAYNVVNGKHFQAGEKPVSIINNTVYTKVDGVDVEIFENKTTLPVNVAFELWAKRN
IKPVPEIKILNNLGVDIAANTVIWDYKRESPAYISTIGVCTMTDIAKKPTENACSSLT
VFFDGRVDGQVDSFRNARNGVLTIEGSKGLNPSKGGPPQASLNGVTLIGESVKTQFNY
FKKVDGVVQQLPETYFTQSRSLDDFKPRSQMEVDFLQLAMDEFIERYKLEGYAFEHIV
YGDVSHGQLGGLHLMIGLAKRSLESLLKLEDFIPIDSTVKNYFVTDATGSSKCVCSV
IDLDDDFVEI IKSQDLSVSVKVVTVTIDYAEISFMLWCKDGHVETFYPKLQANQWQ
PGVAMPNLYKMQRMLLDKCDLHNYGENAVIPKGIMMNVAKYTQLCQYLNTLTIAVPYN
MRVIHFAGSDKGVAPGSAVLKQWLPVGTLLVSDINDFVSDADSTLIGDCSTVYTAN
KWDLIISDMYDPKTKHILKENDSKEGFFTYLCGFIKQKLALGGSVAIKITEHSWNADL
YKLMGYFSWWTAFVTNVNASSEAFVIGVNYLGKQKESIDGYTMHANYIFWRNTNPIQ
LSSYSLFDMSKFPLKLRGTAVMSLKDNDQINDMICSLLEKGRLLIRENNKVVFSDDVLV
NN"

ORIGIN

1 gtgaaatggt **t**atgtgtggc 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	GTGAA	ATGGT	T	ATGTGTGGCGG		22	
Sbjct	15260	GTGAA	ATGGT	T	ATGTGTGGCGG		15281	
CDS: ORFlab [Bat coro	5024	S	E	M V	M	C	G	G

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

GenBank: [GU190215.1](#)

[GenBank](#) [FASTA](#)

[Link To This View](#) | [Feedback](#)



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	GTGAAATGGT T ATGTGTGG	19		
Sbjct 2494	GTGAAATGGT T ATGTGTGG	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	GTGAAATGGT T ATGTGTGG	19		
Sbjct 2335	GTGAAATGGT T ATGTGTGG	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	GTGAAATGGT T ATGTGTGG	19		
Sbjct 2918	GTGAAATGGT T ATGTGTGG	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	GTGAAATGGT T ATGTGTGG	19		
Sbjct 2496	GTGAAATGGT T ATGTGTGG	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	GTGAAATGGT T ATGTGTGG	19		
Sbjct 1327	GTGAAATGGT T ATGTGTGG	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	GTGAAATGGT T ATGTGTGG	19		
Sbjct 1327	GTGAAATGGT T ATGTGTGG	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	GTGAAATGGT T ATGTGTGG	19		
Sbjct 1327	GTGAAATGGT T ATGTGTGG	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	GTGAAATGGTATGTGTGGCGG	22
Sbjct		15393	GTGAAATGGTCATGTGTGGCGG	15414
CDS:ORF1ab polyprote		5056	S E M V M C G G	

5.2 Ricerca BLAST sulla SEQUENZA di RNA

“GTGAGATGGTCATGTGTGGCGG” del Gene RdRp del “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	Bat SARS-like coronavirus	44.1	44.1	100%	0.037	100.00%	377

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 141470 RNA-dependent RNA polymerase (RdRp) gene, partial cds	Bat SARS-like coronavirus	44.1	44.1	100%	0.037	100.00%	364
MN312653.1	Bat SARS-like coronavirus isolate 141464 RNA-dependent RNA polymerase (RdRp) gene, partial cds	Bat SARS-like coronavirus	44.1	44.1	100%	0.037	100.00%	374
MN312652.1	Bat SARS-like coronavirus isolate 141460 RNA-dependent RNA polymerase (RdRp) gene, partial cds	Bat SARS-like coronavirus	44.1	44.1	100%	0.037	100.00%	327
MN312651.1	Bat SARS-like coronavirus isolate 141456 RNA-dependent RNA polymerase (RdRp) gene, partial cds	Bat SARS-like coronavirus	44.1	44.1	100%	0.037	100.00%	366
MN312650.1	Bat SARS-like coronavirus isolate 141455 RNA-dependent RNA polymerase (RdRp) gene, partial cds	Bat SARS-like coronavirus	44.1	44.1	100%	0.037	100.00%	371
MN312645.1	Bat SARS-like coronavirus isolate 141376 RNA-dependent RNA polymerase (RdRp) gene, partial cds	Bat SARS-like coronavirus	44.1	44.1	100%	0.037	100.00%	379
MN312644.1	Bat SARS-like coronavirus isolate 141375 RNA-dependent RNA polymerase (RdRp) gene, partial cds	Bat SARS-like coronavirus	44.1	44.1	100%	0.037	100.00%	362
MN312643.1	Bat SARS-like coronavirus isolate 141369 RNA-dependent RNA polymerase (RdRp) gene, partial cds	Bat SARS-like coronavirus	44.1	44.1	100%	0.037	100.00%	362
MN312642.1	Bat SARS-like coronavirus isolate 141352 RNA-dependent RNA polymerase (RdRp) gene, partial cds	Bat SARS-like coronavirus	44.1	44.1	100%	0.037	100.00%	349
MN312641.1	Bat SARS-like coronavirus isolate 141350 RNA-dependent RNA polymerase (RdRp) gene, partial cds	Bat SARS-like coronavirus	44.1	44.1	100%	0.037	100.00%	367
MN312640.1	Bat SARS-like coronavirus isolate 141344 RNA-dependent RNA polymerase (RdRp) gene, partial cds	Bat SARS-like coronavirus	44.1	44.1	100%	0.037	100.00%	367
MN312639.1	Bat SARS-like coronavirus isolate 141342 RNA-dependent RNA polymerase (RdRp) gene, partial cds	Bat SARS-like coronavirus	44.1	44.1	100%	0.037	100.00%	346
MN312638.1	Bat SARS-like coronavirus isolate 141341 RNA-dependent RNA polymerase (RdRp) gene, partial cds	Bat SARS-like coronavirus	44.1	44.1	100%	0.037	100.00%	359
MN312637.1	Bat SARS-like coronavirus isolate 141335 RNA-dependent RNA polymerase (RdRp) gene, partial cds	Bat SARS-like coronavirus	44.1	44.1	100%	0.037	100.00%	376
MN312606.1	Bat SARS-like coronavirus isolate 7330 RNA-dependent RNA polymerase (RdRp) gene, partial cds	Bat SARS-like coronavirus	44.1	44.1	100%	0.037	100.00%	402
MN312605.1	Bat SARS-like coronavirus isolate 7326 RNA-dependent RNA polymerase (RdRp) gene, partial cds	Bat SARS-like coronavirus	44.1	44.1	100%	0.037	100.00%	403
MN312604.1	Bat SARS-like coronavirus isolate 7325 RNA-dependent RNA polymerase (RdRp) gene, partial cds	Bat SARS-like coronavirus	44.1	44.1	100%	0.037	100.00%	405
MN312603.1	Bat SARS-like coronavirus isolate 6533 RNA-dependent RNA polymerase (RdRp) gene, partial cds	Bat SARS-like coronavirus	44.1	44.1	100%	0.037	100.00%	406
MN312602.1	Bat SARS-like coronavirus isolate 6530 RNA-dependent RNA polymerase (RdRp) gene, partial cds	Bat SARS-like coronavirus	44.1	44.1	100%	0.037	100.00%	399
MN312601.1	Bat SARS-like coronavirus isolate 6301 RNA-dependent RNA polymerase (RdRp) gene, partial cds	Bat SARS-like coronavirus	44.1	44.1	100%	0.037	100.00%	400
MN312599.1	Bat SARS-like coronavirus isolate 6278 RNA-dependent RNA polymerase (RdRp) gene, partial cds	Bat SARS-like coronavirus	44.1	44.1	100%	0.037	100.00%	371
MN312598.1	Bat SARS-like coronavirus isolate 6266 RNA-dependent RNA polymerase (RdRp) gene, partial cds	Bat SARS-like coronavirus	44.1	44.1	100%	0.037	100.00%	378
MN312597.1	Bat SARS-like coronavirus isolate 6255 RNA-dependent RNA polymerase (RdRp) gene, partial cds	Bat SARS-like coronavirus	44.1	44.1	100%	0.037	100.00%	402
MN312596.1	Bat SARS-like coronavirus isolate 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	Severe acute respiratory syndrome coronavirus 2	44.1	44.1	100%	0.037	100.00%	29833

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

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

MT308984.1	SARS coronavirus Urbani	44.1	44.1	100%	0.037	100.00%	29730
MK211378.1	Coronavirus BtRs-BetaCoV/YN2018D, complete genome	44.1	44.1	100%	0.037	100.00%	30213
MK211377.1	Coronavirus BtRs-BetaCoV/YN2018C, complete genome	44.1	44.1	100%	0.037	100.00%	29689
MK211376.1	Coronavirus BtRs-BetaCoV/YN2018B, complete genome	44.1	44.1	100%	0.037	100.00%	30256
MK211375.1	Coronavirus BtRs-BetaCoV/YN2018A, complete genome	44.1	44.1	100%	0.037	100.00%	29698
MK211374.1	Coronavirus BtRI-BetaCoV/SC2018, complete genome	44.1	44.1	100%	0.037	100.00%	29648
MK062184.1	SARS coronavirus Urbani isolate icSARS-C7-MA, complete genome	44.1	44.1	100%	0.037	100.00%	29874
MK062183.1	SARS coronavirus Urbani isolate icSARS-C7, complete genome	44.1	44.1	100%	0.037	100.00%	29874
MK062182.1	SARS coronavirus Urbani isolate icSARS-C3-MA, complete genome	44.1	44.1	100%	0.037	100.00%	29727
MK062181.1	SARS coronavirus Urbani isolate icSARS-C3, complete genome	44.1	44.1	100%	0.037	100.00%	29727
MK062180.1	SARS coronavirus Urbani isolate icSARS-MA, complete genome	44.1	44.1	100%	0.037	100.00%	29727
MK062179.1	SARS coronavirus Urbani isolate icSARS, complete genome	44.1	44.1	100%	0.037	100.00%	29727
MG772934.1	Bat SARS-like coronavirus isolate bat-SL-CoVZXC21, complete genome	44.1	44.1	100%	0.037	100.00%	29732

Bat SARS-like coronavirus isolate bat-SL-CoVZXC45, complete genome

MG772933.1	Bat SARS-like coronavirus isolate bat-SL-CoVZXC1 RNA-dependent RNA polymerase gene, partial cds	44.1	44.1	100%	0.037	100.00%	29802
MG772932.1	Bat SARS-like coronavirus isolate bat-SL-CoVZXC2 RNA-dependent RNA polymerase gene, partial cds	44.1	44.1	100%	0.037	100.00%	437
MG772931.1	Bat SARS-like coronavirus isolate bat-SL-CoVZXC4 RNA-dependent RNA polymerase gene, partial cds	44.1	44.1	100%	0.037	100.00%	381
MG772930.1	Bat SARS-like coronavirus isolate bat-SL-CoVZXC5 RNA-dependent RNA polymerase gene, partial cds	44.1	44.1	100%	0.037	100.00%	405
MG772929.1	Bat SARS-like coronavirus isolate bat-SL-CoVZXC6 RNA-dependent RNA polymerase gene, partial cds	44.1	44.1	100%	0.037	100.00%	438
MG772928.1	Bat SARS-like coronavirus isolate bat-SL-CoVZXC7 RNA-dependent RNA polymerase gene, partial cds	44.1	44.1	100%	0.037	100.00%	430
MG772927.1	Bat SARS-like coronavirus isolate bat-SL-CoVZXC12 RNA-dependent RNA polymerase gene, partial cds	44.1	44.1	100%	0.037	100.00%	407
MG772924.1	Bat SARS-like coronavirus isolate bat-SL-CoVZXC21 RNA-dependent RNA polymerase gene, partial cds	44.1	44.1	100%	0.037	100.00%	437
MG772921.1	Bat SARS-like coronavirus isolate bat-SL-CoVZXC22 RNA-dependent RNA	44.1	44.1	100%	0.037	100.00%	428
MG772920.1		44.1	44.1	100%	0.037	100.00%	418

	polymerase gene, partial cds								
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

VISTA DI ALLINEAMENTO VERTICALE AGGANCIATO A QUERY CON LETTERE D'IDENTITA'

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

<u>MN312643.1</u>	130	GTGAGATGGTCATGTGTGGCGG	151
<u>MN312642.1</u>	131	GTGAGATGGTCATGTGTGGCGG	152
<u>MN312641.1</u>	130	GTGAGATGGTCATGTGTGGCGG	151
<u>MN312640.1</u>	130	GTGAGATGGTCATGTGTGGCGG	151
<u>MN312639.1</u>	129	GTGAGATGGTCATGTGTGGCGG	150
<u>MN312638.1</u>	130	GTGAGATGGTCATGTGTGGCGG	151
<u>MN312637.1</u>	128	GTGAGATGGTCATGTGTGGCGG	149
<u>MN312606.1</u>	122	GTGAGATGGTCATGTGTGGCGG	143
<u>MN312605.1</u>	122	GTGAGATGGTCATGTGTGGCGG	143
<u>MN312604.1</u>	127	GTGAGATGGTCATGTGTGGCGG	148
<u>MN312603.1</u>	127	GTGAGATGGTCATGTGTGGCGG	148
<u>MN312602.1</u>	120	GTGAGATGGTCATGTGTGGCGG	141
<u>MN312601.1</u>	126	GTGAGATGGTCATGTGTGGCGG	147
<u>MN312599.1</u>	129	GTGAGATGGTCATGTGTGGCGG	150
<u>MN312598.1</u>	128	GTGAGATGGTCATGTGTGGCGG	149
<u>MN312597.1</u>	122	GTGAGATGGTCATGTGTGGCGG	143
<u>MN312596.1</u>	122	GTGAGATGGTCATGTGTGGCGG	143
<u>MN312595.1</u>	126	GTGAGATGGTCATGTGTGGCGG	147
<u>MT434760.1</u>	15403	GTGAGATGGTCATGTGTGGCGG	15424
<u>MT308984.1</u>	15361	GTGAGATGGTCATGTGTGGCGG	15382
<u>MK211378.1</u>	15361	GTGAGATGGTCATGTGTGGCGG	15382
<u>MK211377.1</u>	15359	GTGAGATGGTCATGTGTGGCGG	15380
<u>MK211376.1</u>	15360	GTGAGATGGTCATGTGTGGCGG	15381
<u>MK211375.1</u>	15362	GTGAGATGGTCATGTGTGGCGG	15383
<u>MK211374.1</u>	15324	GTGAGATGGTCATGTGTGGCGG	15345
<u>MK062184.1</u>	15362	GTGAGATGGTCATGTGTGGCGG	15383
<u>MK062183.1</u>	15362	GTGAGATGGTCATGTGTGGCGG	15383
<u>MK062182.1</u>	15361	GTGAGATGGTCATGTGTGGCGG	15382
<u>MK062181.1</u>	15361	GTGAGATGGTCATGTGTGGCGG	15382
<u>MK062180.1</u>	15361	GTGAGATGGTCATGTGTGGCGG	15382
<u>MK062179.1</u>	15361	GTGAGATGGTCATGTGTGGCGG	15382
<u>MG772934.1</u>	15352	GTGAGATGGTCATGTGTGGCGG	15373
<u>MG772933.1</u>	15418	GTGAGATGGTCATGTGTGGCGG	15439
<u>MG772932.1</u>	144	GTGAGATGGTCATGTGTGGCGG	165
<u>MG772931.1</u>	87	GTGAGATGGTCATGTGTGGCGG	108
<u>MG772930.1</u>	124	GTGAGATGGTCATGTGTGGCGG	145
<u>MG772929.1</u>	144	GTGAGATGGTCATGTGTGGCGG	165
<u>MG772928.1</u>	136	GTGAGATGGTCATGTGTGGCGG	157
<u>MG772927.1</u>	124	GTGAGATGGTCATGTGTGGCGG	145
<u>MG772924.1</u>	145	GTGAGATGGTCATGTGTGGCGG	166
<u>MG772921.1</u>	134	GTGAGATGGTCATGTGTGGCGG	155
<u>MG772920.1</u>	124	GTGAGATGGTCATGTGTGGCGG	145
<u>MG772913.1</u>	124	GTGAGATGGTCATGTGTGGCGG	145
<u>MG772911.1</u>	144	GTGAGATGGTCATGTGTGGCGG	165
<u>MG772910.1</u>	124	GTGAGATGGTCATGTGTGGCGG	145
<u>MT782115.1</u>	15343	GTGAGATGGTCATGTGTGGCGG	15364
<u>MT782114.1</u>	15343	GTGAGATGGTCATGTGTGGCGG	15364
<u>MT439595.1</u>	15403	GTGAGATGGTCATGTGTGGCGG	15423
<u>MW276628.1</u>	15377	GTGAGATGGTCATGTGTGGCGG	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

CDS

```
/gene="ORFlab"  
<1..>22  
/gene="ORFlab"  
/ribosomal_slippage  
/codon_start=3  
/product="ORFlab polyprotein"  
/protein_id="QJQ27876.1"  
/translation="MESLVPGFNEKTHVQLSLPVLQVRDVLVRGFGDSVEEVLSEARQ  
HLKDGTCGLVEVEKGVLPQLEQPYVFIKRS DARTAPHGHVMVELVAELEGIQYGRSGE  
TLGVLVPHVGEIPVAYRKVLLRKNGKNGAGGHSYGADLKSFDLGDELGTDPYEDFQEN  
WNTKHSSGVTRELMRELNGGAYTRYVDNNFCGPDGYPLECIKDLLARAGKASCTLSEQ  
LDFIDTKRGVYCCREHEHEIAWYTERSEKSYELQTPFEIKLAKKFDTFNGECPNFVFP  
LNSIIKTIQPRVEKKKLDGFMGRIRSVYPVASPNECNQMCLSTLMKCDHCGETSWQTG  
DFVKATCEFCGTENLTKEGATTCGYLPQNAVVKIYCPACHNSEVGPESHSLAEYHNESG  
LKTILRKGGRITAFGGCVFSYVGCHNKCAWVPRASANIGCNHTGVVGESEGLNDNL  
LEILQKEKVNINIVGDFKLNEEIAIILASFSASTSAFVETVKGLDYKAFKQIVESCGN  
FKVTKGKAKKGAWNIGEQKSLSPLYAFASEAARVRSIFSRLETAQNSVRVLQKAA  
ITILDGISQYSLRLIDAMMFTSDLATNNLVVMAYITGGVVQLTSQWLTNIFGTVYEKL  
KPVLDWLEEKFEKGEVFLRDGWEIVKFI STCACEIVGGQIVTCAKEIKESVQTFFKLV  
NKFLALCADSIIIGGAKLALNLGETFVTHSKGLYRKC VKSREETGLLMPLKAPKEII  
FLEGETLPTPEVLTEEVV LKTGDLQPLEQPTSEAVEAPLVGTPVCINGLMLEIKDTEK  
YCALAPNMMVTNNTFTLKGGA PTKVTFGDDTVIEVQGYKSVNITFELDERIDKVLNEK  
CSAYTVELGTEVNEFACVVADAVIKTLQPVSELLTPLGIDLDEWSMATYYLFDSEGEF  
KLASHMYCSFYPPDEDEEEGDC EEEEFEPSTQYEGTEDDYQGKPLEFGATSAALQSE  
EEQEEDWLDDDSQQT VGGQDGS EDNQT TTIQTIVEVQPQLEMELTPVVQTI EVNSFSG  
YLKLT DNVIK NADIVEEAKVKPTVVVNAANVYLKHGGGVAGALNKATNNAMQVESD  
DYIATNGPLKVGGS CVLSGHNLA KHCLHVGPVNKGEDIQLLKSAYENFNQHEVLLA  
PLLSAGIFGADPIHSLRVCVDTVRTNVYLA VFDKNLYDKLVSSFLEMKSEKQVEQKIA  
EIPKEEVKPFITESKPSVEQRKQDDKIKACVEEVTTLEETKFLTENLLYIDINGN  
LHPDSATLVSDIDITFLK KDAPYIVGDVVQEGVLTAVVIPTKKAGGTTEMLAKALRKV  
PTDNYITTYPGQGLNGYTV EEAKTVLKCKSAFYIILPSIISNEKQEILGTVSWNLREM  
LAHAEE TRKLMFVCVETKAI VSTIQRKYKGIKIQEGVVDYGARFYFYTSKTTVASLIN  
TLNDLNETLV TMLPGYVTHGLNLEEAARYMRS LKVPATVSVSSPDAVTAYNGYLTSSS  
KTPEEHFIETISLAGSYK DWSYSGQSTQLGIEFLKRGDKSVYYSNPTTFHLDGEVIT  
FDNLKTL LSLREVRTIKVFTTVDNINLHTQVVDM SMTYGGQFGPTYLDGADVTKIKPH  
NXHEGKTFYVLPND DTLRVEAFEYHHTDPSFLGRYMSALNH TKXXXXXXXXXGLTSIK  
WADNNCYLATALLT LQQIELKFNPPALQDAYYRARAGEAANFCALILAYCNKTVGELG  
DVRE TMSYLFQHANLDSCKRVLNVVCKTCGQQQTTLKGVEAVMYMG TSYEQFKKGVQ  
IPCTCGKQATKYL VQQESPFVMMSAPPAQYELKHGTF TCASEYTGNYQCGHYKHITSK  
ETLYCIDGALLTKSSEYKGPITDVFYKENS YTTTIKPVTYKLDGVVCTEIDPKLDNYY  
KKDNSYFTEQPIDLV PNQYPNASFDNFK FVCDNIKFADDLNQLTG YKKPASRELKVT  
FFPDLNGDVVAIDYKH YTPSFKKGAKLLHKPIVWHVNNATNKATYKPN TWCIRCLWST  
KPVETSNSFDVLKSEDAQGM DN LACEDLKPVSEEVVENPTIQKDVLECNVKTTEVVG D  
IILKPANN SLKITEEVGHTDLMAAYVDNSSLT IKKPNELSRV LGLKTLATHGLAAVNS
```

VPWDTIANYAKPFLNKVSTTTNIVTRCLNRVCTNYMPYFFTLQLCTFTRSTNSRI
KASMPPTIAKNTVKSVMGKFCLEASFNYLKS PNF SKLINIIWFLLLSVCLGSLIYSTA
ALGLVMSNLGMPSYCTGYREGYLNSTNVTIATYCTGSI PCSVCLSGLDLSDTYP SLET
IQITISSFKWDLTAFGLVAEWFLAYILFTRFFYVGLAAIMQLFFSYFAVHFISNSWL
MWLIINLVQMAPISAMVRMYIFFASFYVWKS YVHVVDGCNSSTCMCCYKRNRATRVE
CTTIVNGVRRSFYVYANGGKGFCKLHNWNCVNCDFCAGSTFISDEVARDSLQFKRP
INPTDQSSYIVDSVTVKNGSIHLYFDKAGQKTYERHSLSHFVNLDNLRANNTKGS LPI
NVIVFDGKSKCEESSAKSASVYYSQ LMCQPI LLLDQALVSDVGD SAEVAVKMF DAYVN
TFSSTFNVPMEK LKTLVATAEAE LAKNVSLDNVLS TFI SAARQGFVDS DVE TKDVVEC
LKLSHQSDIEVTGDCSNMYMLTYNKVENMTPRDLGACIDCSARHINAQVAKSHNIALI
WNVKDFMSLSEQLRKQIRSAAKNNLPFKLT CATTRQVNVVTTKIALKGGKIVNNWL
KQLIKVTLVFLFVA AIFYLITPVHVM SKHTDFSSEIIGYKAIDGGVTRDIASDTDCFA
NKHADFDTWFSQRGGSYTNDKACPLIAAVITREVG FVVPGLPGTILRTTNGDFLHFLP
RVFSAVGNICYTPSKLIEYTD FATSACVLA AECTIFK DASGKVPVYCYDTNVLEGSVA
YESLRPDTRYV LMDGSI IQFPNTYLEG SVRVVTTFDSEYCRHGT CERSEAGVCVSTSG
RWWLNNDYRSLPGVFCGVDVAVNLLTNMFTPLIQPIGALDISASIVAGGIVAIVVTCL
AAYFMRFRRA FGEYSHVVA FN TLLFLMSFTVLC LTPVYSFLPGVYSVIYLYLTFYLTN
DVSFLAHIQWVMVMTPLVPFWIT IAYIICISTKH FYWFFSNYLKRRVVFNGVSFSTFE
EAALCTFLLNKEMYLKLRSDVLLPLTQYNRYLALYNKYK YFSGAMDTTSYREACCHL
AKALNDFSNSGSDVLYQPPQTSITS AVLQSGFRKMAFP SGKVEGCMVQVTCGTTTLNG
LWLDDVVYCPRHVIC TSEDMLNPNYEDLLIRKSNHNFLVQAGNVQLRVIGHSMQNCVL
KLKVD TANPKTPKYKFVRIQPGQTF SVLACYN GSPSGVYQCAMRPNFTIKGSFLNGSC
GSVGFNIDYDCVSFCYMHMELPTGVHAGTDLEGNFYGPFVDRQTAQAAGDTTITVN
VLAWLYAAVINGDRWFLNRFTTTLNDFNLVAMKYN YEPLTQDHVDILGPLSAQTGIAV
LDMCASLKELLQNGMNGRTI LGSALLEDEFTPF DVVRQC SGVTFQSAVKRTIKGTHHW
LLLTILTSLLVLVQSTQWSLFFFX YENAF LPFAMGI IAMS AFAMMFVKHKHAF LCLFL
LPSLATVAYFNMYMPASWVMRIMTWLDMVDTSLSGFKLKD CVMYASAVVLLILMTAR
TVYDDGARRVWTL MNVLT LVYKVYYGNALDQAISMWALIISVTSNYS GVVTTVMFLAR
GIVFMCEYCP IFFITGNTLQCIMLVYCF LGYFCTCYFGLFCLLNRYFRLTLGVYDYL
VSTQEFRYMSQGLLPPKNS IDAFKLN IKLLGVGGKPCIKVATVQSKMSDVKCTSVVL
LSVLQQLRV ESSSKLWAQCVQLHNDILLAKDTTEAF EKMSL LSVLLSMQGA VDINKL
CEEMLDNRATLQAIASEFSS LPSYAAFATAQEAYEQAVANGDSEVVLK LKLSLNVAK
SEFDRDAAMQRKLEK MADQAMTQMYQARSEDKRAKVT SAMQTM LFTMLRKLNDALN
NIINNARDGCVPLNIIPLTTAAKLMVVIPDYNTYKNTCDGTTFTYASALWEIQQV VDA
DSKIVQLSEI SMDNSPNLAWPLI V T ALRANS AVKLQNNELSPVALRQM SCAAGTTQTA
CTDDNALAYYNTTKGGRFVLALLSDLQDLKWARFPKSDGTGTIYTELEPPCRFVDTDP
KGPVKYLYFIKGLNNLRGMVLGSLAATVRLQAGNATEVPANSTVLSFCFAVDAAK
AYKDYLASGGQPI TNCVKMLCTHTGTGQAITVTP EANMDQESFGGASCCLYCRCHIDH
ENPKGFCDLKGKYVQIPTTCANDPVGFTLKN TVCTVCGMWKYGCSCDQLRE PMLQSA
DAQSFLNRVCGVSAARLTPCGTGTSTDVVYRAFDIYNDKVAGFAKFLKTNCCRFQEKD
EDDNLIDSYFVVKRH TFSNYQHEETIYNLLKDCPAVVKHDFKFRIDGDMVPHISRQR
LTKYTMADLVYALRHFDEGNCDTLKEILV TYNCCDDYFNKKDWYDFVENPDILRVYA
NLGERVRQALLKTVQFC DAMRNAGIVGLTLDNQLDNGNWDYDFGDFIQ TTPGSGVPVV
DSYYSLLMPI LTLTRALTAESHVDTDLTKPYIKWDL LKYDFTEERLKLFD RYFKYWDQ

TYHPNCVNCDDRCILHCANFNVLFSTVFPPTSFGPLVRKIFVDGVPFVSTGYHFRE
 LGVVHNQDVNLHSSRSLFKELLVYAADPAMHAASGNLLLDKRTTCFVAALTNNAFQ
 TVKPGNFNKFDFYDAVSKGFFKEGSSVELKHFFFAQDGNAAISDYDYRYNLPTMCDI
 RQLLFVVEVVDKYFDCYDGGCINANQVIVNNLDKSAGFPFNKWKARLYYDSMSYEDQ
 DALFAYTKRNVIPITITQMNLYAISAKNRARTVAGVSICTMTNRQFHQKLLKSIAT
 RGATVVIIGTSKFGYGGWHNMLKXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
 XXXXXXXXXXXXXXXXXXXXXXXX**EMVMCGG**SLYVKPGGTSSGDATTAYANSVFNICQ
 AVTANVNALLSTDGNKIADKYVRNLQHRLYECLYRNRDVTDFVNEFYAYLRKHFSSM
 ILSDDAVVCFNSTYASQGLVASIKNFKSVLYYQNNVMSEAKCWTEETDLTKGPHEFCS
 QHTMLVKQGGDYVYLPYDPDSRILGAGCFVDDIVKTDGTLMIERFVSLAIDAYPLTKH
 PNQEYADVHLYLQYIRKLHDELGHMLDMYSVMLTNDNTSRYWEPEFYEAMYPHTV
 LQAVGACVLCNSQTSLRGACIRRPFLCCKCCYDHVISTSHKLVLSVNPYVCNAPGCD
 VTDVTLQYLGLGMSYCKSHKPPISFPLCANGQVFGLYKNTCVGSDNVTFDFAIATCDW
 TNAGDYILANTCTERLKLFAAETLKATEETFKLSYGIATVREVLSRELHLSWEVGKP
 RPPLNRNYVFTGYRVTKNSKVQIGEYTFEKGDYGDVVYRGTTTYKLVNGDYFVLTSH
 TVMPLSAPTLPQEHYVRITGLYPTLNI SDEFSSNVANYQKVMQKYSTLQGFPGTGK
 SHFAIGLALYPSARIVYTACSHAARDALCEKALKYLPIDKCSRIIPARARVECFDKF
 KVNSTLEQYVFCVNALPETTADIVVFDEISMATNYDLSVNVARLRAKHVYIGDPAQ
 LPAPRTLLTKGTLEPEYFNVCRLMKTIGPDMFLGTCRRCPAEIVDTSALVYDNKLG
 AHKDKSAQCCKMFYKGVITHDVSSA INRPQIGVVREFLTRNPAWRKAVFISPYNSQNA
 VASKILGLPTQTVDSQSEYDYVIFTQTETAHSCNVNRFNVAITRAKVGILCIMS
 RDLYDKLQFTSLEIPRRNVATLQAENVTLGFKDCSKVITGLHPTQAPTHLSVDTKFKT
 EGLCVDIPGIPKDMTYRRLISMMGFKMNYQVNGYPNMFITREEAIRHVRAWIGFDVEG
 CHATREAVGTNLPLQLGFSTGVNLVAVPTGYVDTPNNTDFSRVSAKPPPGDQFKHLIP
 LMYKGLPWNVRIKIVQMLSDTLKNLSDRVVFLWAHGFELTSMKYFVKIGPERTCCL
 CDRRATCFSTASDTYACWHHSIGFDYVYNPFMIDVQQWGFIGNLQSNHDLYCQVHGNA
 HVASCAIDMTRCLAVHECFVKRVDWTIEYPIIGDELKINAACRQVQHMVKAALLADK
 FVLHDIGNPKAIKCVFQADVEWKFYDAQPCSDKAYKIEELFYSYATHSDKFTDGVCL
 FWNVNDRYPANSIVCRFDTRVLSNLSLPGCDGGSLYVNKHAFHTPAFDKSAFVNLKQ
 LPPFFYSDSPCESHGKQVSDIDYVPLKSATCITRCNLGGAVCRHHANEYRLYLDAYN
 MMISAGFSLWVYKQFDTYNLWNTFTRLQSLNVAFNVNKGFHDGQGEVPSIINNT
 VYTKVDGVDVELFENKTTLPVNVAFELWAKRNIKPVPEVKILNNLGVDAANTVIWYD
 KRDAPAHISTIGVCSMTDIKKPTETICAPLTVFFDGRVDGQVDFRNARNGVLITEG
 SVKGLQPSVGPQASLNGVTLIGEAVKTFNYKKVDGVVQQLPETYFTQSRNLQEFK
 PRSQMEIDFLELAMDEFIERYKLEGYAFEHIVYGDFFSHSQLGGLHLLIGLAKRFKESF
 FELEDFIPMDSTVKNYFITDAQTGSSKCVCSVIDLLDDFVEI IKSQDLSVSVKVVV
 TIDYTEISFMLWCKDGHVETFPKQSSQAWQPGVAMPNLYKMQRMLLEKCDLQNYGD
 SATLPKGIMNVAKYTQLCQYLNTLTLAVPYNMRVIHFGAGSDKGVAPGTAVLRQWLP
 TGTLLVDSDLNDFVSDADSTLIGDCATVHTANKWDLIISDMYDPKTKNVTKENDSKEG
 FFTYICGFIQQKALGGSVAIKITEHSWNADLYKLMGHFAWWTAFVTNVNASSEAFLE
 IGCNYLGGKPREQIDGYVMHANYIFWRNTNPIQLSSYSLFDMSKFPLKLRGTAVMSLKE
 GQINDMILSLLSKGRLLIIRENNRVVSSDVLVNN"

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	GTGAG	GATGGTC	ATGTGTGGCGG				22
Sbjct	15403	GTGAG	GATGGTC	ATGTGTGGCGG				15424
CDS: ORF1ab polyprote	5056	X	E	M V	M	C	G	G

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](#)

[GenBank](#) [FASTA](#)

[Link To This View](#) | [Feedback](#)



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
TLGVLVPHVGETPIAYRNVLLRKNKNGKAGGHSYGIDLKSYDLGDELGTDPIDYEQN

WNTKHGSGALRELTRELNGGAVTRYVDNFCGPDGYPLDCIKDFLARAGKSMCTLSEQ
LDYIESKRGVYCCRDHEHEIAWFTERSDKSYEHQTPFEIKSAKKFDTFKGECPKFVFP
LNSKVKVIQPRVEKKKTEGFMGRIRSVYPVASPQECNNMHLSTLMKCNHCDEVSWQTC
DFLKATCEHCCTENLVIEGPTTCGYLPTNAVVKMPCPACQDPEIGPEHSVADYHNHSN
IETRLRKGGRTRCFGGCVFAYVGCYNKRAYWVPRASADIGSGHTGITGDNVETLNEDL
LEILSRERNINIVGDFHLNEEVAILASFSASTSAFIDTIKSLDYKSFKTIVESCGN
YKVTGKGPVKGAWNIGQQRSVLTPLCGFPSQAAGVIRSIFARTLDAANHSIPDLQRAA
VTILDGISEQSLRLVDAMVYTSDLLTNSVIIMAYVTGGLVQQTSQWLSNLLGTTVEKL
RPIFEWIEAKLSAGVEFLKDAWEILKFLITGVFDIVKQIQVASDNIKDCVKCFIDVV
NKALEMCIDQVTIAGAKLRSLNLGEVFAQSKGLYRQCIRGKEQLQLLMLPLKAPKEVT
FLEGDSHDTVLTSEEVVLNKELEALETPVDSFTNGAIVGTPVCVNGLMLEIKDKEQ
YCALSPGLLATNNVFRKGGAPIKGVTFGEDTVWEVQGYKNVRITFELDERVDKVLNE
KCSVYTVESGTEVFACVVAEAVVKTLPVSDLLTNMGIDLDEWSVATFYLFDDAGE
ENFSSRMYSFYPPDEEEEDDAECEEEEIDETCEHEYGTEDDYQGLPLEFGASAETVR
VEEEEEEDWLDDETEQSEIEPEPEPTPEEPVNQFTGYLKLTDNVAIKCVDIVKEAQSA
NPMVIVNAANIHLKHGGVAGALNKATNGAMQKESDDYIKLNGPLTVGGSCLLSGHNL
AKKCLHVGPPLNAGEDIQLLKAAYENFNSQDILLAPLLSAGIFGAKPLQSLQVCVQT
VRTQVYIAVNDKALYEQVMDYLDNLKPRVEAPKQEEPPNTEDSKTEEKSVVQKPV DV
KPKIKACIDEVTTTTLEETKFLTNKLLLFADINGKLYHDSQNMLRGEDMSFLEKDAPYM
VGDVITSGDITCVVIPSCKAGGTTEMLSRALKKVPVDEYITTYPGQCAGYTL EEAKT
ALKKCKSAFYVLPSEAPNAKEEILGTVSWNLREMLAHAEETRKLMPICMDVRAIMATI
QRKYKGIKIQEGIVDYGVRFFFYTSKEPVASIIKLNLSLNEPLVTMPIGYVTHGFNLE
EAARCMRSLKAPAVVSVSSPDAVTTYNGYLTSSSKTSEEHFVETVSLAGSYRDWSYSG
QRTELGEVFLKRGDKIVYHTLESPVEFHLGDEVLSLDKLSLLSLREVKTIKVFTTVD
NTNLHTQLVDMSTYQQFGPTYLDGADVTKIKPHVNHEGKTFVFLPSDDTLRSEAFE
YYHTLDESFLGRYMSALNHTKKWFPQVGGLTSIKWADNNCYLSSVLLALQQLEVKFN
APALQEAYYRARAGDAANFCALILAYSNKTVGELGDVRETMTHLLQHANLES AKRVLN
VVCKHCGQKTTTLTGVEAVMYMGTLSYDNLKTGVSIPCVCGRDATQYLVQQESSFVMM
SAPPAEYKQQGTFLCANEYTGNYQCCHYTHITAKETLYRIDGAHLTKMSEYKGPVTD
VFYKETSYYYYIKPVSYKLDGVTYTEIEPKLDGYYKKNAYYTEQPIDLVPTQPLPNA
SEDNFKLTCNFKFADDLNQMTGFTKASRELSVTFPPDLNGDVVAIDYRHSASFKK
GAKLLHKPIVWHINQATTKTTFKPNTWCLRCLWSTKPVDTSNSFEVLAVEDTQGMNDL
ACESQQPTSEEVVENPTIQKEVIECDVKTTEVVGNVILKPSDEGVKVTQELGHEDLMA
AYVENTSITIKKPNELSLALGLKTIATHGIAAINSVPWSKILAYVKPFLGQAAITTSN
CAKRLAQRVFNMPYVFTLLFQLCTFTKSTNSRIRASLPTTIAKNSVKSVAKLCLDA
GINYKSPKFSKLFTIAMWLLLLSICLGS LICVTAAFGVLLSNFGAPSYCNGVRELYL
NSSNVTTMDFCEGSFPCSICLSGLDSLDSYPALETIQVTISSYKLDLITLGLAAEWVL
AYMLFTKFFYLGLSAIMQVFFGYFASHFISNSWLMWFIISIVQMAPVSAMVRMYIFF
ASFYYIWKSYVHIMDGCTSSTCMMCYKRN RATRVECTTIVNGMKRSFYVYANGGRGFC
KTHNWNCLNCDTFCSTFISDEVARDLSLQFKRPINPTDQSSYIVDSVAVKNGALHL
YFDKAGQKTYERHPLSHFVNLDNLRANNTKGS LPINIVFDGKSKCEDESAKSASVYY
SQLMCQPIILLDQVLVSDVGDSTEVSVKMFDAYVDTFSATFSVPMEKLLKALVATAHSE
LAKGVALDGVSTFVSAARQGVVDTDVDTKDVIECLKLSHHSLEVTGDSCNNFMLTY
NKVENMTPRDLGACIDCNARHINAQVAKSHNVSLIWNVKDYMSLSEQLRKQIRSAAKK

NNIPFRLTCATTRQVVNVIITKISLKGKIVSTCFKMLKATLLCVLAAALVCIYVMPV
HTLSIHDGYTNEIIGYKAIQDGVTRDIISTDDCFANKHAGFDAWFSQRGGSYKNDKSC
PVVAAIITREIGFIVPGLPGTVLRAINGDFLHFLPRVFSAVGNICYTPSKLIEYSDEFA
TSACVLAAECTIFKDDAMGKPVPCYDTNLEGSISYSELRPDTRYVLMGDSIIQFPNT
YLEGSVRVVTTFDAEYCRHGTCESEVGCICLSTSGRWVLNNEHYRALSGVFCGV DAMN
LIANIFTPLVQPVGALDVSASVAVAGGIIAILVTCAAYYFMKFRRVFGEYNHVVAANAL
LFLMSFTILCLVPAYSFLPGVYSVFYLYLTFYFTNDVSVFLAHLQWFAMFSPVFPWIT
AIYVFCISLKHCHWFFNNYLKRVMFNGVTFSTFEEAALCTFLLNKEMYLKLRSETLL
PLTQYNRYLALYNYKYFSGALDTTSYREAACCHLAKALNDFSNNGADVLYQPPQTSI
TSAVLQSGFRKMAFSPGKVEGCMVQVTCGTTTLNGLWLDDEVYCPRHVICTAEDMLNP
NYEDLLIRKSNHSFLVQAGNVQLRVIGHSMQNCLLRLKVDTSNPKTPKYKFVRIQPGQ
TFSVLACYNGSPSGVYQCAMRPNYTIKGSFLNGSCGSGVFNIDYDCVSVFCYMHMELP
TGVHAGTDLEKGFYGPVDRQTAQAAGTDTTITLNVLAWLYAAVINGDRWFNRFRTT
LNDNFVAMKYNIEPLTQDHDVILGPLSAQTGIAVLDMCAALKALLQNGMNGRTILGS
TILEDEFTPFVVRQCSGVTFQGKFKKIVKGTHHWMLLFTLSLLILVQSTQWSLFFF
VYENAFLPFTLGIMAIACAMLLVKHKHAFCLCLLPSLATVAYFNMVMPASWVMRI
MTWLELADTSLSGYRLKDCVMYASALVLLIIMTARTVYDDAARRVWTLMNVI TLVYKV
YYGNALDQAI SMWALVISVTSNYSVGVVTTIMFLARAI VFCVEEYPLLFITGNTLQCI
MLVYCFGLYCCCCYFGLFCLLNRYFRLTLGVYDYLVSQTQEFYRMSQGLLPKSSIDA
FKLNKILGIGGKPCIKVATVQSKMSDVKCTSVVLLSVLQQLRVESSSKLWAQCVQLH
NDILLAKDTTEAFEKMSVLLSVLLSMQGAVDINRLCEEMLDNRATLQAIASEFSSLP
YAAAYATAQEAYEQAVANGDSEVVLKLLKSLNVAKSEFDRDAAMQRKLEKMAQAMTQ
MYQARSEDKRAKVT SAMQTMLFTMLRKLNDALNNIINNARDGCVPLNIIPLTAAK
LMVVVDPYGYKNTCDGNTFTYASALWEIQVVDADSKIVQLSEINMNSPNLAWPLI
VTALRANSVAVKQNNELSPVALRQMSCAAGTQTACTDDNALAYYNSKGGRFVLLAL
SDHQDLKWARFPKSDGTGTIYAELEPPCRFVTDTPKGPVKVYLYFIKGLNNLNRGMVL
GSLAATVRLQAGNATEVPANSTVLSFCAFAVDPAKAYKDYLASGGQPITNCVKMLCTH
TGTGQAITVTPEANMDQESFGGASCCLYCRCHIDHPNPKGFCDLKGYVQIPTTCAND
PVGFTLRNTVCTVCGMWKGYGCSCDQLREPLMQSADASTFLNRVCGVSAARLTPCGTG
TSTDVVYRAFDIYNEKVAGFAKFLKTNCCRQEKDEEGNLLDSYFVVKRHTMSNYQHE
ETIYNLVKDCPAVAVHDFKFRVDGDMVPHISRQRLTKYTMADLVYALRHFDEGNCDT
LKEILVTYNCCDDYFNKKDWYDFVENPDI LRVYANLGERVRQSLLKTVQFCDAMRDA
GIVGVLTLDNQDLNGNWDYDFGDFVQVAPGCGVPIVDSYSLMPLITLTRALAAESHM
DADLAKPLIKWDLKYDFTEERLCLFDRYFKYWDQTYHPNCINCLDDRCILHCANFN
LFSTVFPPTSFGPLVRKIFVDGVPFVSTGYHFRELGVVHNQDVNLHSSRSLFKELLV
YAADPAMHAASGNLLLDKRTTCFSVAALTNNVAFQTVKPGNFNKDFYDFAVSKGFFKE
GSSVELKHFFFAQDGNAAISDYDYRYRNLPMTCDIRQLLFVVEVVDKYFDCYDGGCIN
ANQVIVNNLDSAGFPFNKWKARLYYDSMSYEDQDALFAYTKRNVIPITITQMNLYA
ISAKNRARTVAGVSICTMTNRQFHQKLLKSAATRATVIGTSKFGYGGWHNMLKTV
YSDVETPHLMGWDPKCDRAMPNMLRIMASLVLARKHNTCCNLSHRFYRLANCAQVL
SEMVMCGGSLYVKPGGTSSGDATTAYANSVFNICQAVTANVNALLSTDGNKIADKYVR
NLQHRLYECLYRNRDVEHDFVDEFYAYLRKHFSMMILSDDAVVCYNSNYAAQGLVASI
KNFKAVLYYQNNVMSEAKWTEETDLTKGPHEFCSQHTMLVKQDDYVYLPYDPDSRI
LGAGCFVDDIVKTDGTLMIERFVSLAIDAYPLTKHPNQEYADVHFHLYLQYIRKLHDEL

TGHMLDMYSVMLTNDNTSRWEPFYEAMYTPHTVLQAVGVCVLCNSQTSLRGACIR
RFPLCCKCCYDHVISTSHKLVLSVNPYVCNAPGCDVTDVTLQYLGGMSYYCKSHKPI
SFPLCANGQVFLYKNTCVGSDNVTDNFNAIATCDWTNAGDYILANTCTERLKLFAAET
LKATEETFKLSYGIATVREVLSDRELHLSWEVKGKPRPPLNRNYVFTGYRVTKNSKVQI
GEYTFEKGDYGDAVVYRGTTTYKLVNGDYFVLTSHVTMPLSAPTLPQEHYVRITGLY
PTLNI SDEFSSNVANYQKVG MQYSTLQGGPGTGKSHFAIGLALYPSARIVYTACSH
AAVDALCEKALKYLPIDKCSRIIPARARVECFDKFKVNSTLEQYVFCVNALPETTAD
IVVFDEISMATNYDLSVVNARLRAKHVYVYIGDPAQLPAPRTLLTKGTLEPEYFNSVCR
LMKTI GPFDMFLGTCRCPAEIVDTVSALVYDNKLAHKDKSAQCCKMFKYKGVITHDVS
SAINRPQIGVREFLTRNPAWRKAVFISPYNSQNAVASKILGLPTQTV DSSQGSEYDY
VIFTQTETETAHSCNVNRFNVAITRAKIGILCIMS DRDLYDKLQFTSLEIPRRNVATLQ
AENV TGLFKDCSKIIITGLHPTQAPTHLSVDIKFKTEGLCVDIPGIPKDMTYRRLISMM
GFKMNYQVNGYPNMFITREEAIRHVRAWIGFDVEGCHATRDAVGTNLPQLGFSTGVN
LVAVPTGYVDTENNETEFTRVNAKPPPGDQFKHLIPLMYKGLPWNVVRIVQMLSDTL
KGLSDRVVFLWAHGFELTSMKYFVKIGPERTCCLCDKRATCFSTSSDYACWNHSVG
FDYVYNPFMIDVQWQGTGNLQSNHDQHCQVHGNAHVASCDAIMTRCLAVHECFVKRV
DWSVEYPIIGDELRVNSACRKYQHMVVKSAALLADKFPVLHDIGNPKAIKCVPAEVEW
KFYDAQPCSDKAYKIEELFYSYATHHDKFTDGVCLFWNCNVDRYPANAIVCRFDTRVL
SNLNLPGCDGGSLYVNKHAFHTPAFDKSAFTNLKQLPFFYYSDSPCESHGKQVVS D I D
YVPLKSATCITRCNLGGAVCRHHANEYRQYLDAYNMMISAGFSLWIYKQFDTYNLWNT
FTRLQSLENVAYNVVNKGHFDGHAGEAPVSIINNAVYTKVDGIDVEIFENKTTLPVNV
AFELWAKRNKIPVPEIKILNNLGVDIAANTVIWDYKREAPAHVSTIGVCTMTDIAKKP
TESACSSLTVLFDGRVEGQVDLFRNARNGVLITEGSVKGLTPSKGPAQASVNGVTTLIG
ESVKTQFNYPFKKVDGIIQQLPETYFTQSRDLEDFKPRSQMETDFLELAMDEFIQRYKL
EGYAFEHIVYGDFSHGQLGGLHLMIGLAKRSQDSPLKLEDFIPMDSTVKNYFITDAQT
GSSKCVCSVIDLLDDFVEI IKSQDLSVISKVVVKTIDYAEISFMLWCKDGHVETFYF
KLQASQAWQPGVAMPNLYKMQRMLLEKCDLQNYGENAVIPKGIMMNVAKYTQLCQYLN
TLTLAVPYNMRVIHFGAGSDKGVAPGTAVLRQWLP TGTLLVSDLNDFVSDADSTLIG
DCATVHTANKWDLIIISDMYDPRTKHVTKENDSKEGFFTYLCGFIKQKLALGGSIAVKI
TEHSWNADLYKLMGHFSWWTAFVTNVNASSSEAF LIGANYLGKPKQIDGYTMHANYI
FWRNTNPIQLSSYSLFDMSKFPLKLRGTAVMSLKENQINDMIYSLLEKGR LI IRENNR
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

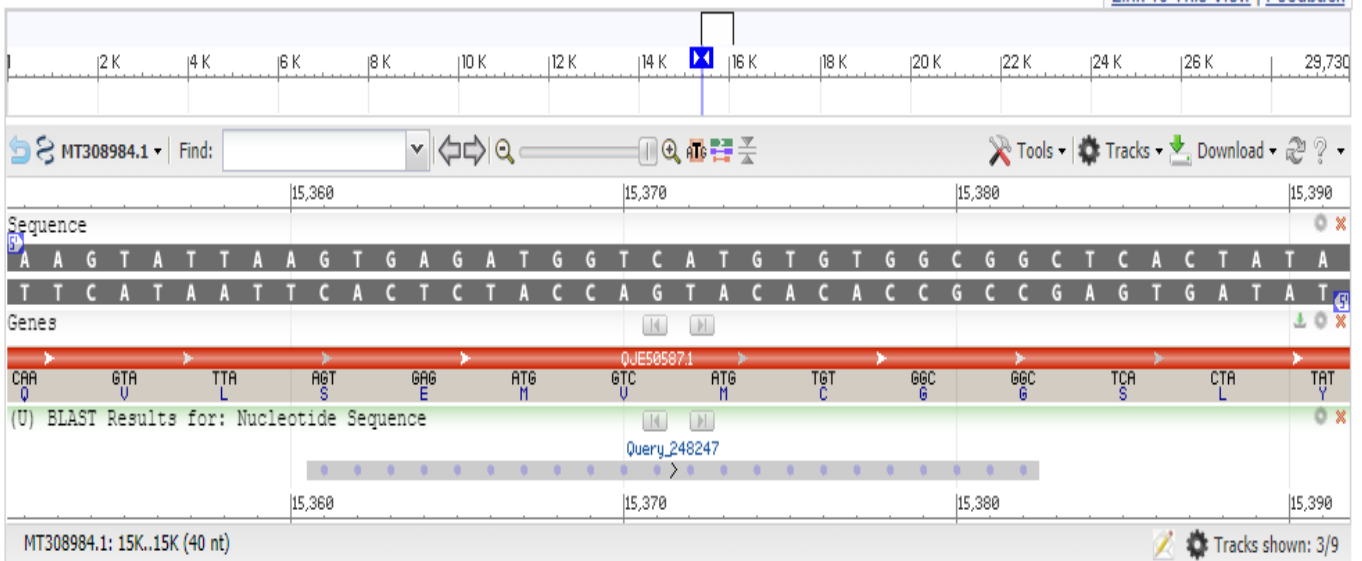
CDS: Putative 1	1	E	M V	M	C	G	G	
Query	1	GTGAG	ATGGTC	ATGTGTGGCGG			22	
Sbjct	15361	GTGAG	ATGGTC	ATGTGTGGCGG			15382	
CDS: ORF1ab [SARS cor	5033	S	E	M V	M	C	G	G

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

GenBank: [MT308984.1](#)

[GenBank](#) [FASTA](#)

[Link To This View](#) | [Feedback](#)



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 lab**"

/protein_id="AVP78030.1"
/translation="MESLVPGFNEKTHGPLSLPVLQVRDVLVRGFGDSVEEALSEARQ
HLKDGTCLVEVEKGVLPQLEQPYVFIKRS DARTAPHGHVMVELVAELDGIQYGRSGE
TLGVLVPHVGEVPVAYRKVLLRKNKNGAGGHSYGADLKSFDLGDDELGTDPIEDFQEN
WNTKHGSGVTRELKRELNGGAYTRYVDNNFCGPDGYPLECIKDLLARAGKASYALSEQ
LDFIDTKRGVYCCREHEHEIAWYTERSEKSYELQTPFEIKLAKKFDTFNGECPNFVFP
LNSTIKTIQPRVEKKKLDGFMGRIRSVYPVASPNECNQMCLSI LMKCDHCGETSWQTG
DFVRATCEFCGTENLTKEGATTCGYLPQNAVVKLYCPACHNPEVGPESHSLAEYHNESG
LKTVLRKGGRTIAYGGCVFAYVGCYNKAYWVPRASANIGCNHTGVVGESES LNDNL
LEILQKEKVNINIVGDFKLNEEIAI ILASFSAISAFVETVKGLDYKTFKQIVESCGN
FKVTKGKAKKGAWNIGEQQSILSPLYAFPSDAAHVRSIFSRLETAHHSVHVLQKAA
I I ILDGISQYSLRLIDAMMFTSDLVTNNLVVMAYITGGVVQMTSQWL TNIFGTVYEKL
KPVLDWLEEKFKEGIEFLRDGWEIVKFISTCACEIVGGQIVTCAKEIKESVQTFKLV
NKFLALCADSIIIGGAKL KALNLGETFVTHSKGLYRKCVRSREETGLLLPLKAPKEII
FLEGETLPTEVLTEEVVLKTGVLQPLEQPTNEAVEAPLIGTPVCINGLMLEIKDTEK
YCALAPNMMVTNNTFTLKGGA PTKVTFGDDTVIEVQGYKSVNITFELDERIDKVLNEK
CSNYTVELGTEVNEFACVVADAVIKTLQPVSELI I PLGIDLDEWSMATYYLFDESGEF
KLSSHMYCSFY PPEDEGEDDCEEGQCEPSTQY EYGTEDDYQGKPLEFGATSFSSSSQE
EEQEEDWLES DSDGQETAVEENKIP SVEVPPVLQVESTPVV TETSEQNNFTGYLKL T
DNVFIKNADIVEEAKKVKPTVVVNAANVYLKHGGGVAGALNKATNNAMQVESDKYITT
NGPLIVGGCVLSGHNLAKNCLHVVGPNVNRGEDIQLLKNAYENFNQHEILLAPLLSA
GIFGADPVHSLRVCVETVHTNVYLVVFDKNLYDKLVSSFLEMKSEKQVEQKVAENPKE
EVKPFPTENKPSVEQRQQAEKKIKASIEEVT TLEETKFLTENLLLYIDINGNLHPD
SATLVKDIDTTF LKKDVPYIVGDVIKEGALTAVVIPTKKAGGTTEMLAKALRKVPTDN
YITTYPGQGLNGYTVEEAKTVLKKCKSAFYILPSI ISNAKQEILGTVSWNLREMLAHA
EETRKLMPVCMETKAIVSTIQRYKGIKIQEGVVDYGARFYFYTSKTTVASLINTLNN
LNETLVTMPLGYVTHGLNLEEAARYMRSLKVPATVSVSSPDAVTAYNGYLTSSSKTPE
EHFIETVSLAGSYKDWSYSGQSTQLGIEFLKRGDKSVYYTSNPTTFHLDGETITFDNL
KTL LSLREVRNIKVF TTVDNINLHTQVVDMSMTYGGQFGPIYLDGADVTKIKPHNSHE
GKTFYVLPND DTLRAEAFEYYHTDPSFLGRYMSALNHTKKWKYPQVNGLTSIKWADN
NCYLATALLT IQQIELKFNPPALQDAYYRARAGEAANFCALILAYCNKTVGELGDVRE
TMNYLFQHANLDSCKRVLNVVCKTCGQQQTTLKGVEAVMYMGTLSEYQLKKGVI PCM
CGKQATQYLVQQESP FVMSAPPAQYELKHGTFVCASEYTGNYQC GHYKHITSKETLY
CIDGALLTKSSEYKGSITDVFYKENSYTTTIKPVTYKLDGVVCTEIDPKLDGYKKN
SYFTEQPIDLVPNQYPNASFDNFK FVCDNTKFADDLNQLSGYKPPASRELKVTFFPD
LNGDVVAIDYKHYTPSFKKGAKLLHKPIVWHVNNATNKATYKPN IWCIRCLWSTKPEVE
TSNSFDALELGDTQGMDNLACEVLKPVSEEVVENPTIQKDILECNVKTTEVVGDII LK
PASDGLKITKEVGHTDLMAAYVDNSSLTIKKPNELSRVLGLKTLATHGLAAINSVPWD
TIANYVKPFLNKVVSTTTNIVTRCLNRVCTNYMPYLF TLLLQLCTFTKSTNSRIRASM
PTTI AKNTVKS VGKFCIEASFNYL KSPNFSKLINIVIFLLLSVCLGSLIYSTAALGV
LMSNLGMP SYCTVYREGYLNSTNVTTATYCTGSI PCSVCLSGLDSLDTYPSLETIQIT
ISSFKWDLTAFGLVAEWFLAYILFTRFFYVLGLAAIMQLFFSYFAVHFISNSWLMWLI
INLVQMAPISAMVRMYIFFASFYYVWKS YVHVVDGCTSSCTMCMCYKRNRATRVECTTI
VNGVRRSFYVYANGGKGFCKLHNWNCINCDTFCAGSTFISDEVARDLSLQFKRPINPT

DQSSYIVDSVTVKNGSIHLYFDKAGQKTYERHLSLSHFVNLDNLRANNTKGS L P I N V I V
FDGKSKCEESSAKSASVYYSQLMCQPILLDDQALVSDVGD SAEVAVKMFDAYVNI FSS
TFNVPMEK L K A L V A T A E A E L A K N V S L D N V L S T F I S A A R Q G F V D S D V E T K D V V E C L K L S
HQSDIEVTGDS C N N Y M L T Y N K V E N M T P R D L G A C I D C S A R H I N A Q V A K S H N I A L I W N V K
DFMSLSEQLRKQIRSAAKNNL P F R L T C A T T R Q V V N V V T T K I A L K G G K I V N N W L K Q L I
KVTLVFLFITVIFYLITPVHVMFKHND F S S E I I G Y K A I D G G V T R D I A S T D T C F A N K H A
DFDSWFSQRGGSYTNDKACPLVA AVITREVG F V V P G L P G T I L R T I N G D F L H F L P R V F S
AVGNICYTPSKLIEYTD F A T S A C V L A A E C T I F K D A S G K P V P Y C Y D T N V L E G S V A Y E S L
R P D T R Y V L M D G S I I Q F P N T Y L E G S V R V V T T F D S E Y C R H G T C E R S E A G I C V S T S G R W V L
N N D Y Y R S L P G V F C G V D A V N L L T N M F T P L I Q P I G A L D I S A S I V A G G V V A I I V T C L A Y Y F
M R F R R A F G E Y S H V V A F N T L L F F M S F T V L C L T P V Y S F L P G V Y S V I Y L Y L T F Y L T N D V S F
L A H I Q W M V M F T P L V P F W M T I V Y V I C I S T K H F Y W F F S N Y L K R R V V F N G V S F S T F E E A A L
C T F L L N K E M Y L K L R S D V L L P L T Q Y N R Y L A L Y N K Y K Y F S G A M D T T S Y R E A A C C H L A K A I
N D F S N S G S D V L Y Q P P Q T S I T S A V L Q S G F R K M A F P S G K V E G C M V Q V T C G T T T L N G L W L D
D V V Y C P R H V I C T S E D M L N P N Y E D L L I R K S N H N F L V Q A G N V Q L R V V G H S M Q N C V L K L K V
D T A N P K T P K Y K F V R I Q P G Q T F S V L A C Y N G S P S G V Y Q C A M R P N F T I K G S F L N G S C G S V G
F N I D Y D C V S F C Y M H H M E L P T G V H A G T D L E G T F Y G P F V D R Q T A Q A A G T D T T I T V N V L A W
L Y A A V I N G D R W F L N R F T T T L N D F N L V A M K Y N Y E P L T Q D H L D I L G P L S A Q T G I A V L D M C
A S L K E L L Q N G M N G R T I L G S A L L E D E F T P F D V V R Q C S G V T F Q S A V K R T I K G T H H W L L L T
V L T S L L V L V Q S T Q W S L F F F V Y E N A F M P F A M G I I A M S A F A M M F V K H K H A F L C L F L L P S L
A T V A Y F N M V Y M P A S W M R I M T W L D I V D T S L S G F K L K D C V M Y A S A V V L L I L M T A R T V Y D
D G A R R V W T L M N V L T L V Y K V Y Y G N A L D Q A I S M W A L I I S V T S N Y S G V V T T V M F L A R G I V F
M C V E Y C P I F F I T G N T L Q C I M L V Y C F L G Y F C T C Y F G L F C L L N R Y F R L T L G V Y D Y L V S T Q
E F R Y M N S Q G L L P P K S I D A F K L N I K L L G I G G K P C I K V A T V Q S K M S D V K C T S V V L L S V L
Q Q L R V E S S S K L W A Q C V Q L H N D I L L A K D T T E A F E K M V S L L S V L L S M Q G A V D I N K L C E E M
L D N R A T L Q A I A S E F S S L P S Y A A Y A T A Q E A Y E Q A V A N G D S E V V L K K L K K S L N V A K S E F D
R D A A M Q R K L E K M A D Q A M T Q M Y Q A R S E D K R A K V T S A M Q T M L F T M L R K L D N D A L N N I I N
N A R D G C V P L N I I P L T T A A K L M V V V P D Y G T Y K N T C D G N T F T Y A S A L W E I Q Q V D A D S K I
V Q L S E I N M D N S P N L A W P L I V T A L R A N S A V K L Q N N E L S P V A L R Q M S C A A G T T Q T A C T D D
N A L A Y Y N N S K G G R F V L A L L S D H Q D L K W A R F P K S D G T G T I Y T E L E P P C R F V T D T P K G P K
V K Y L Y F I K G L N N L N R G M V L G S L A A T V R L Q A G N A T E V P A N S T V L S F C A F A V D P A K A Y K D
Y L A S G G Q P I T N C V K M L C T H T G T G Q A I T V T P E A N M D Q E S F G G A S C C L Y C R C H I D H P N P K
G F C D L K G K Y V Q I P T T C V N D P V G F T L R N T V C T V C G M W K G Y G C S C D Q L R E P M M Q S A D A S T
F L N R V C G V S A A R L T P C G T G T S T D V V Y R A F D I Y N E K V A G F A K F L K T N C C R F Q E K D E E G N
L L D S Y F V V K R H T M S N Y Q H E E T I Y N L V K G C P A V A V H D F F K F R V D G D M V P H I S R Q R L T K Y
T M A D L V Y A L R H F D E G N C D T L K E I L V T Y K C C D D N Y F N K K D W Y D F V E N P D V L R V Y A N L G E
R V R Q A L L K T V Q F C D A M R D A G I V G V L T L D N Q D L N G N W Y D F G D F V Q V A P G C G V P I V D S Y Y
S L L M P I L T L T R A L A A E S H M D A D L A K P L I K W D L L K Y D F T E E R L C L F D R Y F K Y W D Q T Y H P
N C I N C L D D R C I L H C A N F N V L F S T V F P P T S F G P L V R K I F V D G V P F V V S T G Y H F R E L G V V
H N Q D V N L H N S R L S F K E L L V Y A A D P A M H A A S G N L L L D K R T T C F S V A A L T N N V A F Q T V K P
G N F N K D F Y D F A V S K G F F K E G S S V E L K H F F F A Q D G N A A I S D Y D Y Y R Y N L P T M C D I R Q L L
F V V E V V D K Y F D C Y D G G C I N A N Q V I V N N L D K S A G F P F N K W G K A R L Y Y D S M S Y E D Q D A L F
A Y T K R N V I P T I T Q M N L K Y A I S A K N R A R T V A G V S I C S T M T N R Q F H Q K L L K S I A A T R G A T
V V I G T S K F Y G G W H N M L K T V Y S D V E S P H L M G W D Y P K C D R A M P N M L R I M A S L V L A R K H S T

CCNLSHRFYRLANECAQVL**SEMVMCGG**SLYVKPGGTSSGDATTAYANSVFNICQAVTA
NVNALLSTDGNKIADKYVRNLQHRLYECLYRNRDQEFVDEFYAYLRKHFSSMILSD
DAVVCYNSNYAAQGLVASIKNFKAVLYYQNNVMSEAKCWTETDLTKGPHEFCSQHTM
LVKQGDDYVYLPYPDPSRILGAGCFVDDIVKTDGTLMIERFVSLAIDAYPLTKHPNQE
YADVFLYLQYIRKLDHDELGHMLDMYSVMLTNDNTSRYWEPEFYEAMYTPHTVLQAV
GACVLCNSQTSRLRCGACIRRPFLCCKCCYDHVISTSHKLVLSVNPYVCNAPGCDVTDV
TQLYLGGMSYCKSHKPPISFPLCANGQVFLYKNTCVGSDNVDNFNAIATCDWTNAG
DYILANTCTERLKLFAAETLKATEETFKLSYGIATVREVLSRELHLSWEVKGPRPPL
NRNYVFTGYRVTKNSKVQIGEYTFEKGDYGDVAVYRGTTTYKLVNGDYFVLTSHVMP
LSAPTLVPQEHYVRITGLYPTLNI SDEFSSNVANYQKVGMMQKYSTLQGPPTGKSHFA
IGLALYPSARIVYTACSHAAVDALCEKALKYLPIDKCSRIIPARARVECFDKFKVNS
TLEQYVFCVNALPETTADIVVFEI SMATNYDLSVNNARLRKHYVYIGDPAQLPAP
RTLLTKGTLEPEYFNSVCRMLKTI GPDMFLGTCCRCPAEIVDTVSALVYDNKLKAHKE
KSAQCCKMFYKGVITHDVSSAINRPQIGVVREFLTRNPAWRKAVFISPYNSQNAVASK
ILGLPTQTVDSSQGSEYDYVIFTQTETETAHSCNVNRFNVAITRAKIGILCIMS DRDLY
DKLQFMSLEVPRRNVATLQAENV TGLFKDCSKIITGLHPTQAPTHLSVDTKFKTEGLC
VDIPGIPKDMTYRRLISMMGFKNYQVNGYPNMFITREEAIRHVRWAGFDVEGCHAT
RDAVGTNLPLQLGFSTGVNLVAVPTGYVDTEHSTEFTRVNAKPPPGDQFKHLIPLMYK
GLPWNVVRIKIVQMLS DTLKGLSDRVVFLWAHGFELTSMKYFVKIGPERTCCLCDKR
ATCFSTSSDTYACWNHSGFDYVYNPFMIDVQQWGF TGNLQSNHDQHCQVHGNAHVAS
CDAIMTRCLAVHECFVKRVDWSVEYPIIGDELKINAACRQVHMVVKSAALLADKFPVL
HDIGNPKAIRCVPQSEVDWKFYDAQPCSDKAYKIEELFYATHHDKFTDGVCLFWNC
NVDRYPANAI VCRFDTRVLSNLNLP GCDGGS LYVNKHAFHTPAFDKSAFTHLQKLPFF
YYSDSPCESHGKQVSDIDYVPLKSATCITRCNLGGAVCRHHANEYRQYLDAYNMMIS
AGFSLWIYKQFDTYNLWNTFTKLQSL ENVAYNVINKGHFDGQNGEAPVSI VNNNAVYTK
LDGVDVEIFENKTTLPVNVAFELWAKRNIKPVPEIKILNNLGVDIAANTVIWDYKREA
PAHVSTIGVCTMTDI AKKPTESACSSLTVLFDGRVEGQVDFRNARNGVLITEGSVKG
LTPSKGPAQASVNGVTLIGESVKTQFN YFKKVDGI IQQLPETYFTQSRDLEDFKPRSQ
METDFLELAMDEFIERYKLDGYAFEHIVYGD FSHGQLGGLHLLIGLAKRSQD SLLKLE
DFIPMDSTVKNYFITDAQTGSSKCVCSVIDLLLD FVEI IKSQDLSVSVKVKVTIDY
TEISFMLWCKDGHVETFY PKLQSSQAWQPGVAMPNLYKMQRMLLEKCDLQNYGDSAIL
PKGIMMNVAKYTQLCQYLNLT LLA VPYNMRVIHFGAGSDKGVAPGTA VLRQWLPTGTL
LVDSDLNDFVSDADSTLIGHCATVHTANKWDLI ISDMYDPKTKNVTKENDSKEGFFTY
ICGFIQQKLALGGSVAVKITEHSWNADLYKLMGHFAWWTAFVTNVNASSSEAF LIGCN
YLGKPREQIDGYVMHANYIFWRNTNPIQLSSYS LFDMSKFPLKLRGTAVMSLKD GQIN
DMILSLLSKGR LIVRENNRVI ISSDVLVNN "

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	GTGAG	GATGGTC	ATGTGTGGCGG			22	
Sbjct	15418	GTGAG	GATGGTC	ATGTGTGGCGG			15439	
CDS: non-structural p	5052	S	E	M V	M	C	G	G

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

GenBank: MG772933.1

[GenBank](#) [FASTA](#)

[Link To This View](#) | [Feedback](#)

MG772933.1: 15K..15K (40 nt)

Tracks shown: 2/5

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

CDS

```
/gene="ORFlab"  
<1..>22  
/gene="ORFlab"  
/ribosomal_slippage  
/codon_start=3  
/product="ORFlab polyprotein"  
/protein_id="QPF55728.1"  
/translation="MESLVPGFNEKTHVQLSLPVLQVRDVLVRGFGDSVEXXSXXRQ  
XLKDGTCGLVEVEKGVLPQLEQPYVXIKRS DARTAPXGHVMVXXXXXXXXXXXXXXXXXX  
XXXXXXXXPHVGEIXXXXXXXXXXXXXXXXXXXXXSYGADLKSFDLGDELGTDPYEDFQEN  
WNTKHSSGVTRELMRELNGGAYTRYVDNNFCGPDGYPLECIKDLLARAGKASCTLSEQ  
LDFIDTKRGVYCCREHEHEIAWYTERSEKSYELQTPFEIKLAKKFDTFNGECPNFVFP  
LNSIIKTIQPRVEKKKLDGFMGRIRSVYPVASPNECNQMCLSTLMKXXXXXXXXXXXXX  
XXXXXXXXXXXXXXXXXXXXTCGYLPXNAVVKIYCPACHNSEVGEHSLAEYHNESG  
LKTILRKGGRITAFGGCVFSYVGCHNKCAWVPRASANIGCNHTGVVGEGLNDNL  
LEILQKEKVNINIVGDFKLNEEIAIILASFSASTSAFVETVKGLDYKAFKQIVESCGN  
FKVTKGKAKKAWNIGEQKLSPLYAFASEAARVRSIFSRLETAQNSVRVLQKAA  
ITILDGISQYSLRLIDAMMFTSDLATNNLVMAYITGGVVQLTSQWLTNIFGTVYEKL  
KPVLDWLEEKFEKGEVFLRDGWEIVKFI STCAXXXXXXXXXXXXXXXXXXXXXXFKLV  
NKXXXXXXXXXIIGGAKLALNLTGETFVTHSKGLYRKCCKSREETGLLMPLKAPKEII  
FLEGETLPTXLTXVVLKTDGLQPLEQPTSEAVEAPLVGTPVCINGLMLLEIKDTEK  
YCALAPNMMVTNNTFTLKGGAPTKVTFGDDTVIEVQGYKSVNITFELDERIDKVLNEK  
CSAYTVELGTEVNEFACVVADAVIKTLQPVSELLTPLGIDLDEWSMATYYLFDSEGEF  
KLASHMYCSFYPPDEDEEEGDCEEEEFEPSTQYEGTEDDYQGKPLEFGATSAALQPE  
EEQEEDWLDDDSQQT VGGQDGS EDNQT TTIQTIVEVQPQLEMELTPVVQTI EVNSFSG  
YLKLTDNVYIKNADIVEEAKVKPTVVVNAANVYLKHGGGVAGALNKATNNAMQVESD  
DXIATNGPLKVGSCVLSGHNLA KHCLHVGPVNKGEDIQLLKSAYENFNQHEVLLA  
PLLSAGIFGADPIHSLRVCVDTXXXXYXAXFDKNLYDKLVSSFLEMKSEKQVEQKIA  
EIPKEEVKPFITESKPSVEQRKQDDKIKACVEEVTTLEETKFLTENLLYIDINGN  
LHPDSATLVSDIDITFLKKDAPYIVGDVVQEGVLTAVVIPTKKAGGTTEMLAKALRV  
PTDNYITTYPGQLNGYTV EEAKTVLKCKSAFYIILPSIISNEKQEILGTVSWNLREM  
LAHAEEETRKLMPVCVETKAI VSTIQRYKGIKIQEGVVDYGARFYFYT SKTTVASLIN  
TLNDLNETLVTMPLGYVTHGLNLEEAARYMRS LKVPATVSVSSPDAVTAYNGYLTSSS  
KTPEEHFIETISLAGSYKDWYSQSTQLGIEFLKRGDKSVYYSNPTTFHLDGEVIT  
FDNLKTL LSLREVRTIKVFTTVDNINLHTQVVDM SMTYGGQFGPTYLDGADVTKIKPH  
NSHEGKTFYVLPND DTLRVEAFEYHTTDP SFLGRYMSALNHTKKWKYPQVNGLTSIK  
WADNNCYLATALTLQQIELKFNPPALQDAYYRARAGEAANFCALILAYCNKTVGELG  
DVRETM SYLFQHANLDSCKRVLNVVCKTCGQQQTTLKGEAVMYMGTL SYEQFKKG VQ  
IPCTCGKQATKYL VQQESPFVMMSAPPAQYELKHGTF TCASEYTGNYQCGHYKHITSK  
ETLYCIDGALLTKSSEYKGPITDVFYKENSYTTTIKPVTYKLDGVVCTEIDPKLDNYY  
KKDNSYFTEQPIDLVNPQYPNASFDNFKFVCDNIKFADDLNQLTGYKPKPASRELKVT  
FFPDLNGDVVAIDYKHYTPSFKKGAKLLHKPIVWHVNNATNKATYKPNWTCIRCLWST  
KPVETSNSFDVLKSEDAQGM DN LACEDLKPVSEEVENPTIQKDVLECNVKTTEVVG D  
IILKPANNSLKITEEVGHTDLMAAYVDNSSLTIKKNELSRVLGLKTLATHGLAAVNS
```

VPWDTIANYAKPFLNKVSTTTNIVTRCLNRVCTNYMPYFFTLQLCTFTRSTNSRI
KASMPPTIAKNTVKSVGKFCLEASFNYLKS PNF SKLINIIWFLLLSVCLGSLIYSTA
ALGLVMSNLGMP SYCTGYREGYLNSTNVTIATYCTGSI PCSVCLSGLDSDLTYP SLET
IQITISSFKWDLTAFGLVAEWFLAYILFTRFFYVLGLAAIMQLFFSYFAVHFISNSWL
MWLIINLVQMAPISAMVRMYIFFASFYVWKS YVHVVDGCNSSTCMCCYKRNRATRVE
CTTIVNGVRRSFYVYANGGKGFCKLHNWNCVNCDFCAGSTFISDEVARDSLQFKRP
INPTDQSSYIVDSVTVKNGSIHLYFDKAGQKTYERHSLSHFVNLDNLRANNTKGS LPI
NVIVFDGKSKCEESSAKSASVYYSQ LMCQPI LLLDQALVSDVGD SAEVAVKMF DAYVN
TFSSTFNVPMEK LKTLVATAEAE LAKNVSLDNVLS TFI SAARQGFVSDVETKDVVEC
LKLSHQSDIEVTGDCSNMYMLTYNKVENMTPRDLGACIDCSARHINAQVAKSHNIALI
WNVKDFMSLSEQLRKQIRSAAKNNLPFKLTCAATTRQVNVVTTKIALKGGKIVNNWL
KQLIKVTLVFLFVXXIFYLITPVHVMSKHTDFSSEIIGYKAIDGGVTRDIASDTDCFA
NKHADFDTWFSQRGGSYTNDKACPLIAAVITREVG FVVPGLPGTILRTTNGDFLHFLP
RVFSAVGNICYTPSKLIEYTD FATSACVLAECTIFK DASGKVPYCYDTNVLEGSVA
YESLRPDTRYV LMDGSI IQFPNTYLEG SVRVVTTFDSEYCRHGT CERSEVGVCVSTSG
RWV LNNDYRSLPGVFCGVDVAVNLLTNMFTPLIQPIGALDISASIVAGGIVAIVVTCL
AAYFMRFRRAFG EYSHVVAFN TLLFLMSFTVLC LTPVYSFLPGVYSVIYLYLTFYLTN
DVSFLAHIQWVMVMTPLVPFWIT IAYIICISTKH FYWFFSNYLKRRVVFNGVSFSTFE
EAALCTFLLNKEMYLKLRSDVLLPLTQYNRYLALYNKYKYFSGAMDTTSYREACCHL
AKALNDFSNSGSDVLYQPPQTSITSAVLQSGFRKMAFP SGKVEGCMVQVTCGTTTLNG
LWLDDVVYCPRHVIC TSEDMLNP NYEDLLIRKSNHNFLVQAGNVQLRVIGHSMQNCVL
KLKVD TANPKTPKYKFVRIQPGQTF SVLACYNGSPSGVYQCAMRPNFTIKGSFLNGSC
GSVGFNIDYDCVSFCYMHMELPTGVHAGTDLEGNFYGPFVDRQTAQAAGTDTTITVN
VLAWLYAAVINGDRWFLNRFTTTLXXXLVAMKYNIEPLTQDHVDILGPLSAQTGIAV
LDMCASLKELLQNGMNGRTI LGSALLEDEFTPF DVVRQC SGVTFQSAVKRTIKGTHHW
XXXXILTSLLVLVQSTQWSLFFFLYENAF LPFAMGIIAMSAFAMMFVKHKHAF LCLFL
LPSLATVAYFNMYMPASWVRIMTWLDMVDTSLSGFKLDCVMYASAVVLLILMTAR
TVYDDGARRVWTL MNVLT LVYKVYYGNALDQAISMWALII SVTSNYS GVVTTVMFLAR
GIVFMCVEYCP IFFITGNTLQCIMLVYCF LYG YFCTCYFGLFCLLNRYFRLTLGVYDYL
VSTQEFRYMSQGLLPPKNS IDAFKLN IKLLGVGGKPCIKVATVQSKMSDVKCTSVVL
LSVLQQLRV ESSSKLWAQCVQLHNDILLAKDTTEAF EKMSL LSVLLSMQGAVDINKL
CEEMLDNRATLQAIASEFSS LPSYAAFATAQEAYEQAVANGDSEVVLK LKLSLNVAK
SEFDRDAAMQRKLEK MADQAMTQMYQARSEDKRAKVT SAMQTM LFTMLRKLNDALN
NIINNARDGCVPLNIIPLTTAAKLMVVIPDYNTYKNTCDGTTFTYASALWEIQQV VDA
DSKIVQLSEI SMDNSPNLAWPLI V TALRANS AVKLQNNELSPVALRQM SCAAGTTQTA
CTDDNALAYYNTTKGGRFVLALLSDLQDLKWARFPKSDGTGTIYTELEPPCRFVTDTP
KGPKVKYLYFIKGLNNLRGMVLGSLAATVRLQAGNATEVPANSTVLSFCFAVDAAK
AYKDYLASGGQPI TNCVKMLCTHTGTGQAITVTP EANMDQESFGGASCCLYCRCHIDH
ENPKGFCDLKGKYVQIPTTCANDPVGFTLKNVTCTVCGMWKYGCSCDQLRE PMLQSA
DAQSFLNRVCGVSAARLTPCGTGTSTDVVYRAFDIYNDKVAGFAKFLKTNCCRFQEKD
EDDNLIDSYFVVKRHTFSNYQHEETIYNLLKDCPAVAKHDFKFRIDGDMVPHISRQR
LTKYTMADLVYALRHFDEGNCDTLKEILVTYNCCDDYFNKKDWYDFVENPDILRVYA
NLGERVRQALLKTVQFC DAMRNAGIVGLTLDNQLDNGNWDYDFGDFIQTTPGSGVPVV
DSYYSLLMPI LTLTRALTAESHVDTDLTKPYIKWDL LKYDFTEERLKLFD RYFKYWDQ

TYHPNCVNCDDRCILHCANFNVLFSTVFPPLTSFGPLVRKIFVDGVPFVVSTGYHFRE
LGVVHNQDVNLHSSRSLSFKELLVYAAXPAMHAASGNLLLDKRTTCFSVAALTNNAFQ
TVKPGNFNKFDFYDFAVSKGFFKEGSSVELKHFFFAQDGNAAISDYDYRYNLPMTCDI
RQLLFVVEVVDKYFDCYDGGXINANQVIVNNLDKSAGFPFNKWKARLYYDSMSYEDQ
DALFAYTKRNVIPITITQMNLYAISAKNRARTVAGVSICTMTNRQFHQKLLKSIAT
RGATVVIIGTSKFGYGGWHNMLKTVYSDVENPHLMGWDPKCDRAMPNMLRIMASLVLAR
KHTTCCSLSHRFYRLANCAQVL**SEMVMCGG**SLYVKPGGTSSGDATTAYANSVFNICQ
AVTANVNALLSTDGNKIADKYVRNLQHRLYECLYRNRDVTDFVNEFYAYLRKHFSSM
ILSDDAVVCFNSTYASQGLVASIKNFKSVLYYQNNVMSEAKCWTEETDLTKGPHEFCS
QHTMLVKQGGDYVYLPYDPDSRILGAGCFVDDIVKTDGTLMIERFVSLAIDAYPLTKH
PNQEYADVHLYLQYIRKLHDELGHMLDMYSVMLTNDNTSRYWEPEFYEAMYPHTV
LQAVGACVLCNSQTSLRCGACIRPFPLCCKCCYDHVISTSHKLVLSVNPYVCNAPGCD
VTDVTLQYLGLGMSYCKSHKPPISFPLCANGQVFLYKNTCVGSDNVTFDFAIATCDW
TNAGDYILANTCTERLKLFAAETLKATEETFKLSYGIATVREVLSRELHLSWEVGP
RPPLNRNYVFTGYRVTKNSKVQIGEYTFEKGDYGDVVYRGTTTYKLVNGDYFVLTSH
TVMPLSAPTLPQEHYVRITGLYPTLNI SDEFSSNVANYQKVMQKYSTLQGFPGTGK
SHFAIGLALYPSARIVYTACSHAARDALCEKALKYLPIDKCSRIIPARARVECFDKF
KVNSTLEQYVFCVNALPETTADIVVFDEISMATNYDLSVVNARLRAKHVYIGDPAQ
LPAPRTLLTKGTLEPEYFNVCRLMKTIGPDMFLGTCRRCPAEIVDTSALVYDNKLG
AHKDKSAQCCKMFYKGVITHDVSSAINRPQIGVVREFLTRNPAWRKAVFISPYNSQNA
VASKILGLPTQTVDSQSEYDYVIFTQTTETAHSCNVNRFNVAITRAKVGILCIMS
RDLYDKLQFTSLEIPRRNVATLQAENVTLGLFKDCSKVITGLHPTQAPTHLSVDTKFKT
EGLCVDIPGIPKDMTYRRLISMMGFKNYQVNGYPNMFITREEAIRHVRAWIGFDVEG
CHATREAVGTNLPLQLGFSTGVNLVAVPTGYVDTPNNTDFSRVSAKPPPGDQFKHLIP
LMYKGLPWNVRIKIVQMLSDTLKNLSDRVFVLWAHGFELTSMKYFVKIGPERTCCL
CDRRATCFSTASDTYACWHHSIGFDYVYNPFMIDVQQWGFNTGNLQSNHDLYCQVHGNA
HVASCDAIMTRCLAVHECFVKRVDWTIEYPIIGDELKINAACRQVQHMVKAALLADK
FPVLHDIGNPKAIKCVFQADVEWKFYDAQPCSDKAYKIEELFYSYATHSDKFTDGVCL
FWNCNVDRYPANSIVCRFDTRVLSNLNLPDGGSLYVKNHAFHTPAFDKSAFVNLKQ
LPPFFYSDSPCESHGKQVSDIDYVPLKSATCITRCNLGGAVCRHHANEYRLYLDAYN
MMISAGFSLVYKQFDTYNLWNTFTRLQSLNVAFNVVNKGHFDGQGEVPSIINNT
VYTKVDGVDVELFENKTTLPVNVAFELWAKRNIKPVPEVKILNNLGVDAANTVIWYD
KRDAPAHISTIGVCSMTDIKKPTETICAPLTVFFDGRVDGQVDFRNARNGVLITEG
SVKGLQPSVGPQASLNGVTLIGAVKTFNYKKVDGVXQXXXXXXXXTQSRNLQEFK
PRSQMEIDFLELAMDEFIERYKLEGYAFEHIVYGFSDSHSQLGGLHLLIGLAKRFKESF
FELEDFIPMDSTVKNYFITDAQTGSSKCVCSVIDLLDDFVEI IKSQDLSVSVKVVV
TIDYTEISFMLWCKDGHVETFPKQSSQAWQPGVAMPNLYKMQRMLEKCDLQNYGD
SATLPKGIMNVAKYTQLCQYLNTLTLAVPYNMRVIHFGAGSDKGVAPGTAVLRQWLP
TGTLLVDSDLNDFVSDADSTLIGDCATVHTANKWDLIISDMYDPKTKNVTKENDSKEG
FFTYICGFIQKALALGGSVAIKITEHSWNADLYKLMGHFAWWTAFVTNVNASSEAF
LIGCNLYGKPREQIDGYVMHANYIFWRNTNPIQLSSYSLFDMSKFPPLKRGTAVMSLKE
GQINDMILSLLSKGRLLIIRENNRVVISSDVLVNN"

mat peptide

<1..>22

/gene="ORF1ab"

[/product="RNA-dependent RNA polymerase"](#)

ORIGIN

1 gtg**ar**atggt 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](#)

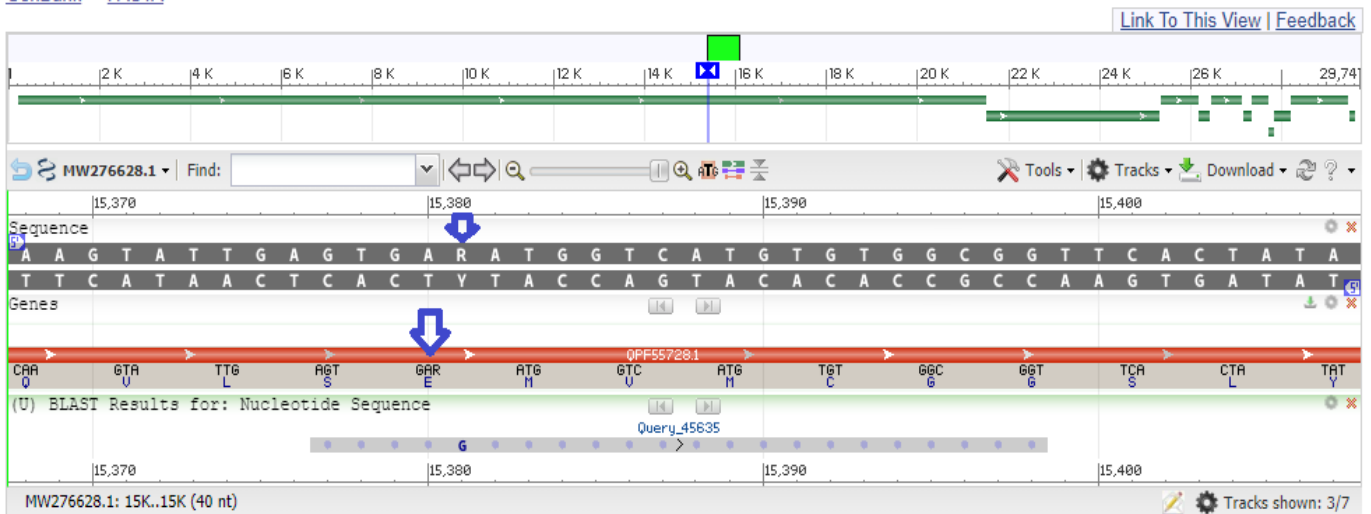
Score	Expect	Identities	Gaps	Strand
40.1 bits(20)	0.59	21/22(95%)	0/22(0%)	Plus/Plus
CDS: Putative 1	1	E M V M C G G		
Query	1	GTG A R A T G G T C A T G T G T G G C G G	22	
Sbjct	15377	GTG A R A T G G T C A T G T G T G G C G G	15398	
CDS:ORF1ab polyprote	5056	S E M V M C G G		

ATTENZIONE: R = A O 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

[GenBank](#) [FASTA](#)



**5.3 Ricerca BLAST su una SEQUENZA di RNA del Gene
 “GTGAGATGGTCATGTGTGGCGG” RdRp del “Bat SARS-like
 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	Rhinolophus monoceros coronavirus	44.1	44.1	100%	0.037	100.00%	375

cds

.....
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	
.....

CAPITOLO VI

**Sostituzioni di BASI
nella SEQUENZA di RNA
“GTGAAATGGTCATGTGTGGCGG”
del GENE RdRp
(RdRp_SARSr-F)
e successive ricerche
BLAST**

OBIETTIVI DEL CAPITOLO VI

Mentre nel capitolo precedente sono state effettuate le ricerche BLAST sulle sequenze “GTGAAATGGT**T**ATGTGTGGCGG” e “GTGAG**A**TGGTCATGTGTGGCGG” (le due sequenze del gene RdRp i cui allineamenti significativi delle ricerche BLAST identificano il *Pipistrello* e il *Pangolino*), in questo capitolo, per non trascurare l’impatto di eventuali altre variazioni di basi azotate nelle stesse posizioni evidenziate in blu nella sequenza “GTGAAATGGT**C**ATGTGTGGCGG” ed identificare eventuali altri “organismi” rilevanti per le finalità di questo studio, sono state eseguite **tutte le sostituzioni di basi possibili**, ottenendo come risultato le seguenti altre 4 sequenze:

“GTGAAATGGT**A**ATGTGTGGCGG”, “GTGAAATGGT**G**ATGTGTGGCGG”

“GTGAT**T**ATGGTCATGTGTGGCGG”, “GTGAC**A**TGGTCATGTGTGGCGG”

Queste sequenze sono state oggetto di **quattro Ricerche BLAST complete**. Quelli che seguono sono gli “**organismi**” che sono stati riscontrati in “**comune**” tra queste *quattro sequenze e altre sequenze* mostrate nei capitoli successivi. Per essere più facilmente distinguibili, sia scorrendo il documento, sia utilizzando la funzione “**Trova**”, gli “**organismi comuni**” sono stati opportunamente *colorati in maniera diversa l’uno dall’altro*.

1) Sequenza “GTGAAATGGT**A**ATGTGTGGCGG”:

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

2) Sequenza “GTGAAATGGT**G**ATGTGTGGCGG”:

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

3) Sequenza “GTGAT**T**ATGGTCATGTGTGGCGG”:

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

4) Sequenza “GTGAC**A**TGGTCATGTGTGGCGG”:

Human DNA.

6.1 Ricerca BLAST sulla SEQUENZA modificata di RNA

“GTGAAATGGT**A**ATGTGTGGCGG” del RdRp Gene – (**Hitlist size: 1000**)

Query: GTGAAATGGT**A**ATGTGTGGCGG

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), 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

VISTA DI ALLINEAMENTO VERTICALE AGGANCIATO A QUERY CON LETTERE D'IDENTITA'

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 Ricerca BLAST sulla SEQUENZA modificata di RNA

“GTGAAATGGT**GATGTGTGGCGG**” del RdRp Gene – (Hitlist size: 1000)

Query: GTGAAATGGT**GATGTGTGGCGG**

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),	Severe acute respiratory syndrome coronavirus 2	36.2	36.2	100%	9.2	95.45%	29813

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	Gene/Region	Species	Length	Score	E-value	Identity	Positives	Bits
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 shepardii	36.2	36.2	81%	9.2	100.00%	52995

VISTA DI ALLINEAMENTO VERTICALE AGGANCIATO A QUERY CON LETTERE D'IDENTITA'

Query range 1: 1 to 22

Query	1	GTGAAATGGTGATGTGTGGCGG	22
LR778259.1	43904207	GTGAAATGGT G ATGTGTGG	43904189
LR778259.1	23381906	TGAAATGGT G ATGTG	23381920
LR722978.1	59908563	TGAAATGGT G ATGTGTGGC	59908581
LR722978.1	33466936	GAAATGGT G ATGTGTGG	33466952
LR722978.1	52466582	AAATGGT G ATGTGTGG	52466597
LR722978.1	88735259	ATGGT G ATGTGTGGCG	88735274
LR722978.1	34565888	AAATGGT G ATGTGTG	34565902
LR722978.1	43597785	TGAAATGGT G ATGTG	43597799
LR722978.1	48408969	GAAATGGT G ATGTGT	48408955
LR722978.1	51014259	GAAATGGT G ATGTGT	51014245
LR722978.1	60018449	TGAAATGGT G ATGTG	60018463
LR722978.1	81496758	GAAATGGT G ATGTGT	81496744
LR722978.1	91961262	TGAAATGGT G ATGTG	91961276
LR537138.1	35644928	TGAAATGGT G ATGTGTGGC	35644946
LR537138.1	2734107	TGGT G ATGTGTGGCG	2734093
LR537138.1	28927756	AATGGT G ATGTGTGG	28927770
CP025673.1	4838415	AAATGGT G ATGTGTGGCGG	4838433
CP025657.1	4115448	AAATGGT G ATGTGTGGCGG	4115466
XR_002546422.1	2426	GTGAAATGGT G ATGTGTGG	2444
XM_022083378.1	2430	GTGAAATGGT G ATGTGTGG	2448
XM_022083377.1	3276	GTGAAATGGT G ATGTGTGG	3294
HG975450.1	59839601	GAAATGGT G ATGTGTGGCG	59839619
HG975450.1	41067225	AAATGGT G ATGTGTGG	41067240
MW444836.1	15393	GTGAAATGGT C ATGTGTGGCGG	15414
CP054686.1	24319927	GTGAAATGGT G ATGTGTG	24319910
CP054686.1	5164724	GTGAAATGGT G ATGTGT	5164740
CP054686.1	24402614	GTGAAATGGT G ATGTGT	24402598
CP054686.1	10548306	TGGT G ATGTGTGGCG	10548292
OC005272.1	4696	TGAAATGGT G ATGTGTGG	4679

6.3 Ricerca BLAST sulla SEQUENZA modificata di RNA

“GTGATATGGTCATGTGTGGCGG” del 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 PREDICTED: Corvus brachyrhynchos uncharacterized LOC108447881	milkfish	38.2	98.6	86%	2.2	100.00%	53690977
XR_001863586.1	(LOC108447881), ncRNA 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	American crow	38.2	38.2	86%	2.2	100.00%	3419
MW444836.1		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

VISTA DI ALLINEAMENTO VERTICALE AGGANCIATO A QUERY CON LETTERE D'IDENTITA'

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	GTGAATGGTCATGTGTGGCGG	15414
OD009150.1	17979	GTGATATGGTCATGTGTG	17996
OA566508.1	78403	GTGATATGGTCATGTGTG	78420

6.4 Ricerca BLAST sulla SEQUENZA modificata di RNA

“GTGACATGGTCATGTGTGGCGG” del 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 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	human	38.2	38.2	86%	2.2	100.00%	184302
MW444836.1		Severe acute respiratory syndrome coronavirus 2	36.2	36.2	100%	9.2	95.45%	29813
.....

VISTA DI ALLINEAMENTO VERTICALE AGGANCIATO A QUERY CON LETTERE D'IDENTITA'

Query range 1: 1 to 22

Query	1	GTGACATGGTCATGTGTGGCGG	22
AL162393.15	144164	GTGACATGGTCATGTGTGG	144182
MW444836.1	15393	GTGA A ATGGTCATGTGTGGCGG	15414

CAPITOLO VII

**PROCEDURA PER LA
MODIFICAZIONE DELLA
SEQUENZA ORIGINARIA
E LA GENERAZIONE DI
NUOVE SEQUENZE**

OBIETTIVI DEL CAPITOLO VII

In questo capitolo è trattato l'importante e rivoluzionario argomento delle *analisi e delle modificazioni di sequenze di DNA o RNA* basato sui principi della **TEORIA TRICROMATICA DELL'EQUILIBRIO DEI SISTEMI (T.T.E.S.)**: la **procedura** attraverso la quale **modificare la Sequenza Originaria** e **generare nuove sequenze di DNA o RNA**.

Con questa procedura si introduce una revisione profonda dell'attuale modalità di intendere la **struttura costitutiva di una sequenza di DNA o RNA** e del suo **rapporto** tra essa e altre diverse e possibili sequenze di DNA o RNA.

In particolare, è spiegato il concetto di **“Trend Non Manifesto”** di una specifica sequenza di DNA o RNA e la modalità attraverso cui esso si moltiplica in **numerosi “Trend Non Manifesti”**.

La **sequenza originaria** differisce da **tutte le nuove sequenze generate** per almeno il **70% di basi** e quindi si riferisce a **materiale biologico molto differente**.

Questo dato è di fondamentale importanza perché in questo scritto si afferma che, nonostante la notevole differenza di basi tra la **sequenza originaria** e le **nuove sequenze generate**, esse hanno **“caratteristiche” comuni** proprio perché i loro TREND sono **STRUTTURALMENTE IDENTICI**.

Il capitolo si conclude con una proposta particolare di applicazione del **codice binario** alla lettura del **codice genetico** attraverso la quale la **complessità di quest'ultimo** aumenta notevolmente, **evidenziando somiglianze e differenze attualmente ignote tra le sequenze**, ma incrementando significativamente anche il numero di informazioni che si potrebbero ottenere se avessimo una *banca dati* che prendesse in considerazione i **Trend** e tutto ciò che è ad essi correlato.

7.1 MODIFICAZIONE PARZIALE DI UNA SEQUENZA ORIGINARIA E GENERAZIONE DI NUOVE SEQUENZE

Come accennato nell'**Introduzione**, la **procedura completa** per analizzare e modificare una sequenza di DNA (o di RNA) prevede **NOVE FASI** (<http://www.ttesystems.eu/Cms/files/Pdf/Introduzione%20DNA%20ITA.pdf>).

A causa del poco tempo avuto a disposizione per la stesura di questo documento e della enorme quantità di dati presentati, alcune *Fasi* sono state omesse e, tra esse, la **Prima Fase**, l'**Analisi della Sequenza Originaria**.

In questo paragrafo è spiegata la procedura per la realizzazione della **Seconda Fase**, la **Modificazione della Sequenza Originaria**.

In questo documento, la **Sequenza Originaria** oggetto di studio approfondito è la *sequenza di RNA* “GTGAAATGGTCATGTGTGGCGG” del Gene *RdRp* (*RdRp_SARSr-F*).

Dopo aver eseguito l'**analisi grafica** (*Prima Fase*) della **sequenza originaria** di DNA o di RNA, il software della **T.T.E.S.** offre la possibilità di modificare la **sequenza originaria** e di generare numerose e diverse **nuove sequenze** di DNA o di RNA che rispettano fedelmente i numerosi e diversi “*trend non manifesti*” della sequenza originaria.

I “**trend non manifesti**” sono *andamenti*, non chiaramente identificabili, presenti in tutte le sequenze di DNA e di RNA.

Generare numerose e diverse **nuove sequenze** di DNA o di RNA è possibile per **due motivi**:

- 1) perché ogni specifica sequenza di DNA o di RNA (*sequenza originaria*) può essere «trasformata» in numerose e diverse *nuove sequenze*, rispettando i diversi e numerosi “*trend non manifesti*” della *sequenza originaria* (si veda la Fig. 7.1 nella pagina successiva);
- 2) perché ogni specifico “*trend non manifesto*” della *sequenza originaria* può generare numerose e diverse nuove sequenze (si veda la Fig. 7.1 nella pagina successiva).

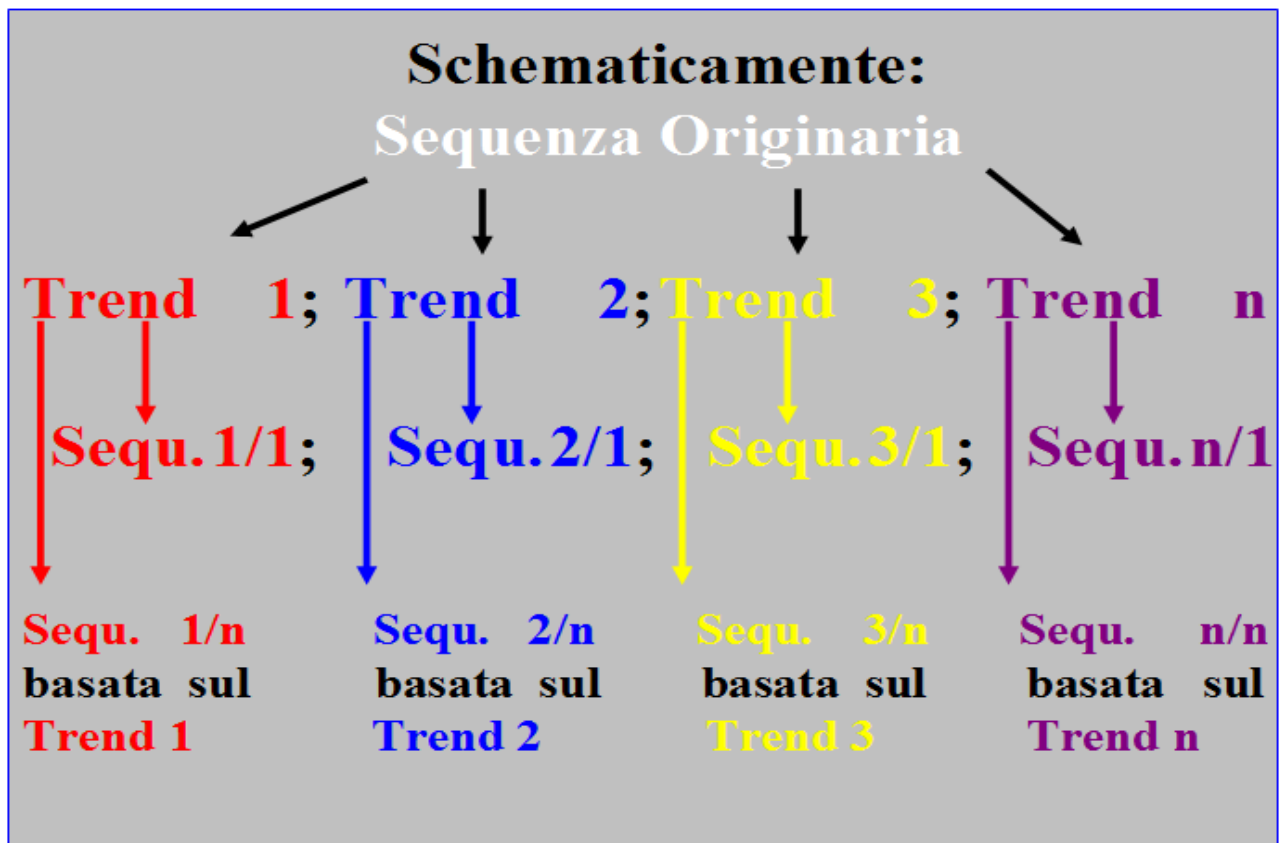


Fig. 7.1

A partire da **ognuno** dei **possibili “Trend Non Manifesti”** della **sequenza originaria** è possibile generare un numero indeterminato di **“nuove sequenze”** (come schematizzato in Fig. 7.1). Così, dal **Trend n°1** è possibile generare le sequenze n°1/1, 1/2, 1/3, ecc. Allo stesso modo dal **Trend n°2** è possibile generare le sequenze n°2/1, 2/2, 2/3, ecc., e così di seguito.

L’identificazione dei **“trend non manifesti”** costituisce il presupposto fondamentale per le modificazioni delle sequenze di DNA e RNA eseguite con il software della **T.T.E.S.**

Per la **MODIFICAZIONE COMPLETA DELLA SEQUENZA ORIGINARIA** e la generazione di **TUTTE LE NUOVE SEQUENZE** è necessario **identificare tutti i possibili TREND NON MANIFESTI** della **sequenza originaria**.

E’ fondamentale sottolineare che, in genere, la **sequenza originaria** differisce da **tutte le nuove sequenze generate** per almeno il **70% di basi** (si tratta quindi di **materiale biologico molto differente**).

7.2 I VENTIQUEATTRO “TREND NON MANIFESTI”

I principali “TREND NON MANIFESTI” sono VENTIQUEATTRO. Essi sono stati identificati attribuendo un numero compreso da 0 a 3 alle quattro basi azotate (T, C, G e A) e calcolando il numero delle Permutazioni senza ripetizione di n (4) elementi secondo la formula $P(k) = n!$ [$P(4) = 4!$].

Qui di seguito sono schematizzati i 24 “Trend Non Manifesti”.

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)

ATTENZIONE:

Come mostrato sopra e come si vedrà nel capitolo successivo, alcuni “Trend Non Manifesti”, anche se corrispondenti a permutazioni differenti, durante le modificazioni della sequenza originale, genereranno le stesse nuove sequenze.

7.3 PROCEDURA PER LA MODIFICAZIONE PARZIALE DELLA SEQUENZA ORIGINARIA

Nel paragrafo 7.1 è stato detto che per la **MODIFICAZIONE COMPLETA DELLA SEQUENZA ORIGINARIA** e la generazione di **TUTTE LE NUOVE SEQUENZE** è necessario **identificare tutti i possibili TREND NON MANIFESTI** della **sequenza originaria**.

In questo esempio considereremo **solo** uno tra i possibili **Trend Non Manifesti**. Per questo motivo, la modificazione della sequenza originaria è definita **“parziale”**.

Supponiamo di voler modificare una **sequenza di 22 basi**. Definiamo tale sequenza **“Sequenza Originaria X”**.

Sequenza Originaria X T A T C C C A T T A C A T G A T A T T A T T

Supponiamo di voler modificare la **Sequenza Originaria X** seguendo il **Trend n° 3**, uno dei **24 Trend Non Manifesti** identificati nel paragrafo precedente.

Modificazione della “Sequenza Originaria X” seguendo il Trend n° 3 (T=2; C=0; G=3; A=1)

Sequenza Originaria 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

Come è mostrato nella tabella sopra riportata, la **prima procedura** da eseguire è quella di **attribuire ad ogni lettera della sequenza il “numero specifico”** che le è stato assegnato secondo quanto stabilito dal *Trend Non Manifesto* selezionato.

Sequenza Originaria 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

La **seconda procedura** da eseguire è quella di **attribuire ad ogni “numero specifico” della sequenza una delle seguenti quattro lettere: “I”, “i”, “D” o “d”**.

Le *quattro lettere* sono stabilite confrontando un **“numero specifico”** con il **“numero specifico” che lo precede nella sequenza** e hanno il seguente significato:

- I = Increase** (Incremento rispetto al precedente **numero specifico** assegnato)
- i = repeated increase** (incremento ripetuto rispetto al precedente **numero specifico** assegnato)
- D = Decrease** (Decremento rispetto al precedente **numero specifico** assegnato)
- d = repeated decrease** (decremento ripetuto rispetto al precedente **numero specifico** assegnato)

Le definizioni **repeated increase (“i”)** e **repeated decrease (“d”)** sono utilizzate quando, **dopo un Incremento o dopo un Decremento**, il **“numero specifico” è ripetuto, è uguale**

a quello che lo precede (come mostrato nella tabella precedente e in quelle mostrate nelle pagine successive).

Queste quattro lettere (“**T**”, “**i**”, “**D**” o “**d**”) rendono **manifesto** il Trend Non Manifesto associato ai quattro “**numeri specifici**” che sono stati assegnati secondo quanto stabilito dal *Trend Non Manifesto* selezionato.

Sequenza Originaria 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

La **terza procedura** da eseguire è quella di creare una **nuova sequenza di “numeri specifici”** diversa da quella precedente, ma assolutamente rispettosa del **Trend n° 3**.

Per ottenere questo risultato è sufficiente **sostituire** ai “**numeri specifici**” del **Trend n° 3** (quelli nella seconda riga nella tabella mostrata qui di sopra) dei “**numeri specifici**” **diversi** che rispettino il **Trend n° 3 (I i D d)**, ovvero numeri il cui **Incremento** o **Decremento** si attenga rigorosamente a quanto stabilito dalla sequenza di “**T**”, “**i**”, “**D**” e “**d**” creata nel corso della **seconda procedura** (si veda la tabella mostrata qui di sopra e quella successiva).

Sequenza Originaria 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
Sequenza 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

La **quarta procedura** da eseguire consiste nella generazione della **nuova sequenza** (sequenza di lettere colorate in rosso nella tabella mostrata qui di sopra), la **Sequenza n° 3/1**.

Per raggiungere questo risultato è sufficiente **sostituire** alla sequenza dei “**numeri specifici**” **diversi** (quelli della quarta riga nella tabella mostrata qui di sopra e corrispondenti al **Trend n° 3/1**) le corrispondenti lettere attribuite secondo quanto stabilito dal *Trend Non Manifesto* selezionato (il **Trend n° 3**), ovvero **T=2**, **C=0**, **G=3** e **A=1**.

Riepilogando:

Sequenza Originaria X di 22 basi:

Sequenza Originaria X	T	A	T	C	C	C	A	T	T	A	C	A	T	G	A	T	A	T	T	A	T	T
-----------------------	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---

Sequenza n° 3/1 di 22 basi

generata dalla modificazione della **Sequenza Originaria X**:

Sequenza 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
--------------------------	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---

Riepilogando:

Sequenza Originaria Y di 22 basi:

Sequenza Originaria Y G T G A A A T G G T C A T G T G T G G C G G

Sequenza n° 3/1 di 22 basi generata dalla modificazione della **Sequenza Originaria Y:**

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

Anche in questo caso, la **Sequenza n° 3/1**, la **nuova sequenza generata** seguendo il **Trend n° 3/1** (Trend che è stato generato seguendo fedelmente il **Trend n° 3**), ha l'**81,82%** di basi (lettere) diverse dalla **Sequenza Originaria Y**, ma la **sequenza originaria** e la **nuova sequenza**, nonostante le differenze delle basi, hanno “caratteristiche” comuni proprio perché i loro **TREND** sono **STRUTTURALMENTE IDENTICI**.

Riprendiamo adesso l'ultima tabella mostrata nella pagina precedente.

Sequenza Originaria 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

Desidero richiamare l'attenzione sulle **lettere e i numeri UGUALI evidenziati in grigio** in entrambe le sequenze.

Essi sono **evidenziati in grigio** per indicare che, in quelle **specifiche posizioni**, le lettere della **Sequenza Originaria** hanno un **Trend** che NON PUO' ESSERE MODIFICATO.

Come possiamo osservare, alle lettere “**CATG**” corrispondono i numeri “**0123**”.

Nell'esempio riportato, alle lettere della **Sequenza Originaria** corrispondono **numeri progressivamente crescenti** (o potrebbero pure essere *crescenti ripetuti*, come “0112223”, a cui corrisponderebbero nell'esempio riportato le lettere ipotetiche “CAATTTG”) di un'unità, a partenza dal **numero minore** (cioè, dallo **0**) fino al **numero maggiore** (cioè, al **3**).

In queste **specifiche condizioni NON SONO MAI CONSENTITE MODIFICAZIONI DELLE LETTERE NELLE NUOVE SEQUENZE RISPETTO ALLE SEQUENZE ORIGINARIE**.

La stessa cosa dicasi se alle lettere della **Sequenza Originaria** corrispondessero **numeri progressivamente decrescenti** (cioè “3201”, o decrescenti ripetuti, come “3221110”, a cui corrisponderebbero nell'esempio riportato le lettere ipotetiche “GTAAAC”) di un'unità, a partenza dal **numero maggiore** (cioè, dal **3**) fino al **numero minore** (cioè, allo **0**).

E' verosimile pensare che **tali particolari raggruppamenti di lettere** assumano un significato biologico importante come segnale o forse anche perché potrebbero essere particolarmente resistenti a modificazioni accidentali durante i processi di **trascrizione**.

La Sequenza Originaria Y,

Sequenza Originaria Y

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

corrisponde alla ormai nota *sequenza di RNA del Gene RdRP (RdRp_SARSr-F)* riscontrata nel SARS-CoV-2 (si veda la sequenza delimitata dalle linee laterali nere e superiori rosse nel grafico mostrato qui di sotto in Fig. 7.2).

Le frecce verdi in Fig. 7.2 indicano rispettivamente gli aminoacidi **Metionina** (Methionine, Met/M, codone **ATG**) e **Valina** (Valine, Val/V, codone **GTC**), mentre la freccia blu indica l'aminoacido **Metionina** (codone **ATG**).

Also in Fig. 7.2, the letters delimited by the blue frame (those 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.

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

GenBank: MW411920.1

[GenBank](#) [FASTA](#)

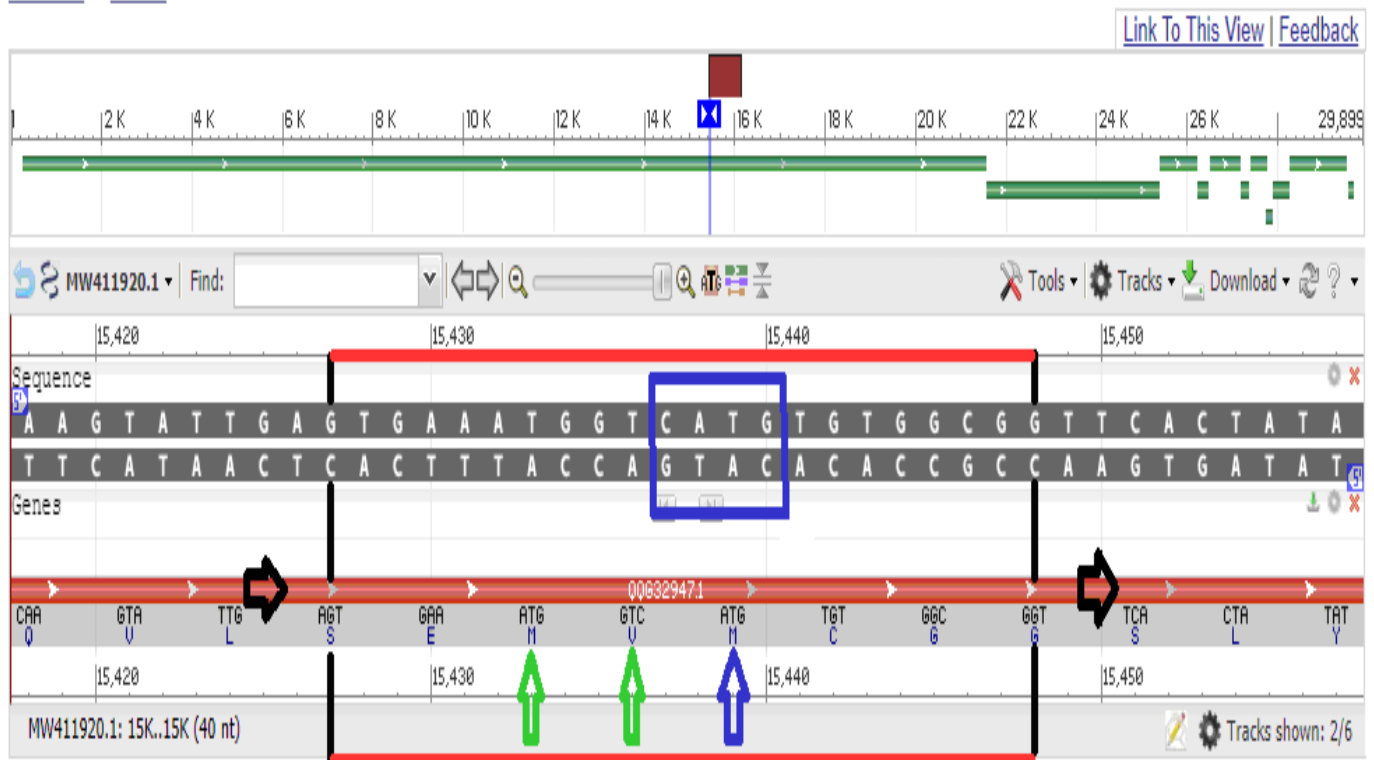


Fig. 7.2

La **Sequenza n° 3/1** dell'ultimo esempio riportato (definita invece **Sequenza Originaria X** nell'esempio mostrato all'inizio di questo paragrafo),

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**

come vedremo nel capitolo successivo, corrisponde realmente alla **Sequenza n° 3/1** (si veda la sequenza delimitata dalle linee laterali nere e superiori rosse nel grafico mostrato qui di sotto in Fig. 7.3), una **nuova sequenza generata** (seguendo il **Trend n° 3/1**) a partire dalla *sequenza di RNA del Gene RdRP (RdRp_SARSr-F)*, la **Sequenza Originaria Y** di cui sopra.

Come si puo' osservare in Fig. 7.3, tra gli **“organismi”** identificati dagli allineamenti significativi (della ricerca BLAST) di questa sequenza figura sorprendentemente il ***Bacillus Anthracis*** (si ricordi quanto detto nella Premessa e si veda quanto scritto nei Capitoli X e XI, nella Discussione Generale e nelle Conclusioni).

E' interessante notare che, allo stesso modo di come mostrato in Fig. 7.2, le frecce verdi mostrate in Fig. 7.3, indicano rispettivamente gli aminoacidi **Metionina** (Methionine, Met/M, codone **ATG**) e **Valina** (Valine, Val/V, codone **GTA**), mentre la freccia blu indica un aminoacido diverso, l'**Istidina** (Histidine, His/H, codone **CAT**).

Per inciso ricordo che, come spiegato nel paragrafo 2.1 del Capitolo 2 (e al quale rimando per approfondimenti), per decidere il senso di lettura del codone bisogna riferirsi al senso di lettura stabilito dalla direzione indicata dalle **Frecce nere** mostrate nelle Fig. 7.2 e 7.3.

Così, mentre in Fig. 7.2 i **codoni** si leggeranno in senso 5' → 3' (da sinistra verso destra) in **“sequence”** (sequenza in alto), nella Fig.7.3 i **codoni** si leggeranno in senso 5' → 3' (da destra verso sinistra) in **“genes”** (sequenza in basso).

Bacillus anthracis strain FDAARGOS_702 chromosome

GenBank: CP054800.1

[GenBank](#) [FASTA](#)

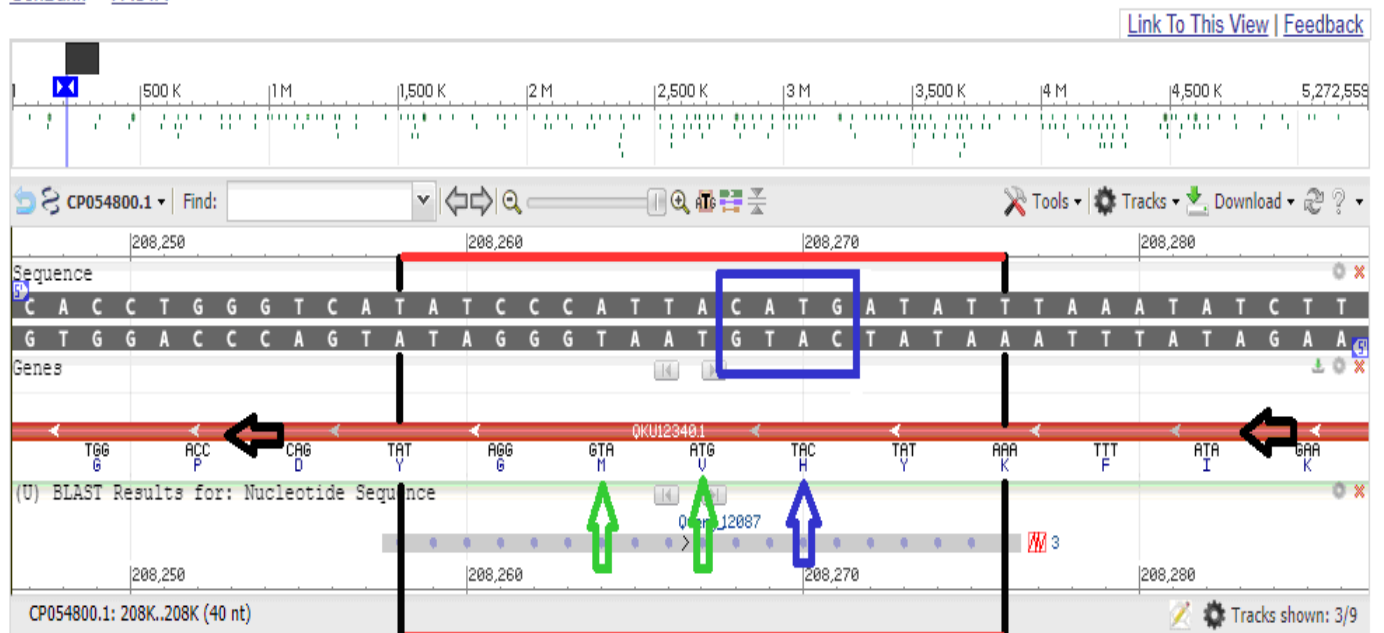


Fig. 7.3

Sempre in Fig. 7.3, le lettere delimitate dalla cornice blu (quelle poste in alto) si riferiscono al **particolare raggruppamento di lettere** (“CATG”) già ampiamente discusso nelle pagine precedenti.

Tre di queste lettere (“ATG”) corrispondono alle tre lettere speculari “TAC” che, lette da destra verso sinistra, codificano l’aminoacido **H** (diversamente da come mostrato nella Fig. 7.2 in cui l’aminoacido in questione è la **M**), mentre la lettera “C” corrisponde alla **prima** lettera speculare (la lettera “G”) dell’aminoacido **V** (“GTA”) che **segue** l’aminoacido **H** indicato dalla freccia blu.

Prima di concludere il capitolo, riprendiamo di nuovo l’ultima tabella mostrata a pagina 142, quella della **Modificazione della “Sequenza Originaria Y” seguendo il Trend n° 3** (T=2; C=0; G=3; A=1).

Sequenza Originaria Y	G	T	G	A	A	A	T	G	G	T	C	A	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

Come è stato accennato nel paragrafo 1.1 del Capitolo I, in **bioinformatica** l’informazione genetica è trattata in **maniera binaria**, cioè come **una serie di 0 e 1**, e trasformata in **algoritmi**.

Se volessimo applicare il **codice binario** ed utilizzare **una serie di 0 e 1** a posto dei numeri **0, 1, 2 e 3** nella **prima** e nella **terza procedura** di **modificazione di una sequenza originaria** (come è stata esposta nelle pagine precedenti), è sufficiente assegnare il *codice binario* “00” allo **0**, “01” al numero **1**, “10” al numero **2** e, infine, il codice binario “11” al numero **3**.

A seguito dell’applicazione del **codice binario** il **Trend n° 3** diventerebbe (T=10; C=00; G=11; A=01).

Sequenza Originaria 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
Sequenza n° 3/1 (81,82%)	T	A	T	C	C	C	A	T	T	A	C	A	T	G

Con questa nuova codificazione gli *aminoacidi* saranno definiti da **sei numeri**.

Così, nella **Sequenza n° 3/1**, al codone “ATG” (evidenziato in grigio scuro nella tabella mostrata qui di sopra), specularmente letto “TAC” (in fig. 7.3, lettere delimitate dalla cornice blu, quelle poste in basso) e codificante “CAT” (codone letto da destra verso sinistra in fig. 7.3), corrisponderanno i numeri “000110” (infatti, per il **Trend n° 3**, C=00, A=01 e T=10) associati all’aminoacido **H** (indicato dalla freccia blu in fig. 7.3 nella pagina precedente).

ATTENZIONE: se, diversamente da come visto sopra, avessimo preso in considerazione la stessa **sequenza originaria** (la **Sequenza Originaria Y**), ma utilizzato per la modificazione il **Trend n° 1** (T=00; C=01; G=10; A=11) invece del Trend n° 3 (T=10; C=00; G=11; A=01), ad un qualsiasi codone “**CAT**” sarebbero corrisposti 6 numeri diversi, ovvero “**011100**” piuttosto che i numeri “000110”.

I 6 numeri “**011100**” sarebbero sempre stati associati all’aminoacido **H**, ma avrebbero avuto un significato diverso, perché associati ad una **sequenza** (la **sequenza n° 1/1**) che avrebbe avuto “caratteristiche” **comuni** alla **sequenza originaria** (perché il **Trend n°1** della **sequenza originaria** e il **Trend n° 1/1** della **sequenza n° 1/1** sono STRUTTURALMENTE IDENTICI) e anche alla **nuova sequenza generata** seguendo il **Trend n° 3/1** (semplicemente per il fatto che anche essa è stata generata a partire dalla **sequenza originaria**), ma anche “caratteristiche” **non comuni** a quest’ultima sequenza, perché essa sarebbe stata generata seguendo il **Trend n° 3/1**, piuttosto che il **Trend n° 1/1**.

Da questo semplice esempio, si capisce come, con una codifica come quella proposta, la **complessità del codice genetico** aumenti notevolmente, *evidenziando somiglianze e differenze attualmente ignote tra le sequenze*, e anche come crescerebbero il numero di informazioni che si potrebbero ottenere se avessimo una *banca dati* che prendesse in considerazione i **Trend** e tutto ciò che è ad essi correlato.

CAPITOLO VIII

**MODIFICAZIONI
DELLA SEQUENZA DI RNA
“GTGAAATGGTCATGTGTGGCGG”
del Gene RdRp (RdRp_SARSr-F)
E GENERAZIONI DI
NUOVE SEQUENZE**

OBIETTIVI DEL CAPITOLO VIII

Questo capitolo è dedicato alle **modificazioni** della sequenza di RNA “GTGAAATGGTCATGTGTGGCGG” del *Gene RdRp (RdRp_SARSr-F)*, una delle sequenze di RNA utilizzate come controllo positivo del tampone rinofaringeo, insieme ad altre sequenze dei geni del *SARS-CoV-2*.

L’obiettivo delle modificazioni di questa sequenza di RNA (definita **sequenza originaria**) è quello di generare un certo numero di **nuove sequenze** che abbiano almeno il **70%** di basi diverse dalla **sequenza originaria**.

Per ottenere questo risultato si applica la procedura spiegata nel capitolo precedente basata sull’identificazione dei diversi TREND NON MANIFESTI della **sequenza originaria**.

Nonostante la notevole quantità di basi diverse, le **nuove sequenze hanno** “**caratteristiche**” **funzionali comuni alla sequenza originaria**.

Questa evidenza può essere opportunamente sfruttata per ottenere risultati biologicamente rilevanti da parte dell’industrie farmaceutiche, biochimiche, biotecnologiche, ecc.

Nello specifico, le **modificazioni della sequenza di RNA** “GTGAAATGGTCATGTGTGGCGG” del *Gene RdRp (RdRp_SARSr-F)* e la **metodologia proposta in questo documento aprono prospettive diverse per avviare studi finalizzati e trattamenti innovativi della pandemia da Covid 19**.

8.1 MODIFICAZIONI della SEQUENZA di RNA “GTGAAATGGTCATGTGTGGCGG” del Gene RdRp (RdRp_SARSr-F) e generazioni di NUOVE SEQUENZE

La sequenza “GTGAAATGGTCATGTGTGGCGG” [del Gene RdRp (*RNA-dependent RNA polymerase*, RdRp_SARSr-F)] è parte dell’ “**ORF1ab**” (lettura aperta frame 1ab di SARS-CoV-2), una delle sequenze di RNA utilizzate come controllo positivo del tampone rinofaringeo, insieme ad altre sequenze dei geni del SARS-CoV-2.

Qui di seguito sono presentate le tabelle con SOLTANTO ALCUNE delle numerose modificazioni di questa importante sequenza (si veda prima il paragrafo 7.3 del capitolo 7).

ATTENZIONE:

Le **coppie di Trend** la cui applicazione genera sequenze uguali (si veda il paragrafo 7.2 del capitolo 7) sono evidenziate con colori simili.

Sequenza Originaria = GTGAAATGGTCATGTGTGGCGG

Modificazioni seguendo il Trend n° 1 (T=0; C=1; G=2; A=3)

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° 1/1	1	0	1	2	2	2	1	3	3	1	2	3	1	3	1	3	1	3	3	2	3	3
Sequenza 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

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° 1/2	1	0	1	2	2	2	1	3	3	0	2	3	1	3	1	3	1	3	3	0	1	1
Sequenza 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

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

Modificazioni seguendo il Trend n° 2 (T=3; C=2; G=1; A=0)

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° 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
Sequenza 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

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° 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
Sequenza 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

Modificazioni seguendo il Trend n° 3 (T=2; C=0; G=3; A=1)

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° 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

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° 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
Sequenza 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

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° 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
Sequenza 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

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° 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
Sequenza 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

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° 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
Sequenza 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

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° 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
Sequenza 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

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

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° 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
Sequenza 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

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

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° 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
Sequenza 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

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° 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
Sequenza 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

Modificazioni seguendo il Trend n° 6 (T=2; C=3; G=0; A=1)

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° 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
Sequenza 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

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° 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
Sequenza 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

Modificazioni seguendo il Trend n° 7 (T=2; C=1; G=0; A=3)

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° 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
Sequenza 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

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° 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
Sequenza 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

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° 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
Sequenza 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

Modificazione seguendo il Trend n° 8 (T=1; C=0; G=3; A=2)

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° 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
Sequenza 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

Modificazioni seguendo il Trend n° 9 (T=2; C=0; G=1; A=3)

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° 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
Sequenza 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

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° 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
Sequenza 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

Modificazione seguendo il Trend n° 10 (T=2; C=1; G=3; A=0)

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° 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
Sequenza 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

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

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° 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
Sequenza 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

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

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° 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
Sequenza 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

Modificazione seguendo il Trend n° 14 (T=1; C=3; G=2; A=0)

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° 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
Sequenza 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

Modificazione seguendo il Trend n° 15 (T=3; C=2; G=0; A=1)

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° 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
Sequenza 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

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

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° 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
Sequenza 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

Modificazioni seguendo il Trend n° 18 (T=0; C=1; G=3; A=2)

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° 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
Sequenza 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

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° 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
Sequenza 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

Modificazione seguendo il Trend n° 19 (T=1; C=2; G=3; A=0)

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° 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
Sequenza 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

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

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° 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
Sequenza 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

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° 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
Sequenza 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

CAPITOLO IX

**RICERCHE BLAST
E DISCUSSIONE DEI
RISULTATI OTTENUTI
DAGLI ALLINEAMENTI
DELLA SEQUENZA N° 1/3**

OBIETTIVI DEL CAPITOLO IX

In questo capitolo sono riportati i risultati degli *allineamenti significativi* della ricerca BLAST effettuata sulla **Sequenza n° 1/3**. Questa sequenza è una delle tante generate dalla **Sequenza Originaria** [la sequenza “GTGAAATGGTCATGTGTGGCGG” del Gene RdRp (*RNA-dependent RNA polymerase*, RdRp_SARSr-F)] e dalla quale differisce per l’**86,36 %** delle basi (si veda il Capitolo precedente per la *procedura di modificazione*).

I risultati degli allineamenti significativi della ricerca BLAST (**Hitlist size 100**) hanno riguardato diversi “**organismi**” comuni (distinguibili con diversi colori nelle pagine successive) agli “**organismi**” identificati dalle ricerche BLAST effettuate sulle sequenze a cui è stata **sostituita una base alla sequenza originaria** (si vedano i Capitoli V e VI).

Estendendo la ricerca BLAST (**Hitlist size 1000**) sono stati individuati moltissimi “**organismi**” comuni, due sequenze del *Pipistrellus pipistrellus* e una sequenza del *Virus dell’HIV-1*.

Quest’ultima sequenza non era sicuramente tra quelle attese, non avendo apparentemente nulla in “comune” con la **sequenza originaria**, se non il fatto che entrambe le sequenze si riferiscono a dei **virus**, ma di natura diversa.

L’**infezione da HIV** è causata da uno di due *retrovirus* simili (*HIV-1* e *HIV-2*) che distruggono i *linfociti T CD4+* e *inibiscono la risposta immunitaria cellulo-mediata*, con aumento del rischio di alcune infezioni e di alcuni tumori. L’infezione primaria può produrre una *sindrome febbrile aspecifica*. Il rischio di successive manifestazioni, correlate allo stato di immunodeficienza, è proporzionale al livello di deplezione del linfocita CD4+ circolante (<https://www.msmanuals.com/it-it/professionale/malattie-infettive/virus-dell-immunodeficienza-umana-hiv/infezione-da-virus-dell-immunodeficienza-umana-hiv>).

Per la peculiarità del *Virus dell’HIV-1*, il suo allineamento significativo è stato oggetto di approfondimento ed è stata proposta anche una IPOTETICA **predizione dei codoni** (*triplette di basi che formano gli aminoacidi*) costitutivi di un **segmento della Sequenza n° 1/3** e di quelli **adiacenti ad esso** (qualora tali *codoni* diventassero codificanti per proteine).

Le ricerche BLAST effettuate su questi **codoni** hanno dato risultati molto interessanti.

Di rilevante importanza sembrano essere i ruoli svolti dalla “**zinc finger protein**” e dalle famiglie di **idrolasi** (in particolare, la **idrolasi serina**).

Come si potrà apprezzare nel paragrafo x, i risultati della ricerca BLAST sulla “**Sequenza n° 1/3 (HIV-1) – Modificata**” hanno identificato un numero notevole di “organismi” e, tra questi, **tre specie di pipistrelli** [*Rhinolophus ferrumequinum*) e *Pipistrellus (pipistrellus e kuhlii)*], **due specie di pangolini** [*Manis javanica (pangolino malese)* e *Manis pentadactyla (pangolino cinese)*] e diversi “organismi” implicati in sintomatologie prevalentemente **infettive, febbrili e polmonari** (in particolare, *Klebsiella sp.*, *Legionella pneumophila* e *Simian immunodeficiency virus*). Imprevedibili anche gli allineamenti con alcune sequenze che si riferiscono a **olfactory receptor 6Y1-like**, in *Manis pentadactyla* e in *Manis javanica*, *un tipo di recettore olfattivo comune anche agli esseri umani*. Un allineamento significativo dei codoni si riferisce alla **neuraminidasi virale** del *Virus dell’Influenza A*. Infine, molto importanti sono l’identificazione di una sequenza che si riferisce al **dominio N-terminale della RNA polimerasi diretta dal DNA** e quella che riguarda la **glicosiltransferasi famiglia 2** del *Bacillus anthracis*. Questo **batterio**, che sarà identificato anche da altre sequenze ottenute dalle modificazioni della **sequenza originaria**, per la sua **letalità** e per il suo massiccio coinvolgimento nel **danno polmonare**, sarà oggetto di ampia trattazione nei capitoli successivi.

SEQUENZA N° 1/3

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

9.1 SEQUENZA 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	blunt-snouted	36.2	66.4	81%	8.9	100.00%	40234402

	assembly, chromosome: 16	clingfish								
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	Claviceps gigantea	34.2	34.2	77%	35	100.00%		40254	

	Lps1, and EasH genes, complete cds									
LT606948.1	Polynucleobacter necessarius isolate PPGSP3 genome assembly, chromosome: I	Polynucleobacter necessarius	34.2	34.2	77%	35	100.00%	1655757		
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		
	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		
XM_018231673.1	PREDICTED: Monodelphis domestica uncharacterized LOC103103228 (LOC103103228), ncRNA	gray short-tailed opossum	34.2	34.2	77%	35	100.00%	760		
XR_461531.2	PREDICTED: Calidris pugnax FK506 binding protein 14 (FKBP14), mRNA	ruff	34.2	34.2	77%	35	100.00%	1807		
XM_014937915.1	Leptosphaeria maculans lepidii ibcn84_scaffold00006 complete sequence	Leptosphaeria maculans 'lepidii' group	34.2	34.2	77%	35	100.00%	1544579		
FO906018.1	Cyprinus carpio genome assembly common carp genome, scaffold 000000475	common carp	34.2	68.4	95%	35	95.24%	1926052		
LN598375.2	Cyprinus carpio genome assembly common carp genome, scaffold: LG1, chromosome: 1	common carp	34.2	64.4	100%	35	95.24%	15297981		
LN590701.1	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		
XM_037263922.1	Homo sapiens DNA, chromosome 18, nearly complete genome	human	34.2	34.2	77%	35	100.00%	77846715		
AP023478.1	PREDICTED: Amphiprion ocellaris transmembrane protein 182-like (LOC111569869), mRNA	clown anemonefish	34.2	34.2	77%	35	100.00%	2086		
XM_023272289.2	Onychomys torridus genome assembly, chromosome: 13	southern grasshopper mouse	34.2	159	77%	35	100.00%	78371027		
LR877200.1	Arvicola amphibius genome assembly, chromosome: 13	Eurasian water vole	34.2	94.6	86%	35	100.00%	75712621		
LR862393.1	PREDICTED: Aphantopus hyperantus pre-mRNA-splicing factor Slu7 (LOC117986526), mRNA	ringlet	34.2	34.2	77%	35	100.00%	1744		
XM_034973366.1	Herbaspirillum rubrisubalbicans Os34, complete genome	Herbaspirillum rubrisubalbicans Os34	34.2	34.2	77%	35	100.00%	6122717		
CP008956.1	Canis lupus familiaris breed Labrador retriever chromosome 27a	dog	34.2	34.2	77%	35	100.00%	46050380		
CP050582.1	Canis lupus familiaris breed Labrador retriever chromosome 27b	dog	34.2	34.2	77%	35	100.00%	46054595		
CP050640.1	Plectropomus leopardus DNA, chromosome 8, nearly complete sequence	leopard coralgroup	34.2	34.2	77%	35	100.00%	35874999		
AP022707.1	Aphantopus hyperantus genome assembly, chromosome: 11	ringlet	34.2	34.2	77%	35	100.00%	15241335		
LR761658.1	PREDICTED: Drosophila virilis octopamine receptor Oamb (LOC6629914), transcript variant X2, mRNA	Drosophila virilis	34.2	34.2	77%	35	100.00%	4221		
XM_032439116.1	PREDICTED: Drosophila virilis octopamine receptor Oamb	Drosophila virilis	34.2	34.2	77%	35	100.00%	5158		
XM_032439114.1										

	(LOC6629914), transcript variant X1, mRNA									
XM_031711235.1	PREDICTED: Actinia tenebrosa serine/threonine-protein kinase Haspin homolog ALK1-like (LOC116302041), mRNA	Australian red waratah sea anemone	34.2	34.2	77%	35	100.00%	4119		
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	greater pipefish	34.2	34.2	77%	35	100.00%	2126		

	SLU7 homolog, splicing factor (slu7), mRNA								
	PREDICTED: Morone saxatilis SLU7 homolog, splicing factor (slu7), mRNA	striped sea-bass	34.2	34.2	77%	35	100.00%	2257	
XM_035668914.1	Ananas comosus genome assembly, chromosome: 23	pineapple	34.2	34.2	77%	35	100.00%	7979763	
LR828303.1	Danio kyathit genome assembly, chromosome: 13	Danio kyathit	34.2	64.4	86%	35	100.00%	63723151	
LR812531.1	Danio aesculapii genome assembly, chromosome: 18	Danio aesculapii	34.2	94.6	86%	35	100.00%	56436977	
LR812514.1									

VISTA DI ALLINEAMENTO VERTICALE AGGANCIATO A QUERY CON LETTERE D'IDENTITA'

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

<u>LR862393.1</u>	51550451	GCAATGACACACAAGAA	51550435
<u>LR862393.1</u>	10046671	GGGCAATGACACACA	10046657
<u>LR862393.1</u>	52001369	AATGACACACAAGAA	52001383
<u>XM 034973366.1</u>	398	GCAATGACACACAAGAA	414
<u>CP008956.1</u>	2013233	GCAATGACACACAAGAA	2013249
<u>CP050582.1</u>	4860768	CTCGGGCAATGACACAC	4860784
<u>CP050640.1</u>	4858108	CTCGGGCAATGACACAC	4858124
<u>AP022707.1</u>	6289195	GCAATGACACACAAGAA	6289211
<u>LR761658.1</u>	1638750	GCAATGACACACAAGAA	1638766
<u>XM 032439116.1</u>	354	CTCGGGCAATGACACAC	370
<u>XM 032439114.1</u>	333	CTCGGGCAATGACACAC	349
<u>XM 031711235.1</u>	1995	GCAATGACACACAAGAA	2011
<u>LR606204.1</u>	18859071	GGGCAATGACACACAAG	18859087
<u>XM 028460065.1</u>	501	GCAATGACACACAAGAA	517
<u>XR 003423601.1</u>	609	GCAATGACACACAAGAA	625
<u>XM 026144372.1</u>	1675	CGGGCAATGACACACAA	1691
<u>CP032600.1</u>	8126049	GGGCAATGACACACAAG	8126065
<u>XM 025865289.1</u>	8	CGGGCAATGACACACAA	24
<u>XM 024828719.1</u>	650	GGCAATGACACACAAGA	666
<u>CP027092.1</u>	45563484	CGGGCAATGACACACAA	45563500
<u>XR 002737306.1</u>	610	GCAATGACACACAAGAA	626
<u>CP022464.2</u>	5016426	CGGGCAATGACACACAA	5016442
<u>XM 016051987.2</u>	477	GCAATGACACACAAGAA	493
<u>XM 016051986.2</u>	477	GCAATGACACACAAGAA	493
<u>XM 016051985.2</u>	538	GCAATGACACACAAGAA	554
<u>XM 008668489.2</u>	735	GGCAATGACACACAAGA	751
<u>XM 019870024.1</u>	502	GCAATGACACACAAGAA	518
<u>XM 016622770.1</u>	73	GGCAATGACACACAAGA	89
<u>XM 016051988.1</u>	538	GCAATGACACACAAGAA	554
<u>XM 014149233.1</u>	1259	GGGCAATGACACACAAG	1275
<u>XM 014149231.1</u>	1260	GGGCAATGACACACAAG	1276
<u>LN595020.1</u>	5226	CTCGGGCAATGACGCACAAGA	5246
<u>LN591285.1</u>	601630	CTCGGGCAATGACGCACAAGA	601650
<u>LN591285.1</u>	602415	CTCGGGCAATGACGCACAAGA	602435
<u>LN590715.1</u>	3440524	CTCGGGCAATGACGCACAAGA	3440544
<u>LN590715.1</u>	3447044	CTCGGGCAATGACGCACAAGA	3447064
<u>XM 037261829.1</u>	486	GCAATGACACACAAGAA	502
<u>XM 035668914.1</u>	519	GCAATGACACACAAGAA	535
<u>LR828303.1</u>	653605	CGGGCAATGACACACAA	653621
<u>LR812531.1</u>	54775021	GCAATGACACACAAGAA	54775037
<u>LR812531.1</u>	33111773	GGGCAATGACACACA	33111787
<u>LR812514.1</u>	54032613	GGCAATGACACACAAGA	54032629
<u>LR812514.1</u>	50154682	GGGCAATGACACACA	50154668
<u>LR812514.1</u>	34011470	AATGACACACAAGAA	34011484

9.2 SEQUENZA 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 assembly, chromosome: 22 Coregonus sp. 'balchen' 32.2 64.4 77% 144 100.00% 56862223
..... LR778272.1 Coregonus sp. 'balchen' genome Coregonus sp. 'balchen' 32.2 62.4 77% 144 100.00% 54335267

	assembly, chromosome: 20									
LR778266.1	Coregonus sp. 'balchen' genome assembly, chromosome: 14	Coregonus sp. 'balchen'	32.2	32.2	72%	144	100.00%	47256133		
.....	
	PREDICTED: Sparus aurata uncharacterized LOC115587570 (LOC115587570), transcript variant X2, ncRNA									
XR_003985082.1		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		
.....	
	PREDICTED: Salmo trutta serine/arginine repetitive matrix protein 1-like (LOC115161037), transcript variant X2, mRNA									
XM_029711715.1		Salmo trutta	32.2	32.2	72%	144	100.00%	5051		
	PREDICTED: Salmo trutta serine/arginine repetitive matrix protein 1-like (LOC115161037), transcript variant X1, mRNA									
XM_029711714.1		Salmo trutta	32.2	32.2	72%	144	100.00%	5054		
.....	
	Myripristis murdjan genome assembly, chromosome: 22									
LR597571.1		Myripristis murdjan	32.2	32.2	72%	144	100.00%	31899175		
	Myripristis murdjan genome assembly, chromosome: 20									
LR597569.1		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		
.....	
	Oryza sativa Indica Group cultivar Shuhui498 chromosome 2 sequence									
CP018158.1		Oryza sativa Indica Group	32.2	32.2	72%	144	100.00%	37764328		
.....	
	Oryza sativa Japonica Group DNA, chromosome 2, cultivar: Nipponbare, complete sequence									
AP014958.1		Oryza sativa Japonica Group	32.2	32.2	72%	144	100.150%	35937250		
	Oryza sativa Indica Group cultivar RP Bio-226 chromosome 2 sequence									
CP012610.1		Oryza sativa Indica Group	32.2	32.2	72%	144	100.00%	36385228		
.....	
	Solanum pennellii chromosome ch04, complete genome									
HG975443.1		Solanum pennellii	32.2	32.2	72%	144	100.00%	77197300		
.....	
	Oryza sativa Indica Group cultivar Zhenshan 97 chromosome 2									
CP056053.1		Oryza sativa Indica Group	32.2	32.2	72%	144	100.00%	37267338		
	Oryza sativa Indica Group cultivar Minghui 63 chromosome 2									
CP054677.1		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 shepardii	32.2	32.2	72%	144	100.00%	2793		
.....	
LR877236.1	Acomys russatus	Acomys russatus	32.2	122	95%	144	100.00%	53031086		

LR877232.1	genome assembly, chromosome: 25 Acomys russatus genome assembly, chromosome: 21 Acomys russatus genome assembly, chromosome: 14	Acomys russatus	32.2	64.4	81%	144	100.00%	61132863
LR877225.1	Acomys russatus	32.2	92.7	90%	144	100.00%	62749011
LR862372.1	Pipistrellus pipistrellus genome assembly, chromosome: 16 Pipistrellus pipistrellus genome assembly, chromosome: 6	Pipistrellus pipistrellus	32.2	96.6	77%	144	100.00%	46231460
LR862362.1	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 genome assembly, chromosome: 14	Thalassophryne amazonica	32.2	124	95%	144	100.00%	70111961
LR722979.1	PREDICTED: Chanos chanos glycerophosphocholine phosphodiesterase 1 (gpcpd1), mRNA	Chanos chanos	32.2	32.2	72%	144	100.00%	3535
XM_030771950.1
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 genome assembly, chromosome: 4	Myripristis murdjan	32.2	32.2	72%	144	100.00%	42404719
LR597553.1	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

.....	PREDICTED: Branchiostoma belcheri pre-mRNA-splicing factor SLU7-like (LOC109486685), mRNA	Branchiostoma belcheri	32.2	32.2	72%	144	100.00%	1869
XM_019790575.1								
.....	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_008640479.2								
.....	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_017743113.1								
-	PREDICTED: Cebus imitator SLU7 homolog, splicing factor (SLU7), transcript variant X2, mRNA	Cebus imitator	32.2	32.2	72%	144	100.00%	2095
XM_017502118.1								
.....	Oryza sativa Japonica Group DNA, chromosome 6, cultivar: Nipponbare, complete sequence	Oryza sativa Japonica Group	32.2	32.2	72%	144	100.00%	31248787
AP014962.1								
.....	Oryza sativa Indica Group cultivar RP Bio-226 chromosome 6 sequence	Oryza sativa Indica Group	32.2	32.2	72%	144	100.00%	30903862
CP012614.1								
.....	Solanum pennellii chromosome ch10, complete genome	Solanum pennellii	32.2	62.4	81%	144	100.00%	82529941
HG975449.1								
.....	Oryza sativa Indica Group cultivar Zhenshan 97 chromosome 6	Oryza sativa Indica Group	32.2	32.2	72%	144	100.00%	32119910
CP056057.1								
.....	Oryza sativa Indica Group cultivar Minghui 63 chromosome 6	Oryza sativa Indica Group	32.2	32.2	72%	144	100.00%	31921180
CP054681.1								
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
.....	1_Tdi_b3v08	Timema douglasi	32.2	32.2	72%	144	100.00%	1642
OA603930.1								
.....	PREDICTED: Cebus imitator SLU7 homolog, splicing factor (SLU7), transcript variant X1, mRNA	Cebus imitator	32.2	32.2	72%	144	100.00%	2045
XM_017502117.2								
.....	Acomys russatus genome assembly, chromosome: 6	Acomys russatus	32.2	92.7	86%	144	100.00%	89442464
LR877217.1								
.....	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_035891949.1								
.....	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_035891948.1								
.....	PREDICTED: Egretta garzetta lysosomal associated membrane protein 2 (LAMP2), transcript variant X1, mRNA	Egretta garzetta	32.2	32.2	72%	144	100.00%	3718
XM_009636403.2								
.....	Oryza sativa Japonica Group genomic DNA, chromosome 6, BAC clone:OSJNBb0024N18	Oryza sativa Japonica Group	32.2	32.2	72%	144	100.00%	128739
AP004741.3								
.....	Oryza sativa Japonica Group genomic DNA, chromosome 6, PAC clone:P0537F07	Oryza sativa Japonica Group	32.2	32.2	72%	144	100.00%	147724
AP003525.2								

CP056062.1	Oryza sativa Indica Group cultivar Zhenshan 97 chromosome 11	Oryza sativa Indica Group	30.2	30.2	68%	570	100.00%	32624309
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

CONSRM 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**

//

ATTENZIONE:

La sequenza “**ttcttgtgtg tcattg**” mostrata sopra è la *sequenza contraria e speculare* della sequenza “CAATGACACACAAGAA” (sequenza parziale della **Sequenza 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

CONSRMT 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 ttgcc**

//

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 IL SEGMENTO DELLA SEQUENZA N° 1/3 IDENTIFICATO COME CORRISPONDENTE AL VIRUS DELL'HIV-1

Nelle pagine che seguono è presentato e discusso in maniera molto approfondita il risultato della ricerca BLAST relativo all'*allineamento significativo* che ha identificato una sequenza del *Virus dell'HIV-1*.

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="MDFVDPNLEPWNHPGSRPTTPCSKCYCKACCFHCLVCFQNKGLG
 ISYGRKRKRRRRRGTTPQSCEDHQNPVVKQPTPQFKGISTDPEESKKKVESKAETDRFD"

gene <1..2471
 /gene="rev"

CDS <2176..2471
 /gene="rev"
 /codon_start=1
 /product="rev protein"
 /protein id="ACR55222.1"
 /translation="MAGGSGDGEELLRAVRIKILYQSNPPSSRGSRQTRKNRRRR
 WRARQRQIDISERVLISACLGRPPEPVPLQLPPLERLHLDCSEDCGTSQTSQSGVET
 GVGRPQVSGESSDILGEGDKK"

**gene <1..83
 /gene="vpu"**

**CDS <1..83
 /gene="vpu"
 /codon_start=3
 /product="vpu protein"**

**/protein id="ACR55223.1"
 /translation="MSPLEICAIAGLVVALILAIVVWTIVGIEIKQVLKQRKIDRLIE
 RIRERAEDSGNESDGD AEDLSLLVEMGH DNLGYDNNL"**

gene 1..2613

/gene="env"
misc_feature 1..2613
/gene="env"
/note="nonfunctional envelope glycoprotein due to mutation"

ORIGIN

1 at **gagagtga tggggacgca gaggacttgt ccttgcttgt ggagatgggg catgataatc**
61 **ttgggtatga taataattg tag**tggtaca gaaacttgt gggttactgt ctactatggg
121 gtacctgtgt ggaaagatgc agacaccacc ttatthttgtg catcagatgc taaagcatat
181 gacacagaag tgcataatgt ctgggctaca catgcctgtg tgcccacaga ccccaaccca
241 caagagataa atttgagaaa tgtgacagaa gagtttaaca tgtggaaaaa taacatggta
301 gagcagatgc atacagatat aatcagtcta tgggaccaa gtctaaagcc atgtgtacag
361 ttaacccttc tctgcgttac gttaaattgt agagatgatg acgacatcaa tgtcgccaat
421 agcacagaca ccaactaacga agtcaaaaac ggatctgcca atatcacaga tgaaaaaac
481 tgctctttca ctatgaccac tgacctaata gataagctaa ttgaagtatt ttctgttttt
541 tatagacttg atgtttgaca tattgataaa aatagtaata acagtgggta tagattaata
601 aattgaaata ccacaaccgt tacacaggct ggtccaaag 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 gaccttaca caatacaaga aaaggtatac atataggacc agggcaagca
961 ttctatgcaa tgagtaacat aataggagat ataagacaag cacattgtaa cgtcagtaaa
1021 acaaaatgga atgcaacggt atcccaggta gccaaacaat tagggaataa gtttaacaac
1081 agaacaataa tctttaatag ctctcagga ggggatatag aacttacaag acatagtttt
1141 aactgtggag gagagttttt ctattgtaat acatcaggcc tgtttaatag cacttggat
1201 aatgacacct gaatggcaa tgacacaaa gagtcaaatg aaacctcgtc
aaatggcacc
1261 **ataa**ttccac 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 caccaccac ggcaaagaga agagtgggtg agagagaaaa aagagcagct
1561 ggactgggag ttatgttctt tgggttctta ggagcagccg gaagcactat gggcgcgcg
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 ctgtgcctg gaactctagt
1861 tggagtaata gatctcagga ggaaatatgg aataacatga cctggctgca atgggagaaa
1921 gaaattggca attacacag tataatatat agactacttg aagaatcgca gaaccagcag
1981 gaaaagaatg aacaagactt attggcattg gacaagtgg caagtctgtg gaattggttt
2041 agcatatcaa actggctatg gtatattaaa atatttataa tgatagtagg aggcttaata
2101 ggattaagaa tagtttttac tgtgctttct ataataaata gagttaggca gggatactca
2161 cccttgtcgt ttcagacca cccccagtt caaggggatc tcgacagacc cggaagaatc

```

2221 gaagaagaag gtggagagca aggcagagac agatcgattc gattagttag cgggttctta
2281 gcgcttgccct gggacgacct ccggagcctg tgcctcttca gctaccaccg cttgagagac
2341 ttcatcttga ttgcagcgag gactgtggaa cttctgggac acagcagtct caaggggttg
2401 agactggggt gggaaaggcct caagtatctg gggaaatctt tgatatattg ggggaaggag
2461 ataaaaaata gtgctgttaa tttggtgat actatagcaa tagcagttagc tggtggaca
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

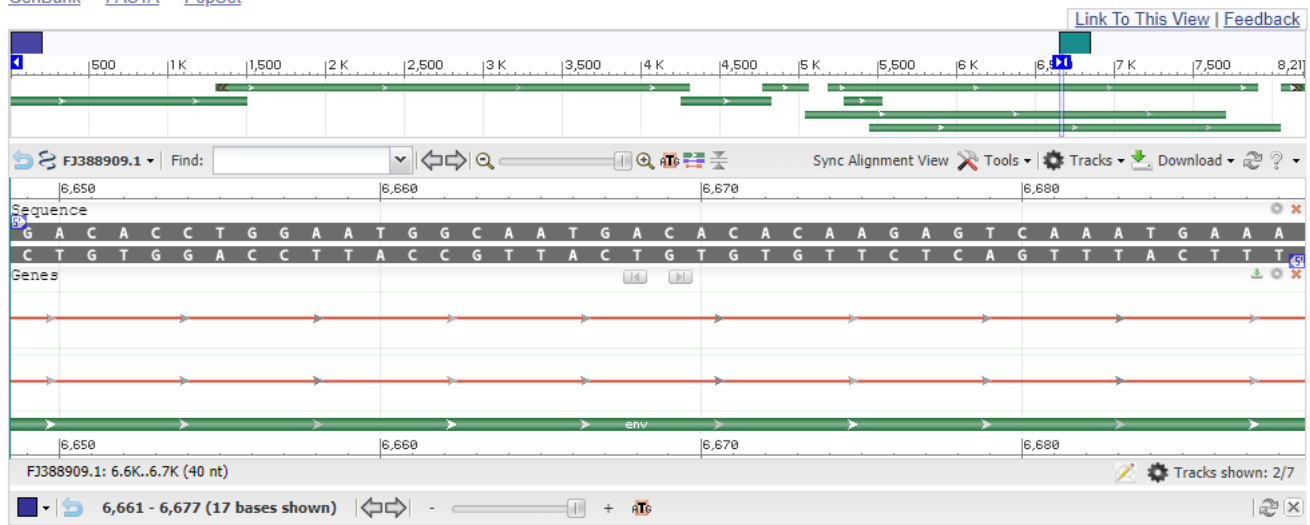
```

Query    5      GGCAATGACACACAAGA  21
          |||
Sbjct   6661   GGCAATGACACACAAGA  6677

```

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
[GenBank](#) [FASTA](#) [PopSet](#)



Come si può osservare a pagina 179, le lettere colorate in rosso (all'interno della sequenza più ampia di lettere evidenziata in giallo) sono parte (“**GGCAATGACACACAAGA**”) della **Sequenza n° 1/3** (“**CTCGGGCAATGACACACAAGAA**”). Questa parte di sequenza è stata identificata come corrispondente al *Virus dell'HIV-1* e si riferisce ad una **regione “misc_feature”** (cioè ad una regione di interesse biologico che non può essere descritta da nessun'altra caratteristica chiave, ovvero una caratteristica nuova o rara) definita “**nonfunctional envelope glycoprotein due to mutation**” (*glicoproteina dell'involucro non funzionale a causa di mutazione*) del **gene Env** (un *gene virale* che codifica per la proteina che forma l'involucro virale) [[https://en.wikipedia.org/wiki/Env_\(gene\)](https://en.wikipedia.org/wiki/Env_(gene))] and [https://en.wikipedia.org/wiki/Viral_envelope].

Un **involucro virale** è lo strato più esterno di molti tipi di virus. Esso protegge il materiale genetico nel suo ciclo di vita quando viaggia tra le *cellule ospiti*. Gli involucri sono tipicamente derivati da *porzioni delle membrane delle cellule ospiti*, ma includono alcune glicoproteine virali. **Possono aiutare i virus a evitare il sistema immunitario dell'ospite**. Le **glicoproteine sulla superficie dell'involucro** servono per identificare e legarsi ai *siti recettori* sulla membrana dell'ospite. L'*involucro virale* si fonde quindi con la membrana dell'ospite, consentendo al *capside* (un altro *strato proteico*, tra l'*involucro* e il *genoma*) e al *genoma virale* di **entrare e infettare l'ospite**. La cellula da cui germogli un virus spesso muore o si indebolisce e rilascia più **particelle virali** per un periodo prolungato (https://en.wikipedia.org/wiki/Viral_envelope).

L'*involucro* del **doppio strato lipidico** di questi virus è relativamente sensibile all'*essiccazione*, al *calore*, al *sapone* e ai *detergenti*, quindi questi virus sono più facili da sterilizzare rispetto ai *virus senza involucro*, hanno una sopravvivenza limitata al di fuori degli ambienti dell'ospite e in genere devono trasferirsi direttamente da ospite a ospite.

Tra i **Virus con involucro** (https://en.wikipedia.org/wiki/Viral_envelope), tra gli altri, annoveriamo i *Virus a DNA* (**Herpesvirus**), i **Retrovirus** (come l'*HIV*) e i **Coronavirus** (*Virus a RNA*). **I virus avvolti possiedono una grande adattabilità, possono cambiare in breve tempo per eludere il sistema immunitario e causare infezioni persistenti** (https://en.wikipedia.org/wiki/Viral_envelope), quelle che si verificano perché il corpo non è in grado di eliminare l'organismo dopo l'infezione iniziale (https://en.wikipedia.org/wiki/Infection#Persistent_infection).

Il prodotto maturo del **gene Env** è la **proteina virale spike** (la cui notorietà ha raggiunto il grande pubblico perché molto nominata dall'inizio della **pandemia del SARS-CoV-2**), che ha **due parti principali**: la *proteina di superficie* (SU) e la *proteina transmembrana* (TM). Il tropismo del virus (costituito dalle cellule e dai tessuti di un ospite che supportano la crescita di un particolare virus o batterio) è determinato dal dominio della **proteina SU** perché è responsabile della *funzione di legame del recettore del virus*. **Il dominio SU determina quindi la specificità del virus per una singola molecola recettoriale** ([https://en.wikipedia.org/wiki/Env_\(gene\)](https://en.wikipedia.org/wiki/Env_(gene))).

Ritornando ad osservare la **sequenza di basi** identificata dalla ricerca BLAST come corrispondente al *Virus dell'HIV-1*, notiamo che la parte iniziale (si vedano le basi colorate in blu a pagina 179) di questa **sequenza** si riferisce alla **parte finale** del *gene Vpu* (**CDS <1..83** /gene="vpu"/codon_start=3/product="vpu protein") che codifica per la **proteina Vpu**.

Vpu è una *proteina accessoria* che nell'HIV è codificata dal *gene vpu*. **Vpu** sta per "Viral Protein U" (https://en.wikipedia.org/wiki/Vpu_protein). La proteina Vpu agisce nella degradazione dei CD4 nel reticolo endoplasmatico (ricordo che l'HIV distrugge i *linfociti T CD4+* e *inibisce la risposta immunitaria cellulo-mediata*) e nell'aumento del rilascio di **virioni** (particelle virali complete costituite da *acido nucleico* circondato da uno strato protettivo di proteine chiamato *capside*) dalla **membrana plasmatica delle cellule infette**. *Vpu* induce la degradazione del recettore virale CD4 e quindi partecipa alla generale sottoregolazione dell'espressione di CD4 durante il corso dell'infezione da HIV. Si ritiene che la degradazione CD4 mediata da Vpu prevenga CD4- Env legandosi nel reticolo endoplasmatico per facilitare il corretto assemblaggio Env nei virioni. **Vpu si trova nelle membrane delle cellule infette, ma non nelle particelle virali stesse** (https://en.wikipedia.org/wiki/Vpu_protein).

E' molto importante puntualizzare che esistono delle **somiglianze strutturali tra Vpu e un'altra piccola proteina virale, M2, codificata dal virus dell'Influenza A**. *Vpu* permea anche le **membrane dei batteri** e, pertanto, è considerato un membro della famiglia delle ***Viroporine***, *proteine virali multifunzionali piccole e solitamente idrofobiche che modificano le membrane cellulari, facilitando così il rilascio del virus dalle cellule infette* (<https://en.wikipedia.org/wiki/Viroporin>).

Le ***Viroporine*** si trovano in una varietà di genomi virali, ma sono particolarmente comuni nei **virus a RNA**. Molti virus che causano malattie umane esprimono *viroporine*. Questi virus includono *virus dell'epatite C*, *HIV-1*, *virus dell'influenza A*, *poliovirus*, *virus respiratorio sinciziale* e *SARS-CoV*. La maggior parte delle *viroporine* non sono essenziali per la **replicazione virale**, anche se la loro assenza riduce significativamente l'efficienza della propagazione virale (le *viroporine* sono quindi considerate *fattori di virulenza*). In alcuni casi gli effetti di *permeabilizzazione di membrana delle viroporine* attivano l'*inflammasoma*, un *complesso proteico* associato all'attivazione dell'**immunità innata** che, se **iperattivo**, può causare **sintomi di malattia**.

Le ***Viroporine*** sono spesso considerate bersagli farmacologici attraenti per lo sviluppo di farmaci antivirali, ma attualmente l'uso clinico è relativamente raro e controverso (<https://en.wikipedia.org/wiki/Viroporin>).

9.4 PREDIZIONE DEI CODONI COSTITUTIVI DEL SEGMENTO DELLA SEQUENZA N° 1/3 E DISCUSSIONE SUI LORO IPOTETICI PRODOTTI

Scendendo alla **riga 1201** della sequenza mostrata a **pagina x** incontriamo le lettere colorate in rosso (**GGCAATGACACACAAGA**, qui sotto riportate) che sono parte della **Sequenza n° 1/3** (**CTCGGGCAATGACACACAAGAA**).

```

1201      gaatggcaa tgacacacaa gagtcaaatg aaacctcgtc
          aaatggcacc
1261      ataa
  
```

Nella sezione “Obiettivi del capitolo IX” si è accennato alla proposta di una **IPOTETICA predizione dei codoni costitutivi di questo segmento della Sequenza n° 1/3 e di quelli adiacenti ad esso** (qualora tali *codoni* diventassero codificanti per proteine). Qui di seguito la *predizione dei codoni costitutivi* e la *sequenza completa dei loro aminoacidi corrispondenti*.

```

INIZIO
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
  
```

Nelle pagine successive sono mostrate *diverse ricerche BLAST* effettuate su differenti **sequenze di aminoacidi** (formate da segmenti dei *codoni costitutivi*) e sono discussi i *principali risultati ottenuti*.

9.5 SEQUENZA DI AMINOACIDI “M A M T H”

In questo paragrafo sono mostrati alcuni risultati della ricerca BLAST effettuata sulla sequenza di aminoacidi **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 sarta]	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	

Proteina di funzione sconosciuta (DUF1317)

Questa famiglia è composta da diverse **ipotetiche proteine batteriche** e **fagiche** di circa 60 residui di lunghezza. **La funzione di questa famiglia è sconosciuta** (<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

Una [*proteina acil-carrier*] *S-maloniltransferasi* è un enzima che appartiene alla famiglia delle **transferasi** [classe di enzimi che attua il trasferimento di gruppi funzionali specifici da una molecola (chiamata donatrice) a un'altra (chiamata accettore). Si veda anche quanto scritto alle pagine 233 e 253 sulle *glicosiltransferasi* e a pag. 241 sull'*Immunoglobulin E-set*, in particolare quelle *aciltransferasi* che trasferiscono gruppi diversi dai gruppi amminoacilici. Questo enzima partecipa alla **biosintesi degli acidi grassi** ([https://en.wikipedia.org/wiki/\(acyl-carrier-protein\)_S-maloniltransferase](https://en.wikipedia.org/wiki/(acyl-carrier-protein)_S-maloniltransferase)).

La proteina ha un'architettura di **tipo α/β** , ma la piega è unica. Il sito attivo dedotto dalla posizione del catalizzatore **Ser92** (*serina 92*) contiene un tipico gomito nucleofilo come osservato nelle **α/β idrolasi**. La **serina 92** è legata all'*idrogeno a His 201* in modo simile a varie **serina idrolasi**. Gli enzimi che partecipano alla **biosintesi degli acidi grassi** sono interessanti **bersagli per gli antibatterici**. **FabD** (*Acyl-Carrier-Protein S-Malonyltransferase*) è un obiettivo ragionevole; tuttavia, non sono stati riportati **inibitori FabD** in letteratura e articoli di revisione su questo argomento. La struttura semplice e l'acidità del *malonato* sembrano consentire pochi approcci alla *sintesi di derivati* (che agiscono come **potenziali inibitori**) che *mantengono il carattere della molecola* ([https://en.wikipedia.org/wiki/\(acyl-carrier-protein\)_S-maloniltransferase](https://en.wikipedia.org/wiki/(acyl-carrier-protein)_S-maloniltransferase)).

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

Gli *SDR (deidrogenasi/reduttasi a catena corta ed enzimi correlati)* rappresentano una delle più *antiche famiglie di proteine* e contribuiscono alle funzioni e alle interazioni essenziali di tutte le forme di vita (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2896744/>, <http://pfam.xfam.org/family/PF00106.25>). La maggior parte degli enzimi di questa famiglia sono noti per essere *ossidoreduttasi dipendenti* da **NAD** o **NADP**. Un'*ossidoreduttasi* è un enzima che catalizza il *trasferimento di elettroni* da una molecola, il *riducente*, chiamato anche donatore di elettroni, a un altro, l'ossidante, chiamato anche *accettore di elettroni*. Questo gruppo di enzimi utilizza solitamente **NADP** + (*nicotinamide adenina dinucleotide fosfato*) o **NAD** + come cofattori. Le *ossidoreduttasi transmembrana* creano catene di trasporto degli elettroni in batteri, cloroplasti e mitocondri, compresi i complessi respiratori I, II e III. Alcuni altri possono associarsi a *membrane biologiche* come *proteine della membrana periferica* o essere ancorati alle membrane attraverso una singola elica trans membrana (<https://en.wikipedia.org/wiki/Oxidoreductase>). Diverse proprietà delle **ossidoreduttasi dipendenti da NAD (P) H** sono state prese di mira dall'**ingegneria delle proteine** e/o da **approcci di evoluzione diretta** per renderle più adatte alle applicazioni. Questi includono il *potenziamento delle proprietà cinetiche degli enzimi*, l'*aumento della loro stabilità* e la *modifica della specificità degli enzimi per il loro substrato o cofattore* (<https://www.sciencedirect.com/science/article/pii/S157096391730273X?via%3Dihub>).

polyketide synthase [Candidatus Entotheonella sarta]

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
```

Polichetidi sintasi (PKSs)

Le **polichetidi sintasi (PKSs)** sono una famiglia di *enzimi multidominio o complessi enzimatici* che producono **polichetidi**, un'ampia classe di **metaboliti secondari**, in batteri, funghi, piante e alcuni lignaggi animali. **Le biosintesi dei polichetidi condividono sorprendenti somiglianze con la biosintesi degli acidi grassi** (https://en.wikipedia.org/wiki/Polyketide_synthase). Le **sintasi di polichetide** sono una fonte importante di *piccole molecole naturali* utilizzate per la **chemioterapia**. Molti degli **antibiotici** comunemente usati, come la **tetraciclina** e i **macrolidi**, sono prodotti dalle **sintasi di polichetide**.

[La **tetraciclina** è usata per trattare una serie di *infezioni*, come acne, colera, brucellosi, peste, malaria e sifilide e alcune **infezioni rare** o esotiche, tra cui l'**antrace**. Il meccanismo d'azione per l'**effetto antibatterico delle tetracicline** si basa sull'**interruzione della traduzione delle proteine nei batteri**, danneggiando così la capacità dei microbi di crescere e ripararsi; tuttavia, la traduzione delle proteine viene interrotta anche nei mitocondri eucariotici, provocando effetti che possono confondere i risultati sperimentali (<https://en.wikipedia.org/wiki/Tetracycline>). In **ingegneria genetica**, la **tetraciclina** è utilizzata in **attivazione trascrizionale**. Essa è stata utilizzata come "**interruttore di controllo**" ingegnerizzato ed è in grado di sviluppare un **retrovirus** che ha indotto un particolare tipo di *leucemia* nei topi e quindi di "**accendere**" e "**spegnerne**" il cancro attraverso la sua *somministrazione* (<https://en.wikipedia.org/wiki/Tetracycline>)].

[I **macrolidi** appartengono alla classe dei **polichetidi** dei prodotti naturali. Il *primo macrolide* scoperto fu l'**eritromicina**. Alcuni macrolidi hanno **attività antibiotica** (soprattutto per trattare le *infezioni* causate da **batteri Gram-positivi**) o **antimicotica** e sono usati come farmaci, tra cui **azitromicina**, **ZITROMAX** (<https://www.drugs.com/zithromax.html>), e **claritromicina**, derivati da *eritromicina modificata chimicamente*. I **macrolidi** sono **batteriostatici** in quanto **sopprimono o inibiscono la crescita batterica piuttosto che ucciderli completamente**. Il meccanismo d'azione dei macrolidi è l'**inibizione della biosintesi delle proteine batteriche** e si pensa che lo facciano *impedendo alla peptidiltransferasi* di aggiungere il *peptide* in crescita attaccato al tRNA al successivo *amminoacido*, nonché inibendo la *traduzione ribosomiale batterica*. Un altro potenziale meccanismo è la *dissociazione prematura del peptidil-tRNA dal ribosoma*. I macrolidi sono attivamente concentrati all'interno dei *leucociti* e quindi vengono *trasportati nel sito dell'infezione* (<https://en.wikipedia.org/wiki/Macrolide>)].

Altri **polichetidi** importanti a livello industriale sono il **sirolimus** (immunosoppressore), la **lovastatina** (farmaco anticolesterolo) e l'**epotilone B** (farmaco antitumorale)

(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>) è presente nello **sperma** e *funziona nella motilità flagellare e nella struttura degli assonemi* (<https://www.omim.org/entry/618146>) [l'*assomema* è una struttura citoscheletrica basata sui microtubuli che forma il nucleo di un cilio o flagello (<https://en.wikipedia.org/wiki/Axoneme>)]. L'*abbattimento di Wdr66* in *T. brucei* ha portato a una **diminuzione della proliferazione cellulare** rispetto al tipo selvatico. Il *knockdown del Wdr66* ha influenzato la *motilità flagellare* e ha provocato la *disorganizzazione dell'assonema*. E' stato dimostrato che il *dominio C-terminale* è necessario per la funzione della proteina e che **WDR66** era necessario per la **formazione della guaina mitocondriale attorno all'assonema durante la flagellogenesi** (<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	

Si veda quanto scritto sopra.

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*) è un gene per la codifica delle proteine. Tra le sue vie correlate vi sono l'*espressione genica* e l'*infezione da virus Herpes simplex 1*. Le annotazioni Gene Ontology (GO) relative a questo gene includono il *legame degli acidi nucleici* e l'*attività del fattore di trascrizione che lega il DNA*. Un importante paralogo di questo gene è **ZNF33B**. **ZNF658** Media la *repressione trascrizionale in risposta allo zinco*. **Reprime diversi geni**, inclusi SLC30A5, SLC30A10 e CBWD1, legandosi all'elemento regolatore trascrizionale dello zinco (ZTRE) (5'-C [AC] C [TAG] CC [TC] -N (0-50) - [GA] G [ATC] G [TG] G-3') trovato nella regione del promotore. Può svolgere un ruolo nel *controllo della biogenesi dei ribosomi*, regolando prevalentemente *i livelli di rRNA*, nonché quelli di diverse *proteine ribosomiali*, coordinando così questo processo altamente esigente di zinco con l'apporto di zinco disponibile (<https://www.genecards.org/cgi-bin/carddisp.pl?gene=ZNF658>). Sulle **Zinc Finger Protein** si legga anche quanto riportato a pag. 232.

Dai risultati mostrati nelle pagine precedenti si è osservato che la sequenza **M A M T H** **non ha evidenziato allineamenti significativi** con sequenze di *Pipistrelli*, *Pangolini* o altri "organismi" considerati importanti per l'oggetto di interesse di questo studio.

9.6 LA MODIFICAZIONE DELLA SEQUENZA DI AMINOACIDI “M A M T H” NELLA SEQUENZA DI AMINOACIDI “M V M T H”

Osservando la sequenza **M A M T H** ho notato la presenza di **due aminoacidi di Metionina (M)** le cui posizioni nella sequenza mi hanno richiamato alla mente un breve segmento delle sequenze (già mostrate nei capitoli precedenti e che ripropongo qui di seguito) del “**Bat SARS-like coronavirus**” e del “**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	GTGAG ATGGTCATG TGTGGCGG	22	
Sbjct	15418	GTGAG ATGGTCATG TGTGGCGG	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	GTGAG ATGGTCATG TGTGGCGG	22	
Sbjct	15377	GTGAR ATGGTCATG TGTGGCGG	15398	
CDS:ORF1ab polyprote	5056	S E	M V M C G	G

ATTENZIONE: R = A o G

Incuriosito da questa “coincidenza”, ho fatto una ricerca BLAST di approfondimento sulla sequenza di 5 aminoacidi “**M V M C G**” del “**Bat SARS-like coronavirus**” e del “**Severe acute respiratory syndrome coronavirus 2**”. Tralasciando le sequenze già nominate, tutti gli altri *allineamenti significativi* riscontrati si riferiscono alle seguenti sequenze del *Batterio della specie Pseudomonas*:

non-ribosomal peptide synthetase [Pseudomonas sp. MWU13-2860]Sequence ID: [RBH56499.1](#) Length: 10329 Number of Matches: 1Range 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

```

Sintetasi peptidiche nonribosomiali

I peptidi nonribosomiali (NRP) sono una classe di *metaboliti secondari peptidici*, solitamente prodotti da microrganismi come *batteri* e *funghi*. I peptidi nonribosomiali si trovano anche negli organismi superiori, come i *nudibranchi*, ma si pensa che siano prodotti dai **batteri** all'interno di questi organismi. Sebbene esista un'ampia gamma di peptidi che non sono sintetizzati dai *ribosomi*, il termine **peptide nonribosomiale** si riferisce tipicamente a un insieme molto specifico di questi (https://en.wikipedia.org/wiki/Nonribosomal_peptide). I **peptidi non ribosomiali** sono sintetizzati da **sintetasi peptidiche nonribosomiali** che, a differenza dei ribosomi, sono **indipendenti dall'RNA messaggero**. Ciascuna sintetasi peptidica nonribosomiale può sintetizzare solo un tipo di peptide. A volte, la disidratazione viene eseguita sulle *serine*, con conseguente *deidroalanina*. I **peptidi nonribosomiali** sono una famiglia molto diversificata di prodotti naturali con una gamma estremamente ampia di **attività biologiche e proprietà farmacologiche**. Sono spesso tossine, siderofori o pigmenti. Gli antibiotici peptidici nonribosomiali, i citostatici e gli immunosoppressori sono in uso commerciale (https://en.wikipedia.org/wiki/Nonribosomal_peptide).

non-ribosomal peptide synthase/polyketide synthase [Pseudomonas asplenii]Sequence ID: [WP_121140405.1](#) Length: 10302 Number of Matches: 1[See 4 more title\(s\)](#) [See all Identical Proteins\(IPG\)](#)

Related Information

[Identical Proteins](#)-Identical proteins to WP_121140405.1Range 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

```

Polichetidi sintasi (PKSs)

Si veda quanto già scritto a pag 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_090208474.1](#) Length: 10286 Number of Matches: 1

[See 1 more title\(s\)](#) [See all Identical Proteins\(IPG\)](#)

Related Information

[Identical Proteins](#)-Identical proteins to WP_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
```

Peptide sintetasi batterica non ribosomiale che sintetizza jessenipeptina

La **jessenipeptina** è un lipopeptide ciclico nonribosomiale regolato dal quorum sensing che mostra una potente attività inibitoria sinergica contro lo *Staphylococcus aureus* resistente alla meticillina clinicamente rilevante (MRSA) (<https://www.pnas.org/content/115/15/3758>).

Riguardo le caratteristiche del *Batterio Pseudomonas* e le sue significative implicazioni nelle *infezioni polmonari* rimando alla letteratura specializzata e a un mio studio originale e molto articolato [([Analisi e modificazioni di Sequenze di DNA o di RNA con la T T E S](#) (Capitolo I° - Parte Prima) e pagina 302 di [Analisi e modificazioni di Sequenze di DNA o di RNA con la T T E S](#) (Capitolo I - Parte Seconda)].

A questo punto, per scoprire eventuali collegamenti tra la sequenza “**M V M C G**” e la sequenza “**M A M T H**” ho sostituito l’aminoacido **Alanina** (codone **GCA** nel segmento “**ATGGCAATGACACAC**” della **Sequenza n° 1/3**) con l’aminoacido **Valina** (sostituendo nel codone **GCA** la base “**C**” con la base “**T**”), ottenendo la sequenza “**M V M T H**” e la sequenza “**ATGGTAATGACACAC**”.

Qui di seguito e nella pagina successiva sono mostrati **due** risultati della ricerca BLAST sulla sequenza di aminoacidi “**M V M T H**”.

immunoglobulin heavy chain junction region [Homo sapiens]
Sequence ID: [MOP80811.1](#) Length: 16 Number 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
  
```

Regione di giunzione della catena pesante delle emoglobine

La **catena pesante delle immunoglobuline (IgH)** è la grande *subunità polipeptidica* di un **anticorpo** (*immunoglobulina*). Nel genoma umano, i loci del *gene IgH* si trovano sul *cromosoma 14*. Un tipico **anticorpo** è composto da *due catene pesanti di immunoglobuline (Ig)* e *due catene leggere di Ig*. Esistono diversi tipi di catene pesanti che definiscono la classe o l’isotipo di un anticorpo. Questi tipi di catene pesanti variano tra i diversi animali. Tutte le *catene pesanti* contengono una serie di *domini immunoglobulinici*, solitamente con un *dominio variabile* (V_H) importante per il legame dell’antigene e diversi domini costanti (C_H1, C_H2, ecc.). **La produzione di una catena pesante vitale è un passaggio chiave nella maturazione delle cellule B** (https://en.wikipedia.org/wiki/Immunoglobulin_heavy_chain#:~:text=Each%20heavy%20chain%20has%20two,hinge%20region%20for%20added%20flexibility).

Qui di seguito la *sequenza delle basi* (e degli *aminoacidi*) corrispondenti alla **regione di giunzione della catena pesante delle emoglobine**:

ATG GTC ATG ACC CAC
M V M T H

Nonostante gli aminoacidi restino gli stessi (ricordo che codoni diversi possono codificare lo stesso aminoacido), le *due lettere* in grassetto giallo nella sequenza mostrata sopra sono **differenti** rispetto a quelle corrispondenti nel segmento “ATG GTA ATG ACA CAC” della **Sequenza 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="CARHAA MVMTH PPDYW"

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 è stato preso in considerazione perché già identificato dalla ricerca BLAST fatta precedentemente sulla sequenza **M A M T H**. Purtroppo, di questo allineamento significativo non sono riuscito a trovare il file in cui è possibile controllare le basi.

Per fare un controllo incrociato, ho eseguito una ricerca BLAST sulla sequenza di basi “ATGGTAATGACACAC”. Qui di seguito e nella pagina successiva sono mostrati i risultati.

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		

Come si osserva dai due risultati della ricerca BLAST sopra riportati, la sequenza di basi “ATGGTAATGACACAC” si riferisce a due sequenze dell’HIV-1, ma gli aminoacidi sono diversi da quelli attesi perché *differente è la corrispondenza di codifica dei codoni*. Nonostante la *non corrispondenza dei codoni*, è osservabile una struttura simile alla Query **M V M T H**, ovvero *tre aminoacidi di cui due laterali uguali e uno centrale diverso (N G N)*.

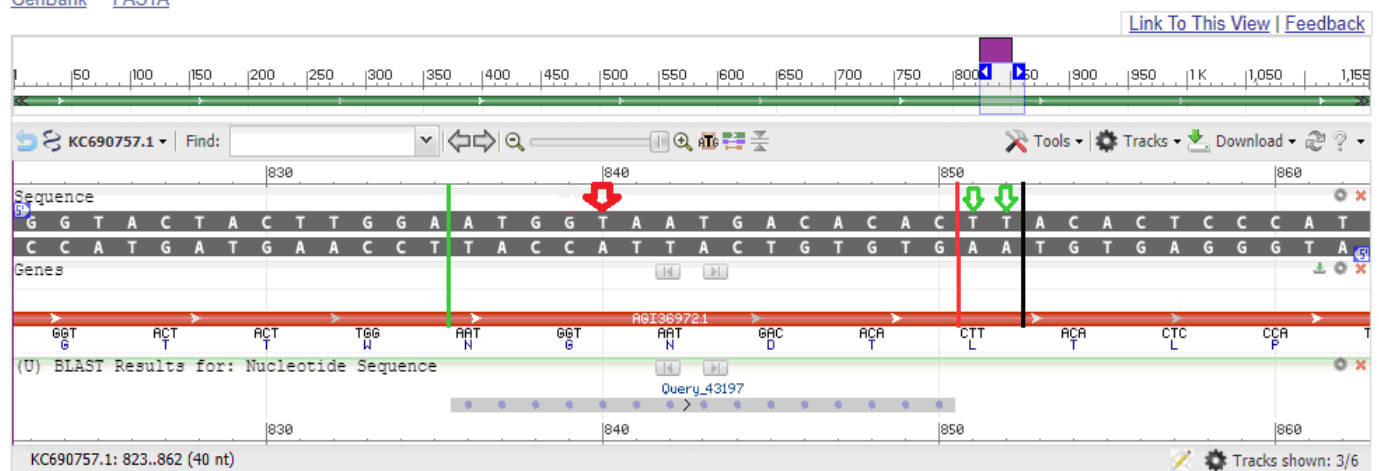
Essendo chiaramente non soddisfatto di questi risultati e spinto dalla necessità di chiarire il ruolo di questa *ripetuta configurazione* (tre aminoacidi di cui *due laterali uguali e uno centrale diverso*, configurazione che verrà ritrovata in altre sequenze mostrate nei capitoli successivi) nella breve sequenza “ATGGTAATGACACAC”, ho approfondito ulteriormente lo studio alla ricerca di informazioni che potessero darmi risposte più chiare e consentirmi di proseguire o di abbandonare definitivamente questa indagine.

Osservando il grafico presentato qui di seguito della sequenza “HIV-1 isolate KRC414204 from South Korea envelope glycoprotein (env) gene, partial cds” (si veda la pagina precedente) ho notato che la sequenza proseguiva con le basi “TT” (si vedano le frecce verdi nel grafico).

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

GenBank: KC690757.1

[GenBank](#) [FASTA](#)



Ho fatto una nuova ricerca BLAST aggiungendo **due basi “T”** finali alla sequenza **“ATGGTAATGACACAC”** per rendere la sequenza più lunga e avere più opportunità di allineamenti significativi potenzialmente interessanti. Oltre le prevedibili sequenze relative a **HIV-1 isolate KRC414204 from South Korea envelope glycoprotein (env) gene** e a **HIV-1 isolate 17 from Zambia envelope glycoprotein (env) gene**, dalla ricerca emergono *altri interessanti* risultati: **tre** allineamenti significativi della sequenza **“ATGGTAATGACACACTT”** si riferiscono a **tre sequenze (una del cromosoma 13 e due del cromosoma 7) di due Pipistrelli.**

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      TGGTAATGACACACTTT 17
             |||
Sbjct 14470538 TGGTAATGACACACTTT 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      GTAATGACACACTTTT 18
             |||
Sbjct 69199040 GTAATGACACACTTTT 69199054
```

“**Audentes fortuna iuvat**” (“*Il destino favorisce chi osa*”) è la celebre *locuzione latina* comunemente tradotta in italiano “**La fortuna aiuta gli audaci**”. Mai motto fu così appropriato!

Rifacendo la ricerca BLAST sulla sequenza “**ATG GTA ATG ACA CAC TT**” per confermare gli *allineamenti significativi identificati*, commetto un banalissimo *errore di trascrizione delle lettere della sequenza* e scrivo: “**ATG GTA ATG ACA CAT T**”, omettendo una “**C**” prima delle due “**T**”.

Non si tratta quindi di una vera e propria (involontaria) modificazione della sequenza capace di incidere sui suoi “**Trend Non Manifesti**” (se non marginalmente nella sua parte finale), ma di una **riduzione della sua lunghezza** prima dell’aggiunta delle due basi di **Timina (T)**. Un’effettiva modificazione della sequenza era già avvenuta con la *sostituzione* della lettera “**C**” con la lettera “**T**” nel codone **GCA**, alla sequenza “**G GCA ATG ACA CAC AAG A**” (la sequenza identificata come corrispondente al *Virus dell’HIV-1* e riferita alla *regione “misc_feature”*). Inoltre, l’*omissione* della “**C**” può non incidere significativamente sulla formazione dell’*ultimo codone* perché i codoni “**CAC**” e “**CAT**” codificano entrambi l’aminoacido **Istidina (H)**. Naturalmente, quest’ultima circostanza si realizzerà **solo** se la codifica partirà da uno dei due potenziali codoni **ATG** presenti nella sequenza o da altri precedenti che rispettino questo ordine di formazione dei codoni.

I risultati della ricerca BLAST presentati nel paragrafo successivo rappresentano alcune delle più importanti rivelazioni di questo lunghissimo e laborioso documento.

9.7 “AUDENTES FORTUNA IUVAT”: LA SEQUENZA “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 ferrumequinum F-box and WD repeat domain containing 11 (FBXW11), transcript variant X4, mRNA	Rhinolophus ferrumequinum	30.2	30.2	93%	606	100.00%	4197
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 E3 ubiquitin protein ligase (BTRC), transcript variant X6, mRNA	Rhinolophus ferrumequinum	30.2	30.2	93%	606	100.00%	6368
XM_033130372.1	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),	Manis pentadactyla	30.2	30.2	93%	606	100.00%	5257

**transcript variant X1,
mRNA**

Accession	Description	Species	Length	Score	Identity	GC	Start	End
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	Pipistrellus kuhlii	30.2	30.2	93%	606	100.00%	6422

XM_036437875.1	(BTRC), transcript variant X4, mRNA 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 kuhlii beta-transducin repeat containing E3 ubiquitin protein ligase (BTRC), transcript variant X2, mRNA	Pipistrellus kuhlii	30.2	30.2	93%	606	100.00%	6516
XM_036437873.1	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%	5167227 9
LR862369.1	Pipistrellus pipistrellus genome assembly, chromosome: 13	Pipistrellus pipistrellus	30.2	30.2	93%	606	100.00%	5328154 6
LR862367.1	Pipistrellus pipistrellus genome assembly, chromosome: 11	Pipistrellus pipistrellus	30.2	120	93%	606	100.00%	6938614 4
LR862365.1	Pipistrellus pipistrellus genome assembly, chromosome: 9	Pipistrellus pipistrellus	30.2	30.2	93%	606	100.00%	7254207 2
LR862363.1	Pipistrellus pipistrellus genome assembly, chromosome: 7	Pipistrellus pipistrellus	30.2	30.2	93%	606	100.00%	8454810 4
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	Legionella	30.2	30.2	93%	606	100.00%	3516334

Accession	Organism	Length	Score	E-value	Identities	Gaps	Strand
	genome pneumophila 2300/99 Alcoy						
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	<u>Simian immunodeficiency virus complete genome</u> 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

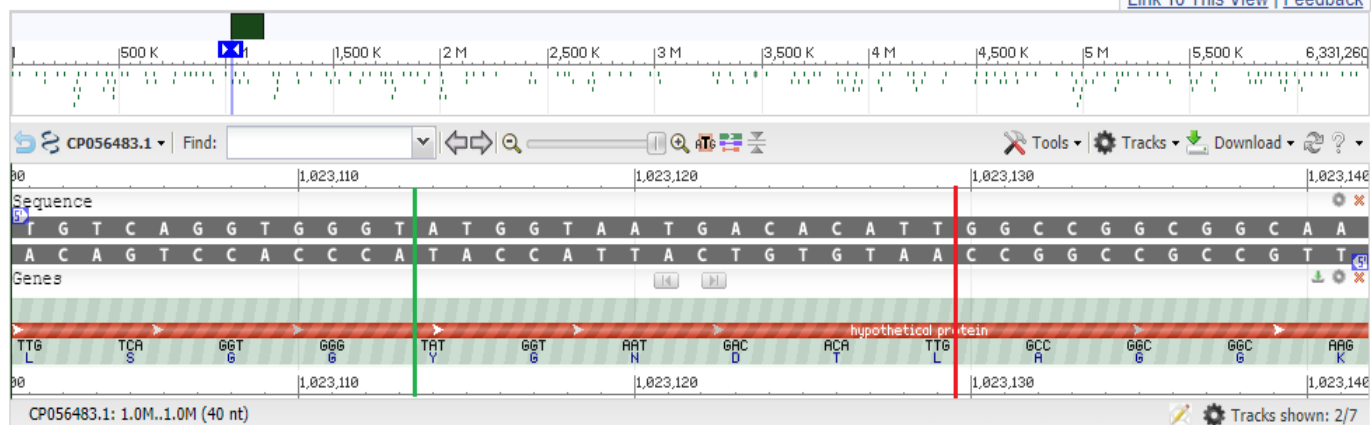
CDS: Putative 1 1 **G** N D T
Query 1 **ATGGTAATG**ACACATT 16
|||||
Sbjct 1023114 **ATGGTAATG**ACACATT 1023129
CDS: **hypothetica** 93 **Y** G N D T **L**

Klebsiella sp. RHBSTW-00464 chromosome, complete genome

GenBank: CP056483.1

[GenBank](#) [FASTA](#)

[Link To This View](#) | [Feedback](#)



La sequenza di aminoacidi **Y G N D T L** di **Klebsiella sp.** è quasi totalmente simile (differisce solamente per il primo aminoacido) alla sequenza **N G N D T L** relativa a **HIV-1 isolate KRC414204 from South Korea envelope glycoprotein (env) gene** e a **HIV-1 isolate 17 from Zambia envelope glycoprotein (env) gene** (si vedano le pagine 194, 195 e 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

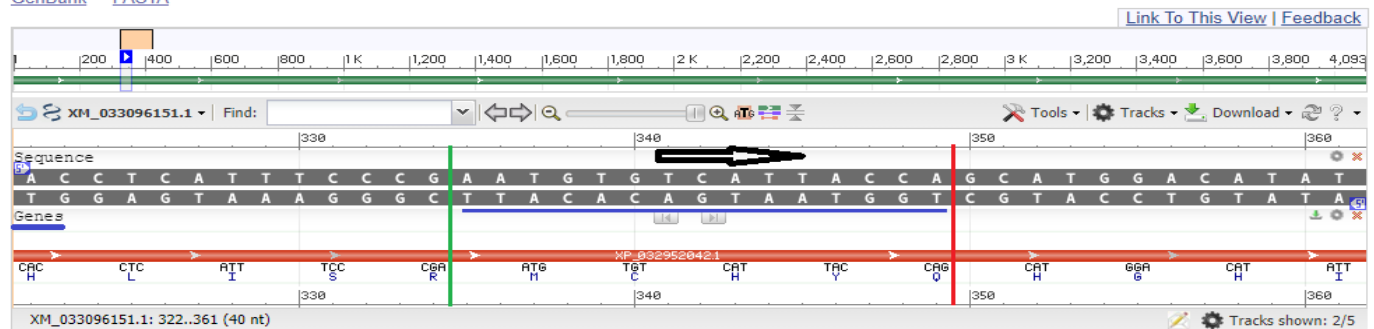
Range 1: 335 to 349 [GenBank](#) [Graphics](#)

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		

PREDICTED: *Rhinolophus ferrumequinum* F-box and WD repeat domain containing 11 (FBXW11), transcript variant X9, mRNA

NCBI Reference Sequence: XM_033096151.1

[GenBank](#) [FASTA](#)



β TrCP2 (ripetizione beta-trasducina contenente proteina 2, nota anche come **Fbxw11** o HOS) è una proteina che nell'uomo è codificata dal gene *FBXW11* (*F-box e WD repeat domain containing 11*) (https://en.wikipedia.org/wiki/FBXW11#cite_note-Koch, A. et al.-38). Questo gene codifica per un membro della famiglia della *proteina F-box* che è caratterizzata da un motivo strutturale di circa 40 residui, la **F-box**. I membri della famiglia *WDR11* sono coinvolti in una varietà di processi cellulari, tra cui la progressione del ciclo cellulare, la trasduzione del segnale, l'apoptosi e la regolazione genica (<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>). Le proteine **F-box** costituiscono una delle quattro sub unità di **ubiquitina proteina ligasi complessa** chiamata *SCFs* (proteine Skp1-Cul1-F-box), che spesso, ma non sempre, riconoscono substrati in maniera *fosforilazione-dipendente*. La β TrCP umana (riferito sia a β TrCP1 che a β TrCP2) è stata originariamente identificata come una **ubiquitina ligasi cellulare** che è legata dalla *proteina virale HIV-1 Vpu* [proteina virale accessoria che nell'*HIV* è codificata dal *gene Vpu*. Questo gene si trova esclusivamente nell'*HIV-1* e nell'*HIV-1-related Simian Immunodeficiency Virus (SIV)*] **per eliminare il CD4** (Cluster di differenziazione 4, glicoproteina che si trova sulla superficie delle cellule immunitarie come le cellule T helper, i monociti, i macrofagi e le cellule dendritiche) collegandolo al meccanismo proteolitico (<https://www.sciencedirect.com/science/article/abs/pii/S1286457908001755#:~:text=Vpu%20is%20an%20accessory%20viral,of%20virion%20release%20from%20cells>). β TrCP si comporta come un **oncoproteina** in alcuni tessuti. Livelli elevati di espressione di β TrCP sono stati trovati nel **cancro del colon-retto, pancreatico, epatoblastoma**, e al **seno** (https://en.wikipedia.org/wiki/FBXW11#cite_note-Koch, A. et al.-38).

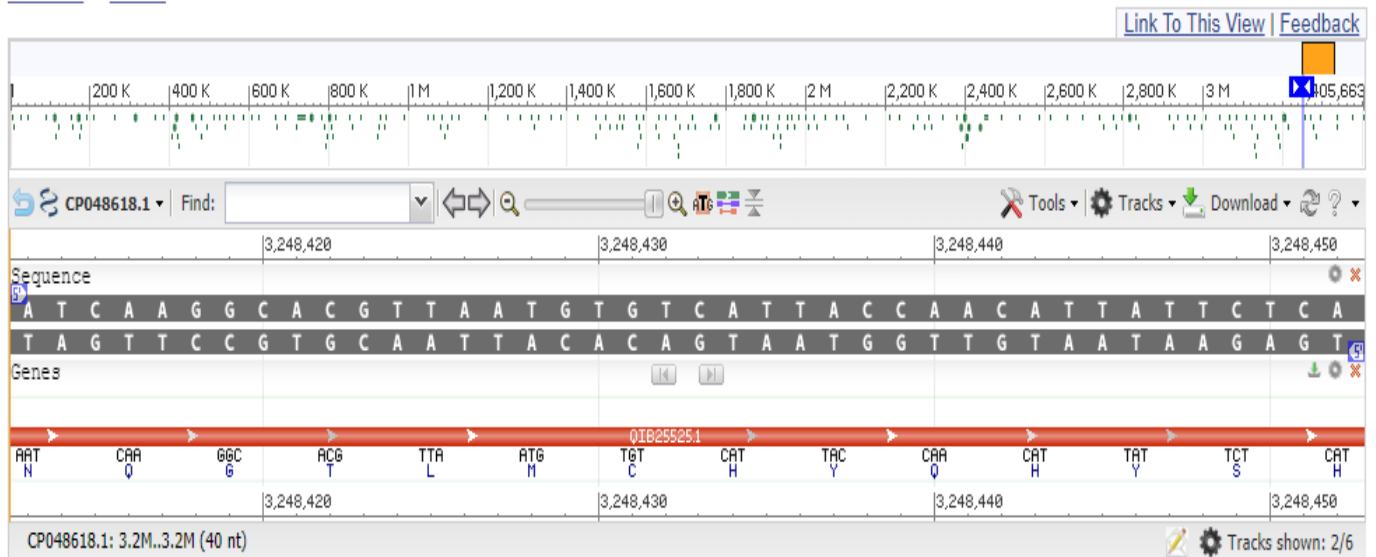
Legionella pneumophila strain ERS1305867 chromosome, complete genome
Sequence ID: CP048618.1 Length: 3405663 Number of Matches: 1
 Range 1: 3248426 to 3248440 [GenBank](#) [Graphics](#)

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		

Legionella pneumophila strain ERS1305867 chromosome, complete genome

GenBank: CP048618.1

[GenBank](#) [FASTA](#)



Metiltransferasi S-adenosil-L-metionina-dipendente (SAM-dependent methyltransferase)

Questa voce rappresenta la superfamiglia della *metiltransferasi S-adenosil-L-metionina-dipendente (SAM MTasi)*. Le *metiltransferasi* trasferiscono un **gruppo metile da un donatore a un accettore**. Le *metiltransferasi* leganti il SAM utilizzano l'onnipresente *donatore di metile SAM* come cofattore per metilare proteine, piccole molecole, lipidi e acidi nucleici. Tutte le *SAM MTasi* contengono un dominio di legame SAM strutturalmente conservato costituito da un foglio beta centrale a sette filamenti che è affiancato da tre alfa-eliche per lato del foglio (<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](#)

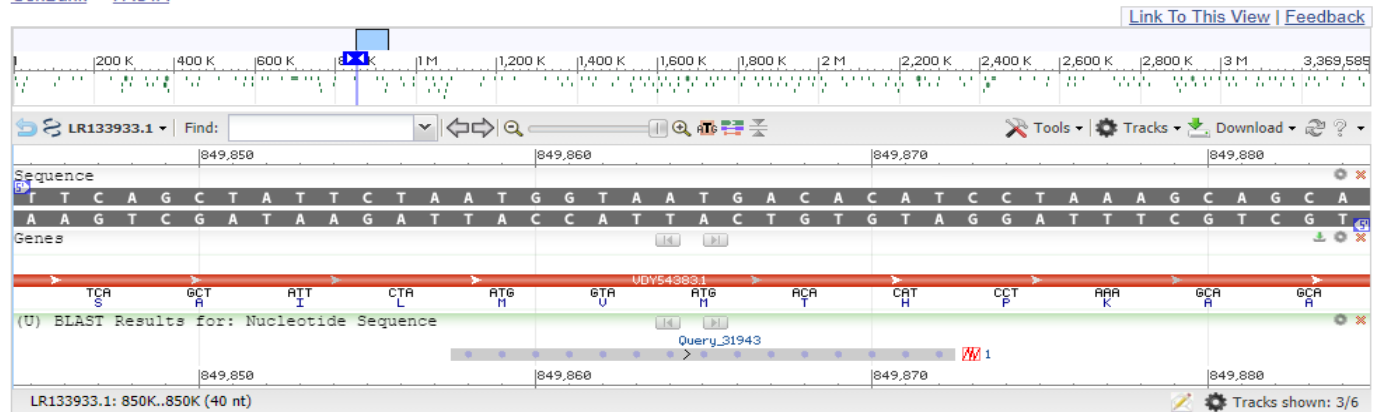
Score	Expect	Identities	Gaps	Strand
30.2 bits(15)	606	15/15(100%)	0/15(0%)	Plus/Plus
CDS: Putative 1	1	M V M T H		
Query	1	ATGGTAATGACA CAT 15		
Sbjct	849858	ATGGTAATGACA CAT 849872		
CDS:putative NAD dep	215	M V M T H		

ATTENZIONE:

E' interessante osservare che il codone **ACA**, codificante la **Treonina (T)** nella sequenza della **Legionella pneumophila**, nella sequenza di **RNA "GT GAA ATG GTC ATG TGT GGC GG"** [tratta dal Gene **RdRp (pSARS-F)** e oggetto del nostro studio] presenta (nella stessa posizione) *basi speculari (TGT)* codificanti la **Cisteina (C)**.

Legionella pneumophila strain NCTC12180 genome assembly, chromosome: 1

GenBank: [LR133933.1](#)
[GenBank](#) [FASTA](#)



Epimerasi/disidratasi dipendente da NAD

Questo dominio si trova nelle **proteine che utilizzano NAD** come cofattore e utilizzano *substrati di nucleotidi di zucchero* per una varietà di reazioni chimiche. Una delle più studiate di queste *proteine* è l'**UDP-galattosio 4-epimerasi** che catalizza la *conversione dell'UDP-galattosio in UDP-glucosio* durante il *metabolismo del galattosio* (<https://www.ebi.ac.uk/interpro/entry/InterPro/IPR001509/>).

Legionella pneumophila subsp. pascullei strain NCTC12273 genome assembly, chromosome: 1
Sequence ID: [LR134380.1](#) Length: 3458917 Number 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	ATGGTAATGACA CAT 15		
Sbjct	864956	ATGGTAATGACA CAT 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: [CP021286.1](#) Length: 3508676 Number 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		

E' interessante notare come **Legionella pneumophila subsp. fraseri strain** utilizzi la *stessa sequenza* per *codificare aminoacidi differenti* che si riferiscono rispettivamente ad **attività catalitica** (*NAD-dependent dehydratase*) e ad una *metiltransferasi* (*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](#)

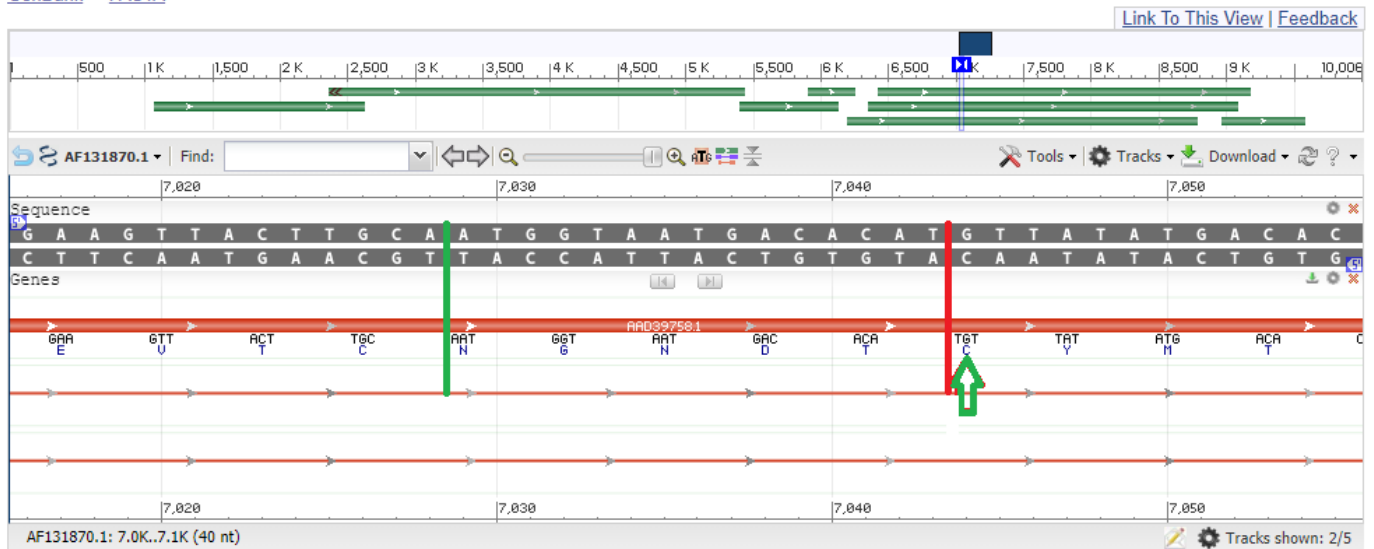
Score	Expect	Identities	Gaps	Strand
30.2 bits(15)	606	15/15(100%)	0/15(0%)	Plus/Plus
CDS: Putative 1	1	G N D T		
Query	1	ATGGTAATGACACAT	15	
Sbjct	7029	ATGGTAATGACACAT	7043	
CDS: envelope glycopr	203	N G N D T		

La sequenza di aminoacidi **N G N** D T di **Simian immunodeficiency virus** è quasi totalmente simile (differisce solamente per l'ultimo aminoacido, si veda la freccia verde nel grafico mostrato qui di seguito) alla sequenza **N G N** D T **L**, sequenza relativa a **HIV-1 isolate KRC414204 from South Korea envelope glycoprotein (env) gene** e a **HIV-1 isolate 17 from Zambia envelope glycoprotein (env) gene** (si vedano le pagine 194,195 e 196).

Simian immunodeficiency virus complete genome

GenBank: AF131870.1

[GenBank](#) [FASTA](#)



Delle **5000** sequenze selezionate dalla ricerca BLAST, le sequenze nella lista appena mostrata sono **solo** alcune di quelle ritenute più importanti e significative per questo studio. Prima di proseguire, desidero ricordare che la sequenza di basi “**ATG**G**TAATGACA**C**ATT**” (su cui è stata effettuata quest’ultima ricerca BLAST) è ottenuta dalla sequenza “**GGCAATGACACAC**” [parte della **Sequenza n°1/3** (**CTCGGGCAATGACACACAAGAA**)] identificata come corrispondente al *Virus dell’HIV-1* e riferita alla *regione “misc_feature”*, alla quale è stata sostituita una lettera (la “**C**” con la “**T**”, così da ottenere un aminoacido di Valina a posto di uno di Alanina) e a cui è stata omessa la “**C**” finale e poi aggiunte due “**T**”. Come sappiamo la **Sequenza n°1/3** è una delle *sequenze modificate* della **Sequenza Originaria**, la sequenza di **22 basi di RNA** “**GTGAAATGGTCATGTGTGGCGG**” tratta dal **Gene RdRp del SARS-CoV-2** (si veda la tabella riassuntiva mostrata qui di seguito).

Sequenza Originaria	G T G A A A T G G T C A T G T G T G G C G G
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
Sequenza 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) - Modificata	A T G G T A A T G A C A C A T T

I risultati della ricerca BLAST sulla “**Sequenza n° 1/3 (HIV-1) – Modificata**” (si veda la tabella mostrata nella pagina precedente) sono particolarmente interessanti. Le **16 basi** considerate hanno identificato un numero notevole di “organismi” e, tra questi, i **due “organismi” oggetto della presente ricerca: tre specie di pipistrelli [Rhinolophus (ferrumequinum) e Pipistrellus (pipistrellus e kuhlii)] e due specie di pangolini [Manis javanica (pangolino malese) e Manis pentadactyla (pangolino cinese)].**

Questi importantissimi *allineamenti significativi* svelano la presenza di collegamenti significativi tra le sequenze del **Gene RdRp del SARS-CoV-2**, dei **Pipistrelli**, dei **Pangolini** e, *inaspettatamente e inusulamente*, anche e soprattutto di una *regione “misc_feature” del Virus dell’HIV-1 (regione definita “nonfunctional envelope glycoprotein due to mutation” del gene Env).*

Molti altri *allineamenti significativi* andrebbero riportati, ma mi limito a quelli che seguono perché gli “organismi” ai quali si riferiscono sono implicati in sintomatologie prevalentemente **infettive, febbrili e polmonari: Klebsiella sp., Legionella pneumophila (subsp. pascullei, subsp. pneumophila, subsp. fraseri, subsp. pneumophila ATCC 33215, subsp. pneumophila str. Thunder Bay, subsp. pneumophila ATCC 43290, sp. PC1000, subsp. pneumophila ATCC 43283, str. Corby, str. Paris e 2300/99 Alcoy) e Simian immunodeficiency virus.**

Klebsiella è un *batterio Gram-negativo* che può causare *cambiamenti distruttivi ai polmoni umani e animali se aspirato*, in particolare agli alveoli, con conseguente gelatina sanguinante, brunastra o gialla come l’espettorato. *Klebsiella* colpisce *uomini di mezza età e anziani con malattie debilitanti* (https://en.wikipedia.org/wiki/Klebsiella_pneumoniae). Si ritiene che questa popolazione di pazienti abbia *difese respiratorie dell’ospite compromesse*, comprese persone con *diabete, alcolismo, tumori maligni, malattie del fegato, malattie polmonari ostruttive croniche, terapia con glucocorticoidi, insufficienza renale e alcune esposizioni professionali*. Molte di queste infezioni si verificano quando una persona è in ospedale per qualche altro motivo. La condizione più comune causata dai batteri *Klebsiella* al di fuori dell’ospedale è comunque la *polmonite*, tipicamente sotto forma di broncopolmonite e anche bronchite. Questi pazienti hanno una maggiore tendenza a sviluppare ascessi polmonari, cavitazione, enfisema e aderenze pleuriche. Ha un tasso di mortalità intorno al 50%, anche con terapia antimicrobica. Gli individui con polmonite da *Klebsiella* tendono a tossire con un caratteristico espettorato, oltre ad avere febbre, nausea, tachicardia e vomito (https://en.wikipedia.org/wiki/Klebsiella_pneumoniae). Il trattamento per la polmonite da *Klebsiella* è costituito da antibiotici come **aminoglicosidi** e **cefalosporine**, la scelta dipende dalle condizioni di salute della persona,

dall'anamnesi e dalla gravità della malattia. Oltre alla *polmonite*, ***Klebsiella*** può anche causare infezioni del tratto urinario, delle vie biliari inferiori e delle ferite chirurgiche, tromboflebite, diarrea, infezione del tratto respiratorio superiore, osteomielite, meningite, batteriemia e sepsi. La sepsi e lo shock settico possono seguire l'ingresso dei batteri nel sangue (https://en.wikipedia.org/wiki/Klebsiella_pneumoniae).

La ***Legionella pneumophila*** è un **batterio Gram-negativo** del genere *Legionella* (il principale batterio patogeno umano in questo gruppo) ed è l'agente eziologico della malattia del legionario, nota anche come **legionellosi**

(https://en.wikipedia.org/wiki/Legionella_pneumophila). Per ottenere *amminoacidi* (gli amminoacidi sono la principale fonte di carbonio ed energia di *L. pneumophila*, che ha quasi 12 classi di *trasportatori ABC*, *permeasi amminoacidici* e molte *proteasi* per servirsene), *L. pneumophila* utilizza l'**effettore F-Box AnkB** (https://en.wikipedia.org/wiki/AnkB_F-Box), una proteina escretta dalla *Legionella pneumophila* che contiene il *dominio F-box*. Le proteine F-box sono coinvolte nell'**ubiquitinazione delle proteine destinate alla degradazione da parte del proteasoma**. **AnkB F-box** è una *proteina* che assembla *proteine poliubiquitinate* della cellula ospite sul lato citoplasmatico del *Vacuolo Contenente la Legionella* (LCV) (https://en.wikipedia.org/wiki/AnkB_F-Box). Questo effettore è necessario per la proliferazione intracellulare all'interno delle cellule umane e dell'ameba e per la proliferazione intrapolmonare della Legionella nei topi (https://en.wikipedia.org/wiki/AnkB_F-Box). La presenza del *dominio F-box* è un esempio di **mimetismo molecolare** (https://en.wikipedia.org/wiki/Molecular_mimicry, argomento di importanza cruciale nello studio dell'*autoimmunità* e del quale consiglio un approfondimento specifico anche per una maggiore comprensione dell'infezione da SARS-CoV-2) dell'*F-box eucariotico* mediante il quale la *Legionella* sfrutta la capacità di degradazione proteasomale della cellula ospite per il proprio bene (https://en.wikipedia.org/wiki/Molecular_mimicry). Tuttavia, la **promozione della degradazione proteasomica** per l'ottenimento di amminoacidi potrebbe non essere l'unica *strategia di virulenza* per ottenere fonti di carbonio ed energia dall'ospite. Gli **enzimi degradativi secreti di tipo II** possono fornire una *strategia aggiuntiva* per generare fonti di carbonio ed energia. I **macrolidi** (*azitromicina* o *claritromicina*) o **fluorochinoloni** (*levofloxacin* o *moxifloxacin*) sono il trattamento standard per la polmonite da Legionella negli esseri umani, con **levofloxacin** considerata la prima linea per aumentare la resistenza all'**azitromicina** (https://en.wikipedia.org/wiki/Legionella_pneumophila).

Simian immunodeficiency virus (SIV), il *Virus dell'Immunodeficienza delle Scimmie*, è una specie di retrovirus (tipo di *virus* che inserisce una copia del suo genoma RNA nel DNA di una cellula ospite che invade, cambiando così il genoma di quella cellula) che causa infezioni persistenti in almeno 45 specie di *primati non umani africani* ([https://en.wikipedia.org/wiki/Simian_immunodeficiency_virus#:~:text=Simian%20immu%20nodeficiency%20virus%20\(SIV\)%20is,of%20African%20non%2Dhuman%20primates](https://en.wikipedia.org/wiki/Simian_immunodeficiency_virus#:~:text=Simian%20immu%20nodeficiency%20virus%20(SIV)%20is,of%20African%20non%2Dhuman%20primates)).

Ceppi di due virus di queste specie di primati si ritiene abbiano *attraversato la barriera di specie in esseri umani*, con conseguente **HIV-2** e **HIV-1**, rispettivamente, i **due virus dell'immunodeficienza umana**. La via più probabile di **trasmissione dell'HIV-1 agli esseri umani** riguarda il contatto con il **sangue degli scimpanzé** che vengono spesso cacciati per la carne selvatica in Africa. A differenza delle infezioni da HIV-1 e HIV-2 negli esseri umani, le infezioni da SIV nei loro *ospiti non umani scimmieschi africani naturali* sembrano in molti casi *non patogene* a causa dell'adattamento evolutivo degli ospiti al virus. Tuttavia, se il virus infetta un *macaco rhesus asiatico o indiano*, questi primati scimmieschi non africani svilupperanno anche l'**AIDS scimmiesco (SAIDS)**, poiché, come gli esseri, non hanno avuto una storia prolungata con il virus. Nel 2012, i ricercatori hanno riferito che l'infezione iniziale delle scimmie rhesus da ceppi SIV resistenti alla neutralizzazione potrebbe essere parzialmente prevenuta attraverso l'uso di un **vaccino anti-SIV_{SME543} che include obbligatoriamente gli antigeni della proteina Env**

([https://en.wikipedia.org/wiki/Simian_immunodeficiency_virus#:~:text=Simian%20immunodeficiency%20virus%20\(SIV\)%20is,of%20African%20non%2Dhuman%20primates](https://en.wikipedia.org/wiki/Simian_immunodeficiency_virus#:~:text=Simian%20immunodeficiency%20virus%20(SIV)%20is,of%20African%20non%2Dhuman%20primates)).

E' necessario puntualizzare che la

Sequ. n° 1/3 (HIV-1) - Modificata

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

che qui è stata utilizzata per identificare **Simian immunodeficiency virus** (e tutti gli altri "organismi" di cui si è già discusso) è una *modificazione minima* di una sequenza di una *regione "misc_feature"* definita "**nonfunctional envelope glycoprotein due to mutation**" (*glicoproteina dell'involucro non funzionale a causa di mutazione*) del **gene Env** (un *gene virale* che codifica per la proteina che forma l'**involucro virale**) del genoma di **Homo sapiens** infetto da **HIV-1** e non del genoma di una scimmia. A sua volta, la sequenza del **gene Env** di cui sopra è stata identificata a partire da una **sequenza modificata** (seguendo il "*Trend Non Manifesto*" n°1) dell'**RNA del SARS-CoV-2**. Queste relazioni dovranno essere necessariamente oggetto di studi e approfondimenti futuri.

9.8 DALLA SEQUENZA DI AMINOACIDI

M V M T H ALLA SEQUENZA DI AMINOACIDI M V M T H F

In questo paragrafo sono mostrati alcuni risultati della ricerca BLAST effettuata sulla sequenza di aminoacidi **M V M T H F**, sequenza ottenuta dall'aggiunta di un aminoacido di *Fenilalanina* (**F**) alla già nota sequenza **M V M T H** (sequenza che, come visto nei paragrafi precedenti, deriva dalla modificazione della sequenza **M A M T H**). La scelta di prolungare la sequenza **M V M T H** con un altro aminoacido è giustificata dalla ricerca di potenziali aminoacidi che potessero far emergere *relazioni significative* tra la sequenza **M V M T H** e il **pipistrello**, il **pagolino** o altri **“organismi”** di interesse per lo studio in atto.

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](#)

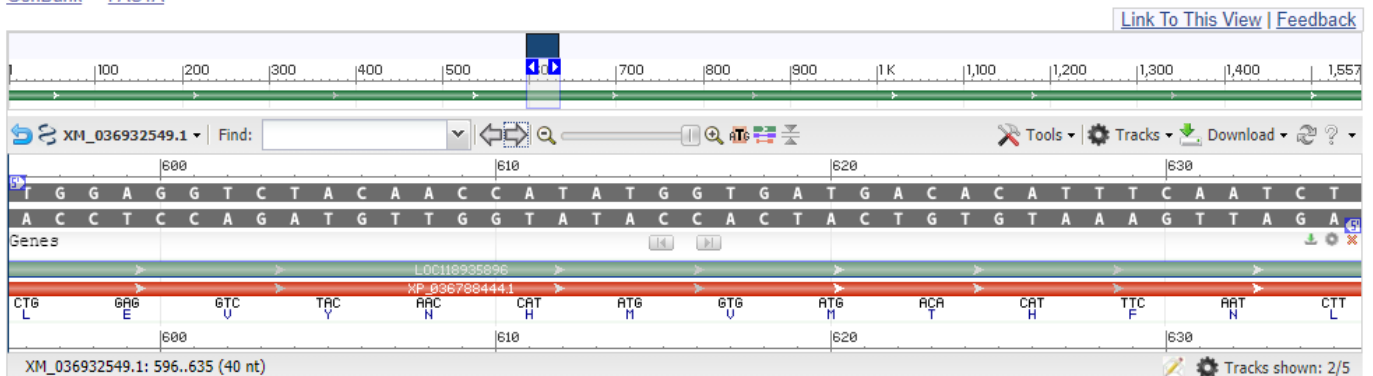
Score	Expect	Identities	Positives	Gaps
26.1 bits(54)	463	6/6(100%)	6/6(100%)	0/6(0%)

Query	1	MVM THF	6
		MVMTHF	
Sbjct	205	MVM THF	210

PREDICTED: Manis pentadactyla olfactory receptor 6Y1-like (LOC118935896), mRNA

NCBI Reference Sequence: XM_036932549.1

[GenBank](#) [FASTA](#)



Olfactory receptor **6Y1**-like (Recettore olfattivo simile a 6Y1)

I **recettori olfattivi (ORs)** sono espressi nelle *membrane cellulari dei neuroni dei recettori olfattivi* e sono responsabili della rilevazione di odori che danno origine al *senso dell'olfatto*. I *recettori olfattivi* attivati innescano gli impulsi nervosi che trasmettono informazioni sull'odore al cervello. Questi recettori sono membri della *famiglia dei recettori accoppiati a proteine G (GPCR) di classe A simile alla rodopsina*. I *recettori olfattivi* formano una famiglia multigenica composta da circa 800 geni nell'uomo e 1400 geni nei topi (https://en.wikipedia.org/wiki/Olfactory_receptor). La **sottofamiglia dei recettori olfattivi 6B e le proteine correlate sono membri della famiglia di classe A dei recettori accoppiati a proteine G a sette transmembrana (7tmA_OR6B-like)** e include i recettori olfattivi umani 6B, 6A, **6Y**, 6P e proteine correlate in altri mammiferi, sauropsidi e anfibi. Più del 60% degli OR umani sono *pseudogeni non funzionali* rispetto a solo il 20% nel topo. Ogni OR può riconoscere odori strutturalmente simili e un singolo odore può essere rilevato da più OR. Il legame di un odorizzante al recettore olfattivo induce un *cambiamento conformazionale* che porta all'attivazione della **proteina G olfattiva specifica (Golf)**. La *proteina G* (Golf e/o Gs) a sua volta stimola l'*adenilato ciclasi* a produrre cAMP (<https://www.ncbi.nlm.nih.gov/Structure/cdd/cddsrv.cgi?uid=320352>). Il cAMP apre canali ionici ciclici guidati dai nucleotidi, che consentono l'afflusso di ioni calcio e sodio, con conseguente depolarizzazione del neurone recettore olfattivo e innesco di un potenziale d'azione che trasmette queste informazioni al cervello. Un sistema di nomenclatura di consenso basato sulla divergenza evolutiva viene qui utilizzato per classificare la famiglia dei recettori olfattivi. La nomenclatura inizia con il nome della radice OR, seguito da un numero intero che rappresenta una famiglia, una lettera che indica una sottofamiglia e un numero intero che rappresenta il gene individuale all'interno della sottofamiglia (<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](#)

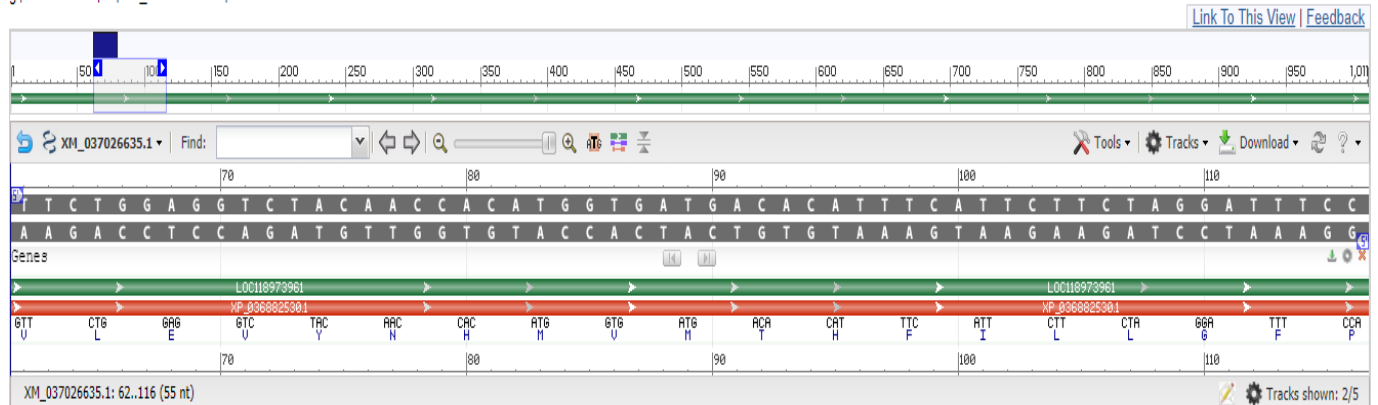
Score	Expect	Identities	Positives	Gaps
26.1 bits(54)	467	6/6(100%)	6/6(100%)	0/6(0%)

```

Query    1      MVMTHF    6
          MVMTHF
Sbjct   28      MVMTHF    33
  
```

PREDICTED: *Manis javanica* olfactory receptor 6Y1-like (LOC118973961), mRNA

gj11926243006[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	MVM THF 6			
	MVMTHF			
Sbjct 32	MVM THF 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	MVM THF 6			
	MVM HF			
Sbjct 362	MVM AHF 367			

Il dominio F-BAR

Il **dominio F-BAR** sta emergendo come un attore importante nei percorsi di *rimodellamento della membrana*. Le proteine del dominio F-BAR accoppiano il rimodellamento della membrana con la dinamica dell'**actina** associata alle *vie endocitiche* e alla formazione di *filopodio* (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2889966/>). Le proteine contenenti il dominio F-BAR possono essere classificate in *cinque sottofamiglie* in base alla loro filogenesi che è coerente con i domini proteici aggiuntivi che possiedono, ad esempio *domini RhoGAP*, *siti di legame Cdc42*, *domini SH3* e *domini tirosin chinasi*. *Dynamin1 / 2*, *N-WASP*, *Huntingtin*, *intersectin* e *Cdc42* sono nodi centrali che influenzano la funzione della proteina del dominio F-BAR (<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	MVM THF 6			
	MVM HF			
Sbjct 1	MVM AHF 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	MVM THF 6			
	MVM HF			
Sbjct 1	MVM AHF 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	MVM THF 6			
	MVM HF			
Sbjct 1	MVM AHF 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	MVM THF 6			
	MVM HF			
Sbjct 1	MVM AHF 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	MVM THF 6			
	MVM HF			
Sbjct 1	MVM AHF 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	MVM THF 6			
	MVM HF			
Sbjct 1	MVM AHF 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		MVM THF 6		
		MVM HF		
Sbjct 1		MVM AHF 6		

I risultati della ricerca BLAST sulla sequenza **M V M T H F** hanno evidenziato allineamenti significativi con i **pagolini** (*Manis pentadactyla* e *javanica*) e con i **pipistrelli** (*Rhinolophus ferrumequinum* e *Pipistrellus kuhlii*). Questo risultato acquista rilevanza per **due motivi**.

Il **primo motivo** è che la sequenza di aminoacidi **M V M T H F**, come sappiamo, ha dei legami diretti con la sequenza di aminoacidi **M A M T H** dalla quale deriva [per sostituzione dell'aminoacido di *Alanina* (A) con uno di *Valina* (V) e per l'aggiunta finale di un aminoacido di *Fenilalanina* (F)] e con la sequenza di basi "ATG GCA ATG ACA CAC" ad essa corrispondente. Questa sequenza di basi è parte di un segmento (**GGCAATGACACACAAGA**) della **Sequenza n° 1/3** (**CTCGGGCAATGACACACAAGAA**), sequenza identificata come corrispondente al *Virus dell'HIV-1* [regione "misc_feature" - **nonfunctional envelope glycoprotein due to mutation** del gene Env] e ottenuta dalla modificazione (seguendo il "Trend Non Manifesto" n°1) della sequenza di RNA "GTGAA **ATG GTC ATG TGT GGC** GG" (corrispondente, in un suo segmento, alla sequenza di aminoacidi **M V M C G**) tratta dal **Gene RdRp** del SARS-CoV-2.

Il **secondo motivo** per il quale questo risultato acquista rilevanza è determinato dalla **tipologia** di alcune sequenze riscontrate e riportate in questo paragrafo. In questo caso mi riferisco a **olfactory receptor 6Y1-like** [**Manis pentadactyla**] e [**Manis javanica**], un tipo di recettore olfattivo comune anche agli esseri umani. Dai resoconti clinici di molte persone che si sono ammalate della **Covid-19** emerge un coinvolgimento temporaneo della perdita dell'olfatto. Naturalmente, non ho alcuna prova dell'eventuale coinvolgimento di questo recettore olfattivo nel disturbo lamentato dai pazienti ammalati della Covid-19, ma ho considerato utile segnalare questa "coincidenza" per mettere altri ricercatori interessati in condizioni di indagare ulteriormente.

9.9 SEQUENZA DI AMINOACIDI “M A M T H K”

In questo paragrafo sono mostrati alcuni risultati della ricerca BLAST effettuata sulla sequenza di aminoacidi **M A M T H K**, la sequenza già presa in considerazione nel *paragrafo 9.5* e a cui è stato aggiunto l'aminoacido **Lisina (K)** determinato dal codone successivo, il *codone AAG*.

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	Pipistrellus kuhlii	21.8	21.8	100%	15143	83.33%	2762

	mPipKuh1_008369							
	[Pipistrellus kuhlii]							
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

Una *pseudouridilato sintasi* è un enzima che catalizza una reazione chimica in cui sono coinvolti l'Uracile e il *D-ribosio 5-fosfato*. Questo enzima appartiene alla famiglia delle **liasi** (un enzima che catalizza la **rottura di vari legami chimici** mediante mezzi diversi dall'*idrolisi* e dall'*ossidazione*), in particolare le **idro-liasi**, che scindono i legami carbonio-ossigeno. Questo enzima partecipa al *metabolismo della pirimidina* (https://en.wikipedia.org/wiki/Pseudouridylate_synthase).

S9 family peptidase [Alteromonadaceae bacterium]

Sequence ID: [MAJ70411.1](#) Length: 804 Number 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

```

Peptidasi e famiglia della peptidasi S9

Una **peptidasi** (chiamata anche **proteasi** o **proteinas**) è un enzima che catalizza (aumenta il tasso di) *proteolisi*, la scomposizione delle proteine in polipeptidi più piccoli o singoli amminoacidi. Ottiene questo risultato scindendo i *legami peptidici* all'interno delle proteine mediante **idrolisi**, una reazione in cui l'**acqua** rompe i legami (<https://en.wikipedia.org/wiki/Protease>). Le **proteasi** sono presenti in tutti gli organismi, dai procarioti agli eucarioti ai virus. Questi enzimi sono coinvolti in una moltitudine di reazioni fisiologiche dalla semplice digestione delle proteine alimentari alle cascate altamente regolate (es., la cascata della **coagulazione del sangue**, il *sistema del complemento*, le *vie dell'apoptosi* e la cascata di attivazione della *profenolossidasi invertebrata*). L'attività può essere un *cambiamento distruttivo* (abolire la funzione di una proteina o digerirla nei suoi componenti principali), può essere un'attivazione di una *funzione*, oppure può essere un *segnale in una via di segnalazione*. Le proteasi determinano la *durata di altre proteine* che svolgono un ruolo fisiologico importante come *ormoni*, *anticorpi* o altri *enzimi*. Questo è uno dei meccanismi regolatori più veloci di "accensione" e "spegnimento" nella fisiologia di un organismo. **Mediante un'azione cooperativa complessa le proteasi possono procedere come reazioni a cascata, che si traducono in un'amplificazione rapida ed efficiente della risposta di un organismo a un segnale fisiologico** (<https://en.wikipedia.org/wiki/Protease>). I batteri contengono proteasi responsabili del controllo generale della qualità delle proteine *degradando le proteine dispiegate o mal ripiegate*. Una *proteasi batterica secreta* può anche agire come **esotossina** ed essere un esempio di un **fattore di virulenza nella patogenesi batterica**. I genomi di alcuni **virus** codificano una massiccia poliproteina, che necessita di una proteasi per dividerla in unità funzionali. Queste proteasi hanno un'elevata specificità e scindono solo un insieme molto limitato di sequenze di substrato (<https://en.wikipedia.org/wiki/Protease>). Sono quindi un bersaglio comune per gli **inibitori della proteasi** ([https://en.wikipedia.org/wiki/Protease_inhibitor_\(pharmacology\)](https://en.wikipedia.org/wiki/Protease_inhibitor_(pharmacology))), una classe di **farmaci antivirali** che sono ampiamente utilizzati per il trattamento di **HIV/AIDS** e l'**epatite C**. Gli **inibitori della proteasi** prevengono la **replicazione virale** *legandosi selettivamente alle proteasi virali* (es. *Proteasi dell'HIV-1*) e *bloccando la scissione proteolitica* dei precursori proteici necessari per la produzione di particelle virali infettive ([https://en.wikipedia.org/wiki/Protease_inhibitor_\(pharmacology\)](https://en.wikipedia.org/wiki/Protease_inhibitor_(pharmacology))). Le proteasi esotossiche batteriche distruggono le strutture extracellulari. Le proteasi sono state raggruppate per la prima volta in **84 famiglie** in base alla loro relazione evolutiva e classificate in **quattro tipi catalitici: serina, cisteina, aspartica e metallo proteasi** (<https://en.wikipedia.org/wiki/Protease>). Per gli interessi di studio trattati in questo documento, la **Proteasi a Serina** (la *Serina* è un aminoacido che può essere destinato alla produzione di energia metabolica) è quella più importante. La **famiglia della peptidasi S9** contiene una variegata serie di **peptidasi serina-dipendenti**. Si ritiene che molte peptidasi della famiglia siano importanti per la **degradazione dei peptidi biologicamente attivi**. I membri della famiglia si trovano in *diversi siti cellulari*: l'*oligopeptidasi proliovola* nella proteina intracellulare, *DDP-IV* e di *attivazione dei fibroblasti α* sono espressi sulle superfici cellulari e l'*oligopeptidasi B* viene secreta nel periplasma batterico (<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-serina ammoniacaliasi

L-serina ammoniacaliasi è un enzima che catalizza una reazione chimica che ha un substrato, **L-serina**, e due prodotti, *piruvato* e NH_3 . Questo enzima appartiene alla famiglia delle **liasi**, in particolare *ammoniacaliasi*, che scindono i legami *carbonio-azoto*. Questo enzima partecipa al *metabolismo* della *glicina*, della *serina* e della *treonina* e al *metabolismo della cisteina*. (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
```

L'adenosina deaminasi

L'adenosina deaminasi (o **adenosina amminoidrolasi** o **ADA**) è un enzima coinvolto nel *metabolismo delle purine*. È necessario per la scomposizione dell'adenosina dal cibo e per il ricambio degli acidi nucleici nei tessuti. La sua funzione principale nell'uomo è lo sviluppo e il *mantenimento del sistema immunitario*, ma il ruolo fisiologico completo dell' ADA non è ancora completamente compreso. L'enzima è stato trovato in *batteri*, *piante*, *invertebrati*, *vertebrati* e *mammiferi*, con un'elevata conservazione della sequenza amminoacidica. Lo **zinco** è l'unico cofattore necessario per l'attività. **Livelli elevati di ADA** sono stati associati all'**AIDS**. **Livelli carenti di ADA** sono però stati associati a *infiammazione polmonare*, *morte delle cellule timiche* e *segnalazione difettosa del recettore delle cellule T*. (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
```

Alfa/beta idrolasi

La *superfamiglia alfa/beta idrolasi* è una superfamiglia di **enzimi idrolitici** di origine filogenetica molto diversa e funzione catalitica che condivide una *piega comune* che comprende *proteasi*, *lipasi*, *perossidasi*, *esterasi*, *epossido idrolasi* e *dalogenasi*. Si ritiene che gli enzimi siano divergenti da un antenato comune, mantenendo una *somiglianza di sequenza poco evidente*, ma *preservando la disposizione dei residui catalitici*. Tutti hanno una *triade catalitica*, i cui elementi sono portati su anelli, che sono le caratteristiche strutturali meglio conservate della *piega* (https://en.wikipedia.org/wiki/Alpha/beta_hydrolase_superfamily).

DNA helicase PcrA [**Bacillus cereus**]

Sequence ID: [WP_098544546.1](#) Length: 743 Number 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
```

PcrA DNA Elicasi

Copia plasmidica ridotta (PcrA) è un'elicasi (enzimi vitali coinvolti nella replicazione del DNA , nella trascrizione , nella traduzione , nella ricombinazione, nella riparazione del DNA e nella biogenesi dei ribosomi, tutti processi che implicano la separazione dei filamenti di acido nucleico). L'elicasi è essenziale per la *replicazione del cerchio rotante del plasmide* (processo di replicazione unidirezionale degli acidi nucleici che può sintetizzare rapidamente più copie di molecole circolari di DNA o RNA , come i plasmidi , i genomi dei batteriofagi e il genoma dell'RNA circolare dei dei viroidi, https://en.wikipedia.org/wiki/Rolling_circle_replication) e la *riparazione del danno al DNA* causato dall'esposizione alle radiazioni ultraviolette. L'elicasi catalizza lo svolgimento del DNA plasmidico a doppio filamento che è stato intaccato all'origine della replicazione dalla *proteina di inizio della replicazione*. L'elicasi gioca **un ruolo importante nella sopravvivenza cellulare regolando i livelli di ricombinazione mediata da RecA nei batteri Gram-positivi** (<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**” è una **proteina del dito C2H2 Zn strutturalmente modificata**. Le dita di zinco sono **regioni proteiche specifiche in grado di legare il DNA**. La “**zinc finger protein 131-like isoform X4**” sopra identificata è una **proteina C2H2 Zn finger strutturalmente modificata (Dito di zinco, tipo C2H2)**. Il **dito di zinco C2H2** è un classico dominio delle dita di zinco. Le dita di zinco di tipo C2H2 sono onnipresenti; si prevede che più dell'1% di tutte le proteine dei mammiferi contenga almeno un dito di zinco. Spesso funzionano come *motivi strutturali di legame del DNA o delle proteine*, come nei *fattori di trascrizione eucariotica*, e quindi svolgono ruoli importanti nei **processi cellulari** come lo **sviluppo**, la **differenziazione** e l'**oncosoppressione**. Le proteine delle dita di zinco C2H2 contengono da 1 a più di 30 ripetizioni di dita di zinco (<https://www.ncbi.nlm.nih.gov/Structure/cdd/cddsrv.cgi?uid=275371>). Ci sono **molte superfamiglie di motivi Znf**, che variano sia nella *sequenza* che nella *struttura*. Mostrano una notevole versatilità nelle modalità di legame, anche tra membri della stessa classe (ad esempio, alcuni legano il DNA, altri la proteina), suggerendo che *i motivi Znf* sono **scaffold stabili che hanno sviluppato funzioni specializzate**. Ad esempio, le *proteine contenenti Znf* funzionano nella **trascrizione genica, traduzione, traffico di mRNA, organizzazione del citoscheletro, sviluppo epiteliale, adesione cellulare, ripiegamento delle proteine, rimodellamento della cromatina e rilevamento dello zinco**, solo per citarne alcuni.

I motivi che legano lo zinco sono strutture stabili e raramente subiscono cambiamenti conformazionali quando legano il loro bersaglio (<http://www.ebi.ac.uk/interpro/entry/InterPro/IPR036236/>).

Zinc finger protein 131 svolge un ruolo durante lo **sviluppo** e l'**organogenesi**, nonché nella *funzione del sistema nervoso centrale adulto*. Può essere coinvolto nella *regolazione trascrizionale* come **repressore della segnalazione di ESR1/ER-alfa** (<https://www.ebi.ac.uk/interpro/entry/InterPro/IPR027758/>), meccanismo di repressione trascrizionale indotta da estrogeni utilizzando il **gene Era** (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2738295/>). E' stato dimostrato che gli **ZNF** potrebbero anche agire come *reclutatori di modificatori della cromatina*, come *cofattori* o come *proteine strutturali coinvolte nella migrazione e invasione cellulare*. In particolare, il ruolo degli **ZNF** nello **sviluppo, nella progressione e nelle metastasi del cancro** sta diventando un argomento di ricerca interessante (<https://www.nature.com/articles/cddiscovery201771>). **Zinc finger C2H2** può avere una varietà di funzioni come legare l'RNA e mediare le interazioni proteina-proteina, ma è meglio conosciuta per il suo ruolo in *proteine leganti il DNA di specifiche sequenze* come **Zif268 (Egr1)** (https://en.wikipedia.org/wiki/Zinc_finger). **EGR-1** (Early growth response protein 1) nota anche come **ZNF268 (zinc finger protein 268)** o **NGFI-A** (proteina A indotta dal fattore di crescita nervoso) è una *proteina* che negli esseri umani è codificata dal **gene EGR1**. **Egr1** è una *proteina nucleare e funziona come un regolatore trascrizionale*, soprattutto cerebrale. Oltre alla sua funzione nel *sistema nervoso*, ci sono prove significative che **EGR-1** insieme al suo paralog **EGR-2** è **indotto nelle malattie fibrotiche** (compresa la **fibrosi polmonare**) ha funzioni chiave nella fibrinogenesi, ed è **necessario per la fibrosi indotta sperimentalmente nei topi** (<https://en.wikipedia.org/wiki/EGR1>). La natura modulare del motivo del dito di zinco consente di legare un gran numero di **combinazioni di sequenze di DNA e RNA con un alto grado di affinità e specificità**, ed è quindi **ideale per proteine di ingegneria che possono essere mirate e legare sequenze di DNA specifiche** (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
```

Le glicosiltransferasi (GTF, Gtfs)

Le **GTF** sono enzimi che stabiliscono **legami glicosidici naturali**. Essi catalizzano il trasferimento da un *donatore di glicosile* ad *accettore di glicosile*. Le **GTF** sono state ampiamente utilizzate sia nella sintesi mirata di *glicoconiugati specifici* che nella *sintesi di librerie di farmaci*, sonde biologiche o prodotti naturali glicosilati differenzialmente nel contesto della scoperta e dello sviluppo di farmaci. Enzimi adatti possono essere isolati da fonti naturali o prodotti in modo ricombinante (<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](https://pubmed.ncbi.nlm.nih.gov/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, Pfothenhauerstrasse 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 :: mPipKuhl.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="mPipKuhl"

/db_xref="taxon:[59472](https://taxon.ncbi.nlm.nih.gov/taxonomy/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](https://www.openstreetmap.org/way/123456789)"

/collection date="2017"

/collected_by="Emma Teeling, Andrea Locatelli"

Protein <1..>18

/product="hypothetical protein"

CDS 1..2762

/locus tag="mPipKuhl 008369"

/coded_by="complement(join(JACAGB010000014.1:2240776..2240894,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 **mams**hk

//

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 MAM**T**HK 6
 MAM HK

Sbjct 1976 MAM**S**HK 1981

Tireoglobulina

La **Tireoglobulina (Tg)** è una **dimero glicoproteina** prodotta dalle cellule follicolari della tiroide ed è utilizzata interamente all'interno della *ghiandola tiroidea*. La tireoglobulina è in tutti i vertebrati il principale precursore degli **ormoni tiroidei**, che vengono prodotti quando i residui di **tirosina** della *tireoglobulina* vengono combinati con lo **iodio** e la proteina viene successivamente scissa (<https://en.wikipedia.org/wiki/Thyroglobulin>). È stato dimostrato che la *tireoglobulina* interagisce con la **proteina legante la catena dell'immunoglobulina (BiP)**. **BiP** è situata nel lume del **reticolo endoplasmatico (ER)**, è un componente essenziale del **meccanismo di traslocazione** e svolge un ruolo nel **trasporto retrogrado attraverso la membrana ER di proteine aberranti destinate alla degradazione da parte del proteasoma**. La **BiP** è una proteina abbondante in tutte le condizioni di crescita, ma la sua sintesi è marcatamente indotta in condizioni che portano all'accumulo di **polipeptidi dispiegati nell'ER** (<https://pubmed.ncbi.nlm.nih.gov/10597629/>). BiP ha una potente **attività immunologica**, modula alcune **malattie cardiovascolari** e **neurodegenerative**, migliora il **metabolismo**, è coinvolta nella **replicazione del DNA batterico** e quindi può essere utile per lo sviluppo di antibiotici. Al contrario, **i virus si basano sul BiP ospite per replicarsi con successo**, in gran parte **infettando le cellule attraverso il BiP sulla superficie cellulare**, stimolando l'espressione del **BiP nelle proteine virali chaperone** e **sopprimendo la risposta alla morte da stress nell'ER** (https://en.wikipedia.org/wiki/Binding_immunoglobulin_protein).

Dalle informazioni riportate in questo paragrafo, un aspetto importante determinato dall'aggiunta dell'**aminoacido Lisina (K)** alla sequenza **M A M T H** sembra riguardare il coinvolgimento di **enzimi** che svolgono **funzioni catalitiche**, in particolare la famiglia delle **liasi** (es., **L-Serina**), l'**adenosina deaminasi**, le **peptidasi** (o **proteasi**), le **idrolasi**, le **glicosiltransferasi**, ovvero molti enzimi della **superfamiglia alfa/beta idrolasi**. Appare evidente anche l'importanza di alcune **proteine** implicate nei processi di **replicazione del DNA**, nella **trascrizione**, **ricombinazione** e **riparazione del DNA** e nella **biogenesi dei ribosomi**, tutti processi che prevedono l'azione delle **elicasi** e implicano la separazione dei filamenti di acido nucleico, nonché nei **meccanismi di traslocazione**.

Da segnalare i motivi a **dito di zinco (zinc finger)** tra gli **allineamenti significativi** più volte riscontrati in questo e anche in altri capitoli successivi. Si tratta di proteine che, per la loro natura modulare, consentono di legare un gran numero di **combinazioni di sequenze di DNA e RNA con un alto grado di affinità e specificità**, quindi **ideali per proteine di ingegneria che possono essere mirate e legare sequenze di DNA specifiche** (https://en.wikipedia.org/wiki/Zinc_finger).

Sempre con riferimento ai motivi a **dito di zinco**, è molto importante sottolineare che essi sono stati identificati anche tra gli **allineamenti significativi** [*Egretta garzetta zinc finger protein 639 (ZNF639), transcript variant X1, X2 and X3, mRNA*] della sequenza di RNA **"GTGAAATGGTTATGTGTGGCGG"** del *RdRp Gene del "Bat coronavirus BM48-31/BGR/2008, complete genome"*. Questa sequenza è importante perché essa è quasi totalmente identica alla sequenza **"GTGAAATGGTCATGTGTGGCGG"**, la sequenza di RNA oggetto di studio di questo documento. Inoltre, come già visto nel Capitolo III in Fig. 3.1 (della quale riporto il quadrante interessato nella pagina successiva), la sequenza **"GTGAAATGGTTATGTGTGGCGG"** (nel quadrante, sequenza indicata dalla freccia blu) è una delle due sequenze del **gene RdRp (RNA polimerasi RNA dipendente)**, il motore del processo di **trascrizione del SARS-CoV-2** i cui allineamenti significativi delle ricerche BLAST identificano il **Pipistrello** e il **Pangolino (Malese e Cinese)**.

RdRP gene



In aggiunta a questi “organismi”, l’altro “organismo” identificato è appunto la **Piccola Egretta**, proprio con gli allineamenti significativi *Egretta garzetta zinc finger protein 639* (ZNF639), *transcript variant X1, X2 and X3, mRNA*. Adesso dovrebbe apparire evidente che la presenza frequente della **zinc finger protein** tra gli allineamenti significativi di questo capitolo (e anche degli altri) non può certamente considerarsi casuale, ma determinata dall’azione del “Trend Non Manifesto” della **Sequenza n° 1/3** (e di altre **sequenze modificate**).

9.10 SEQUENZA DI AMINOACIDI “M A M T H K S Q”

In questo paragrafo sono mostrati alcuni risultati della ricerca BLAST effettuata sulla sequenza di aminoacidi **M A M T H K S Q**, la sequenza già presa in considerazione nel *paragrafo precedente* e a cui sono stati aggiunti gli aminoacidi **Serina (S)** e **Glutamina (Q)**, determinati rispettivamente dai *codoni* AGT e CAA. Le basi di questi ultimi due codoni non appartengono alla **Sequenza 1/3** (la sequenza ottenuta dalla modificazione della sequenza di RNA oggetto di studio di questo documento), ma alla sequenza dell'**HIV-1** identificata dalla **Sequenza 1/3** e oggetto di studio approfondito di questi ultimi paragrafi. Per coerenza, si è scelto di continuare a studiare i codoni successivi fino ad incontrare il codone di STOP.

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 FigB [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	MAMTHK S Q 8			
	MAMTHK Q			
Sbjct 133	MAMTHK A Q 140			

Peptidasi e famiglia della peptidasi S9

Si veda quanto discusso a pagina 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	M A MTHKSQ 8			
	M MTHKSQ			
Sbjct 269	M T MTHKSQ 276			

Le *E-set dell'immunoglobulina* sono proteine simili alle **immunoglobuline (Ig)** molto comuni negli animali. I domini dell'insieme "precoce" (**E-set**) sono probabilmente correlati alle superfamiglie di *proteine Ig-like dell'immunoglobulina* ([IPR007110](#)) e/o della *fibronectina tipo III* ([IPR003961](#)). I *domini Ig-like E-set* includono anche parte della superfamiglia di *enzimi* chiamati **O-Glicosil idrolasi (3.2.1.)** [[\(https://www.ebi.ac.uk/interpro/entry/InterPro/IPR014756/\)](https://www.ebi.ac.uk/interpro/entry/InterPro/IPR014756/)]; [[\(https://www.ebi.ac.uk/interpro/entry/InterPro/IPR004185/\)](https://www.ebi.ac.uk/interpro/entry/InterPro/IPR004185/)]; si veda anche quanto scritto a pagina 230 sulla *superfamiglia alfa/beta idrolasi* e a pagina 233 sulle *glicosiltransferasi*] e la famiglia di *proteine Arrestin*. Queste ultime proteine sono importanti per *regolare la trasduzione del segnale nei recettori accoppiati a proteine G*. La **Sequenza ID: [KRX08649.1](#)** "Immunoglobulin E-set" identificata sopra si riferisce a una parte di una sequenza di "*Arrestin (or S-antigen), N-terminal domain*". Gli **Arrestini** sono stati scoperti per la prima volta come parte di un meccanismo conservato in due fasi per *regolare l'attività dei recettori accoppiati a proteine G (GPCR) nella rodopsina visiva* (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
```

Proteina flagellare del corpo basale, N-terminale

Molte *specie batteriche* nuotano attivamente per mezzo di **flagelli**. Un *flagello di tipo batterico* è un complesso motore composto da un filamento proteico elicoidale extracellulare accoppiato a un motore rotante incorporato nell'involucro cellulare (<https://www.ebi.ac.uk/QuickGO/GTerm?id=GO:0071973>). Una delle parti dell'*organello flagello* è il *corpo basale*. Nella *Salmonella typhimurium* e nei relativi organismi è stato dimostrato che il *bastoncino* è costituito da **quattro correlate diverse proteine evolutive**: nella *porzione distale* del bastoncino ci sono circa 26 subunità di **proteina flgG** e nella *porzione prossimale* ci sono circa sei subunità ciascuna di **proteine flgB, flgC e flgF**. Queste **quattro proteine** contengono un *dominio ricco di asparagina altamente conservato al loro terminale N* (<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-serina ammoniaca-liasi

Si veda quanto discusso a pagina 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

Proteina D del gene di controllo tardivo dei fagi (GPD)

Questa famiglia include una *serie di proteine del gene D del controllo tardivo dei fagi* (o *batteriofagi*) e sequenze batteriche correlate. Questa famiglia comprende anche le *proteine Mu P del batteriofago* e le relative sequenze (<https://www.ncbi.nlm.nih.gov/Structure/cdd/PF05954>). Il *batteriofago* è un *virus* che infetta e si replica all'interno dei *batteri* e degli *archeobatteri* (<https://en.wikipedia.org/wiki/Bacteriophage>). I **fagi** sono stati utilizzati come alternativa agli antibiotici e sono visti come una possibile terapia contro ceppi multi-farmaco-resistenti di molti batteri e **utili per contrastare le armi biologiche e le tossine**, come l'**antrace** e il **botulismo**. **I ricercatori sui batteriofagi hanno sviluppato virus ingegnerizzati per superare la resistenza agli antibiotici ingegnerizzando i geni dei fagi responsabili della codifica degli enzimi che degradano la matrice del biofilm**, le proteine strutturali dei fagi e gli enzimi responsabili della lisi della parete cellulare batterica. D'altra parte, è stato dimostrato che alcune specie di fagi complicano i biofilm coinvolti nella *polmonite* e nella *fibrosi cistica* e *proteggono i batteri dai farmaci destinati a sradicare la malattia, promuovendo così un'infezione persistente* (<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 **M**AMTHKSQ 8

M MTHKSQ

Sbjct 548 **M**DMTHKSQ 555

CRISPR/Cas system-associated protein Cas8a1

CRISPR (*Clustered Regularly Interspaced Short Palindromic Repeats*) e le **proteine Cas associate** costituiscono un **sistema per la difesa dell'ospite** ereditabile da parte di cellule procariotiche contro **fago** e altro **DNA estraneo**; grandi proteine, alcune contengono il dominio **Zn-finger**; gene di firma per il sottotipo I-A; conosciuta anche come famiglia TM1802. (<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	MAMTHK K SQ 8			
	MAMTH SQ			
Sbjct 310	MAMTH T SQ 317			

Neuraminidasi (sialidasi)

Le **Neuraminidasi** sono enzimi **idrolasi glicoside** che fendono (tagliano) i legami glicosidici degli *acidi neuraminici*. Gli *enzimi neuraminidasi* sono una *grande famiglia*, presente in una vasta gamma di organismi. La neuraminidasi più nota è la **neuraminidasi virale**, un farmaco bersaglio per la prevenzione della diffusione dell'**infezione influenzale**. Le *neuraminidasi virali* sono frequentemente utilizzate come determinanti **antigenici** presenti sulla **superficie del virus dell'influenza**. **Alcune varianti della neuraminidasi influenzale conferiscono più virulenza al virus rispetto ad altri. Le sialidasi possono agire come fattori patogeni nelle infezioni microbiche** (<https://en.wikipedia.org/wiki/Neuraminidase>). La *Neuraminidasi virale* è un tipo di *neuraminidasi* presente sulla *superficie dei virus dell'influenza* che consente il rilascio del virus dalla cellula ospite. Le *neuraminidasi* sono enzimi che **scindono i gruppi di acido sialico** (chiamato anche *acido neuraminico*) dalle **glicoproteine**. Gli **inibitori della neuraminidasi** sono **agenti antivirali che inibiscono l'attività della neuraminidasi virale dell'influenza e sono di grande importanza nel controllo dell'influenza** (https://en.wikipedia.org/wiki/Viral_neuraminidase). Le *neuraminidasi virali* sono i membri della famiglia dei **glicosidi idrolasi 34 (CAZY GH_34)** che comprende enzimi con una sola attività nota: **sialidasi** o **neuraminidasi (EC 3.2.1.1)**. Le *neuraminidasi* scindono i residui terminali di acido sialico dalle catene di carboidrati nelle *glicoproteine*. L'**acido sialico** è uno zucchero a carica negativa associato alle porzioni proteiche e lipidiche delle lipoproteine (https://en.wikipedia.org/wiki/Viral_neuraminidase). **Per infettare una cellula ospite, il virus dell'influenza si attacca alla superficie cellulare esterna utilizzando l'emoagglutina, una molecola che si trova sulla superficie del virus che si lega ai gruppi di acido sialico.** Gli **acidi sialici** si trovano su varie *glicoproteine* sulla superficie della cellula ospite. Il *virus* si sposta quindi da un *gruppo acido sialico* ad un altro *gruppo acido sialico* fino a quando non trova il **corretto recettore della superficie cellulare** (la cui identità rimane sconosciuta). La *neuraminidasi* consente questo movimento *scindendo i gruppi di acido sialico a cui era attaccata l'emoagglutina*. **Dopo che il virus è entrato nella cellula e si è replicato, nuove particelle virali germogliano dalla membrana della cellula ospite** (https://en.wikipedia.org/wiki/Viral_neuraminidase). L'**emoagglutina** sulle nuove particelle virali rimane attaccata ai gruppi di glicoproteine dell'acido sialico sulla superficie cellulare esterna e sulla superficie di altre particelle virali; la *neuraminidasi* scinde questi gruppi e quindi consente il rilascio di

particelle virali e previene l'autoaggregazione (https://en.wikipedia.org/wiki/Viral_neuraminidase). Gli **inibitori della neuraminidasi (NAI)** sono una classe di farmaci che bloccano l'enzima *neuraminidasi*. Sono comunemente usati come **farmaci antivirali** perché bloccano la funzione delle neuraminidasi virali del virus dell'influenza, impedendone la riproduzione per gemmazione dalla cellula ospite (https://en.wikipedia.org/wiki/Neuraminidase_inhibitor). *Oseltamivir (Tamiflu)*, *Zanamivir (Relenza)*, *Laninamivir (Inavir)* e *Peramivir* appartengono a questa classe di sostanze (https://en.wikipedia.org/wiki/Viral_neuraminidase). Dopo la **pandemia** causata da **H1N1** nel 2009 (il **virus dell'influenza A sottotipo H1N1** è un sottotipo di **virus influenzale A**. I focolai ben noti di ceppi **H1N1** negli esseri umani includono la **pandemia di influenza suina** del 2009, la **pandemia di influenza russa** del 1977 e la **pandemia di influenza** del 1918), l'efficacia del **trattamento precoce con inibitori della neuraminidasi** nel ridurre i casi gravi e le morti è stata segnalata in vari paesi (https://en.wikipedia.org/wiki/Influenza_A_virus_subtype_H1N1; https://en.wikipedia.org/wiki/Neuraminidase_inhibitor). Il **virus dell'influenza A** causa l'influenza negli **uccelli** e in alcuni **mammiferi** ed è l'unica specie del genere *Alphainfluenzavirus* della famiglia dei virus *Orthomyxoviridae*. Ceppi di tutti i sottotipi di **virus dell'influenza A** sono stati isolati dagli **uccelli selvatici**, sebbene la malattia sia rara. Alcuni isolati del **virus dell'influenza A** causano gravi malattie sia nel **pollame domestico** che, raramente, nell'**uomo**. Occasionalmente, i virus vengono trasmessi dagli **uccelli acquatici selvatici** al **pollame domestico** e ciò può causare un'**epidemia** o dare origine a **pandemie influenzali umane** (https://en.wikipedia.org/wiki/Influenza_A_virus). I **virus dell'Influenza A** sono *negativi-sense* e sono virus a RNA a filamento singolo e segmentato. I **vari sottotipi** sono etichettati secondo un **numero H** (per il tipo di emoagglutinina) e un **numero N** (per il tipo di neuraminidasi). Esistono 18 diversi antigeni H noti (da H1 a H18) e 11 diversi antigeni N noti (da N1 a N11). **H17N10** è stato isolato dai **pipistrelli della frutta** nel 2012. **H18N11** è stato scoperto in un **pipistrello peruviano** nel 2013 (https://en.wikipedia.org/wiki/Influenza_A_virus). Ciascun sottotipo di virus è mutato in una varietà di ceppi con differenti profili patogeni; alcuni sono patogeni per una specie ma non per altre, alcuni sono patogeni per più specie. È stato sviluppato un **vaccino contro l'influenza A** filtrato e purificato per l'uomo e molti paesi lo hanno accumulato per consentire una rapida somministrazione alla popolazione in caso di **pandemia di influenza aviaria**. Nel 2011, i ricercatori hanno segnalato la **scoperta di un anticorpo efficace contro tutti i tipi di virus dell'influenza A**. Si tratta di **FI6**, un **anticorpo** che prende di mira una proteina presente sulla superficie di tutti i virus dell'influenza A chiamata *emoagglutinina*. **FI6** è l'unico anticorpo noto che si lega a tutti i 16 sottotipi dell'emoagglutinina del virus dell'influenza A e si spera che sia utile per una **terapia universale contro il virus dell'influenza** (https://en.wikipedia.org/wiki/Influenza_A_virus; [https://en.wikipedia.org/wiki/FI6_\(antibody\)](https://en.wikipedia.org/wiki/FI6_(antibody))).

L'aggiunta degli **aminoacidi Serina (S)** e **Glutamina (Q)** alla sequenza **M A M T H** conferma l'importanza, già evidenziata nel paragrafo precedente, degli **enzimi** che svolgono **funzioni catalitiche**, in particolare le **peptidasi** (o **proteasi**), **L-serina ammoniacaliasi** e la **O-Glicosil idrolasi**. Importante anche la presenza di **allineamenti significativi** che riguardano le **proteine del gene D del controllo tardivo dei fagi** (o **batteriofagi**), organismi quest'ultimi impiegati **nello sviluppo virus ingegnerizzati responsabili della codifica degli enzimi che degradano la matrice del biofilm dei batteri**. Tuttavia, in alcuni casi, è stato dimostrato anche il contrario, cioè che essi complicano i biofilm coinvolti nella **polmonite** e nella **fibrosi cistica** e **proteggono i batteri dai farmaci destinati a sradicare la malattia, promuovendo così un'infezione persistente**. Infine, importantissimo per le finalità del nostro studio, l'allineamento significativo con la **Neuraminidasi (Sialidasi)**, un enzima della famiglia **glicoside idrolasi** implicata nell'**Influenzavirus A**. Alcune varianti della neuraminidasi influenzale conferiscono più virulenza al virus rispetto ad altri. Le sialidasi possono agire come fattori patogeni nelle infezioni microbiche.

9.11 SEQUENZA DI AMINOACIDI “M A M T H K S Q M K”

In questo paragrafo sono mostrati alcuni risultati della ricerca BLAST effettuata sulla sequenza di aminoacidi **M A M T H K S Q M K**, la sequenza già presa in considerazione nel *paragrafo precedente* e a cui sono stati aggiunti gli aminoacidi **Metionina (M)** e **Lisina (K)**, determinati rispettivamente dai *codoni* ATG e AAA.

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: [EOB47646.1](#) 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

I membri di questa famiglia sono imparentati con *PIG-L* un *N-acetilglucosaminilfosfatidilinositolo de-N-acetilasi* che catalizza la *seconda fase della biosintesi di GPI* (<https://pfam.xfam.org/family/PF02585>). Deepak K. Sharm e Michael A. J. Ferguson spiegano che “Gli **ancoraggi di membrana del glicosilfosfatidilinositolo (GPI)** sono presenti negli organismi nella maggior parte delle fasi dell'evoluzione eucariotica, inclusi protozoi, lieviti, muffe melmose, invertebrati e vertebrati, e si trovano su una vasta gamma di *proteine*. Sono i principali responsabili dell’**ancoraggio delle proteine della superficie cellulare nella membrana plasmatica** (o in alcuni casi alla superficie lumenale topologicamente equivalente delle vescicole secretorie) e possono essere considerati un’alternativa all’ancoraggio polipeptidico transmembrana idrofobo della membrana integrale di tipo proteine 1. Molte altre funzioni sono state proposte (anche se alcune rimangono controverse) per gli **ancoraggi GPI**, compresi i ruoli nello **smistamento intracellulare**, nella **segnalazione transmembrana** e nei **nuovi processi endocitici**. La maggior parte di queste funzioni proposte dipendono dall'ipotesi che un **ancoraggio GPI** possa consentire alle proteine di associarsi in **microdomini di membrana specializzati** (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
```

Proteina simile a LmbE

S. Viars, J. Valentine e M. Hernick spiegano che “La **superfamiglia simile a LmbE** è composta da una **serie di enzimi** che utilizzano un *singolo ione metallico catalitico* per catalizzare l’**idrolisi** di vari *substrati*. Questi *substrati* sono spesso **metaboliti** chiave per eucarioti e procarioti, il che rende gli **enzimi simili a LmbE** obiettivi importanti per lo sviluppo di farmaci....questa è la più recente **superfamiglia di metalloidrolasi**....La maggior parte degli **enzimi** identificati fino ad oggi sono *deacetilasi metal-dipendenti* che catalizzano l’**idrolisi** di una *frazione N- acetilglucosamina* sul substrato utilizzando una combinazione di catene laterali di amminoacidi e un singolo ione metallico legato, prevalentemente **zinco**. Lo **zinco catalitico** è coordinato alle proteine tramite il suo sito di **legame His2-Asp**” (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4101496/>). I membri di questa famiglia includono N-acetilglucosaminil-fosfatidilinositolo de-N-acetilasi (3.5.1.89) che catalizza la seconda fase della **biosintesi del glicosilfosfatidilinositolo (GPI)** e *1D-mio-inositolo 2-acetamido-2-deossi-alfa-D-glucopiranoside deacetilasi* (3.5.1.103), coinvolto nella **biosintesi del micotio** (un insolito *composto tiolico* trovato negli **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
```

Canale del potassio MlotiK1

I **canali ionici ciclici regolati dai nucleotidi** (*nucleotide* del *Mesorhizobium loti*, MlotiK1) sono presenti in batteri, piante, vertebrati e esseri umani. Negli organismi superiori, sono strettamente coinvolti nelle **reti di segnalazione della vista** e dell’**olfatto**. Il legame di cAMP o cGMP favorisce l’attivazione di questi canali ionici (<https://www.pnas.org/content/108/51/20802>). Per approfondire lo studio del **ruolo svolto da questi canali ionici nelle funzioni visive e olfattive**, consultare questo interessante articolo: <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
```

RNA polimerasi N-terminale diretto dal DNA (pfam14700)

Questo è il *dominio N-terminale della RNA polimerasi diretta dal DNA*. Questo dominio ha un ruolo nell'interazione con le *regioni del DNA promotore a monte e la catena nascente dell'RNA*, portando alla **processività dell'enzima**. Al fine di produrre **trascritti di mRNA**, la **RNA polimerasi** subisce una *transizione* dalla *fase di inizio* (che produce solo brevi frammenti di RNA) a una *fase di allungamento*. Questo dominio subisce un **cambiamento strutturale** nella transizione dalla fase di inizio a quella di allungamento. Il cambiamento strutturale si traduce nell'**abolizione del sito di legame del promotore**, nella creazione di un **canale che accoglie l'eteroduplex nel sito attivo** e nella **formazione di un tunnel di uscita** attraverso il quale passa il **trascritto dell'RNA dopo il distacco dell'eteroduplex** (<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

L'**emeritina** è una **proteina** oligomerica responsabile del **trasporto di ossigeno** (O₂) nel phyla invertebrati marini di sipunculidi, priapulidi, brachiopodi e in un unico anellide genere di verme, *Magelona*. Prove recenti hanno rivelato che l'emeritina è una **proteina multifunzionale** e contribuisce all'**immunità innata** e alla **rigenerazione dei tessuti** anteriori nei vermi. Le **proteine emeritina** / **HHE** (H-HxxxE-HxxxH-HxxxxD) trovate nei **batteri** sono implicate nella **trasduzione del segnale** e nella **chemiotassi**. Quelle più lontane includono le proteine H-HxxxE-H-HxxxE (inclusa la **ligasi E3**) e le **proteine F-box animali** (H-HExxxE-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
```

Si veda quanto discusso a pagina 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
```

L'istidina chinasi (HK)

Le Istidine chinasi (HKs) sono proteine della **classe di enzimi transferasi** che svolgono un ruolo nella **trasduzione del segnale** attraverso la **membrana cellulare**. Le HKs possono agire come **recettori cellulari per la segnalazione di molecole** in modo analogo ai **recettori della tirosin chinasi (RTK)**. Le **molecole recettoriali multifunzionali come HK e RTK** hanno tipicamente porzioni all'*esterno della cellula* che si legano a *molecole simili a ormoni* o *fattori di crescita*, porzioni che si estendono *sulla membrana cellulare* e porzioni all'*interno della cellula* che prevedono l'attività enzimatica. Oltre all'**attività chinasica**, i domini intracellulari hanno tipicamente regioni che si legano a una molecola effettrice secondaria o a un complesso di molecole che *propagano ulteriormente la trasduzione del segnale all'interno della cellula*. **Questo tipo di enzima è coinvolto nelle vie di trasduzione del segnale a monte di molti processi cellulari tra cui varie vie metaboliche, virulenti e omeostatiche** (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			

famiglia di glicosidi idrolasi 65

La famiglia di glicosidi idrolasi 65 è una famiglia di **glicosidi idrolasi**. Essi sono un gruppo diffuso di enzimi che idrolizzano il legame glicosidico tra due o più carboidrati, o tra un carboidrato e una parte non carboidrata (https://en.wikipedia.org/wiki/Glycoside_hydrolase_family_65). I **glicosidi idrolasi** (chiamati anche **glicosidasi** o **glicosil idrolasi**) catalizzano l'idrolisi dei legami glicosidici negli zuccheri complessi. Sono enzimi estremamente comuni con ruoli in natura inclusa la degradazione della biomassa come cellulosa, emicellulosa e amido, nelle strategie di difesa antibatterica, nei meccanismi di patogenesi (ad esempio, **neuraminidasi virale**) e nella normale funzione cellulare. **Insieme alle glicosiltransferasi, le glicosidasi costituiscono il principale meccanismo catalitico per la sintesi e la rottura dei legami glicosidici** (https://en.wikipedia.org/wiki/Glycoside_hydrolase).

2-methylaconitate cis-trans isomerase PrpF [Duganella sp. AF9R3]

Sequence ID: [WP_169115272.1](#) Length: 393 Number 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-metilaconitato cis-trans isomerasi PrpF

Il ciclo **2-metilcitrato** è una delle almeno **cinque vie di degradazione per propionato** tramite *propionil-CoA*. La degradazione del propionato verso il piruvato consuma ossalacetato e rilascia succinato. Questa famiglia è costituita da **PrpF**, una proteina caratterizzata in modo incompleto che sembra essere una *proteina accessoria essenziale per la 2-metilisocitrato deidrataasi AcnD Fe/S-dipendente*. Questa proteina è correlata ma distinta da **FldA** (parte della famiglia Pfam PF04303), una presunta proteina di degradazione del fluorene di *Sphingomonas* sp. LB126 (https://www.ncbi.nlm.nih.gov/genome/annotation_prok/evidence/TIGR02334/).

pirin family protein [Photobacterium leiognathi]

Sequence ID: [WP_107184309.1](#) Length: 276 Number 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
```

Proteina della famiglia delle Pirine (Pirin_C)

Il **Dominio cupin Pirin C-terminale** si trova alla metà **C-terminale** della proteina **Pirina** (<https://www.ncbi.nlm.nih.gov/Structure/cdd/cddsrv.cgi?uid=331824>; <https://pfam.xfam.org/family/PF05726>).

Le *pirine eucariotiche* sono proteine nucleari altamente conservate che possono **funzionare come regolatori trascrizionali con un ruolo nell'apoptosi**. Sono stati identificati anche *omologhi procarioti*. È stato dimostrato che sia le pirine batteriche che quelle umane possiedono **attività quercetinasasi**, sebbene questo non sia universalmente vero per tutti i membri della famiglia - YhaK (SWISSPROT), per esempio, non mostra tale attività enzimatica. Pirin è composto da *due domini strutturalmente simili* disposti faccia a faccia (<http://pfam.xfam.org/family/PF05726>). Sebbene i due domini siano simili, il *dominio C-terminale della pirina* differisce dal *dominio N-terminale* in quanto non contiene un sito di *legame metallico* e la sua sequenza non contiene i residui conservati di *coordinamento del metallo*. La **Pirina** è considerata un membro della **superfamiglia "cupin"** sulla base della sequenza primaria e della somiglianza strutturale. La **presenza di un sito di legame metallico nel barile beta N-terminale della pirina** può essere significativa nella sua interazione con **Bcl-3** e il **fattore nucleare I (NFI)** e nel ruolo della **regolazione dell'attività del fattore di trascrizione NF-kappaB** (<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 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
```

Glicosiltransferasi famiglia 2

La **famiglia di glicosidi idrolasi 2** è una famiglia di **glicosidi idrolasi**. Le **glicosidi idrolasi** sono un gruppo diffuso di enzimi che idrolizzano il legame glicosidico tra due o più carboidrati, o tra un carboidrato e una parte non carboidrata. La **famiglia 2 della glicoside idrolasi** comprende enzimi con numerose attività note: **beta-galattosidasi; beta-mannosidasi; beta-glucuronidasi**. Questi enzimi contengono un residuo di acido glutammico conservato che è stato dimostrato essere il catalizzatore acido / base generale nel sito attivo dell'enzima. Il dominio catalitico delle beta-galattosidasi ha un **nucleo di barilotto TIM** circondato da molti altri domini in gran parte beta. Il dominio di legame dello zucchero di queste proteine ha una piega gelatinosa. Questi enzimi includono anche un **dominio beta-sandwich simile alle immunoglobuline** (https://en.wikipedia.org/wiki/Glycoside_hydrolase_family_2).

La **Regione 108..408** della **Sequenza ID: PES82498.1** (la sequenza sopra identificata) corrisponde con le **GH18 (glicosil idrolasi, famiglia 18) tipo II**. Queste chinasi idrolizzano la **chitina**, un polimero abbondante di *N-acetilglucosamina beta-1,4-legata (GlcNAc)* che è un componente importante della parete cellulare dei funghi e dell'esoscheletro degli artropodi. Le chitinasi sono state identificate in *virus, batteri, funghi, parassiti protozoi, insetti e piante*. La struttura del dominio GH18 è un **barile beta/alfa a otto filamenti con una fessura del sito attivo pronunciato all'estremità C-terminale del barile beta**. La famiglia GH18 include *chitotriosidasi, chitobiase, hevamina, zimocina-alfa, narbonina, SI-CLP* (proteina simile alla chitinasi interagente stabilin-1), *IDGF* (fattore di crescita del disco immaginale), *CFLE* (enzima corticale frammento litico) *idrolasi chitinasi vegetali di tipo III e tipo V, le endo-beta-N-acetilglucosaminidasi e le chitolectine*. Le **GH85** (glicosil idrolasi, famiglia 85) **ENGases** (endo-beta-N-acetilglucosaminidasi) sono strettamente correlate alle chinasi **GH18** e sono incluse in questo modello di allineamento (<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
```

Glicosiltransferasi

Si veda quanto discusso sopra e alle pagina 233 e 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 (Pattern a macchie nel Centro Germinale)

Un gruppo di ricercatori (K. Richter, S. Brar, M. Ray, P. Pisitkun, S. Bolland, L. Verkoczy, and M. Diaz; DOI: [10.1074 / jbc.M109.014506](#)) hanno "...identificato una **nuova GTPasi, SLIP-GC**, con espressione limitata a **pochi tessuti**, in particolare **le cellule B del centro germinale**. Manca di omologia con nessuna proteina nota, indicando che potrebbe appartenere a una nuova famiglia di GTPasi. SLIP-GC è espresso nelle cellule B del centro germinale e nei linfomi derivati dalle cellule B del centro germinale come i linfomi a cellule B larghe e diffuse. Nelle linee cellulari, *SLIP-GC* è espresso nei *linfomi* che esprimono la **deaminasi indotta dall'attivazione (AID)** e che probabilmente vanno incontro a **ipermutazione somatica**. SLIP-GC è una **proteina nucleare** e si localizza nelle **fabbriche di replicazione**. La riduzione dei livelli di SLIP-GC nella linea cellulare del linfoma di Burkitt Raji e nelle linee cellulari del linfoma non Hodgkin ha provocato un aumento delle rotture del DNA e dell'apoptosi dipendente dall'AID, poiché la riduzione simultanea dell'AID ha abrogato gli effetti deleteri della riduzione SLIP-GC. Questi risultati suggeriscono fortemente che SLIP-GC è una proteina correlata alla replicazione nelle cellule B del centro germinale la cui riduzione è tossica per le cellule attraverso un meccanismo dipendente dall'AID".

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			

Le idrolasi di serina

Le **idrolasi di serina** sono una delle più grandi classi di enzimi conosciute che comprendono circa ~ 200 enzimi o l'1% dei geni nel proteoma umano. Una caratteristica distintiva di questi enzimi è la presenza di una **serina nucleofila** nel loro sito attivo, che viene utilizzata per l'**idrolisi dei substrati**. La *catalisi* procede con la formazione di un intermedio enzimatico acilico attraverso questa **serina**, seguita dalla saponificazione dell'intermedio indotta da acqua / idrossido e dalla rigenerazione dell'enzima. A differenza di altre *serine non catalitiche*, la **serina nucleofila** di queste *idrolasi* è tipicamente attivata da un **relè protonico** che coinvolge una **triade catalitica** costituita dalla **serina**, un residuo acido (es. *aspartato* o *glutammato*) e un **residuo basico** (solitamente *istidina*), sebbene esistano variazioni su questo meccanismo. Le superfamiglie di **serina idrolasi** includono: *Proteasi della serina*, *Extracellulari lipasi*, *Intracellulari lipasi*, *Colinesterasi*, *Tioesterasi*, alcune *Fosfolipasi*, *Proteine e Glicani idrolasi*, alcune *Amidasi*, alcune *Peptidasi* (https://en.wikipedia.org/wiki/Serine_hydrolase).

Tra le sequenze identificate dalla ricerca BAST e riportate in questo paragrafo, ricordo gli **ancoraggi di membrana del glicosil-fosfatidilinositolo (GPI)**, principali responsabili dell'ancoraggio delle proteine della superficie cellulare nella membrana plasmatica e che svolgono funzioni nello *smistamento intracellulare*, nella *segnalazione transmembrana* e nei *nuovi processi endocitici*.

Un ruolo nella *trasduzione del segnale* (e nella *chemiotassi*) è svolto anche dalla sequenza dell'**emeritrina batterica**.

Stessa cosa dicasi per l'**Istidina chinasi**, enzima coinvolto nelle *vie di trasduzione del segnale* a monte di molti processi cellulari tra cui varie vie metaboliche, virulenti e omeostatiche.

Allineamenti significativi con la **famiglia di glicosidi idrolasi 65**, **glicosiltransferasi famiglia 2** [più precisamente, *GH18 (glicosil idrolasi, famiglia 18) tipo II*], **idrolasi di serina** (idrolasi molto importante e che verrà presa in grande considerazione nei capitoli successivi) e con la **proteina simile a LmbE (metalloidrolasi** che implica lo **zinco**) ripropongono di nuovo l'importanza delle famiglie delle **idrolasi** (già discusse nei paragrafi precedenti).

Il dominio **Zn-finger** è presente anche nella sequenza della famiglia **TM1802 (CRISPR e proteine Cas associate)**, famiglia coinvolta nella difesa dell'ospite contro *fagi* o altro *Dna estraneo*.

Un *sito di legame metallico* è presente anche nel *barile beta N-terminale della pirina* [riscontrato in un allineamento significativo con la sequenza di una **proteina della famiglia delle Pirine (Pirin_C)**], la cui presenza può essere significativa nell'interazione con *Bcl-3* e il *Fattore nucleare I (NFI)* e nel ruolo della regolazione dell'attività del fattore di trascrizione *NF-kappaB* (le *pirine eucariotiche* sono proteine nucleari altamente conservate che possono funzionare come *regolatori trascrizionali* con un ruolo nell'**apoptosi**).

L'**apoptosi dipendente dall'AID** (*Deaminasi Indotta dall'Attivazione*) e le **rottture del DNA** sono coinvolte nella **GTPase Nucleare SLIP-GC** (proteina quest'ultima correlata alla *replicazione nelle cellule B del centro germinale*).

Il *prodotto* di un'altra sequenza è invece "**2-metilaconitato cis-trans isomerasi PrpF**". In questo caso, è il ruolo della **deidratasi** che acquista importanza, in quanto **PrpF** è una *proteina accessoria essenziale per la 2-metilisocitrato deidratasi AcnD Fe/S-dipendente*.

L'allineamento con una sequenza del **canale del potassio MlotiK1**, coinvolto nelle **reti di segnalazione della vista** e dell'**olfatto**, sottolinea invece l'importanza del **sistema olfattivo**.

Infine, molto importante è l'identificazione di una sequenza che si riferisce al **dominio N-terminale della RNA polimerasi diretta dal DNA** (dominio che subisce un cambiamento strutturale che favorisce il passaggio del *trascritto dell'RNA*) perché la sequenza di aminoacidi **M A M T H K S Q M K** è correlata, attraverso un segmento della **Sequenza n° 1/3**, alla **sequenza originaria**, la nota sequenza (GTGAAATGGTCATGTGTGGCGG) del **gene RdRp** [o *RNA-dependent RNA polymerase*] del **SARS-CoV-2** (quella oggetto di studio di questo documento).

Prima di concludere desidero richiamare di nuovo l'attenzione sull'allineamento significativo delle sequenze (*che sono numerosissime*, anche se ne ho riportata solo una come esempio per tutte) il cui prodotto è la **glicosiltransferasi famiglia 2**, perché esso si riferisce al **Bacillus anthracis**. Questo **batterio**, che sarà identificato anche da altre sequenze ottenute dalle modificazioni della **sequenza originaria**, per la sua **letalità** e per il suo massiccio coinvolgimento nel **danno polmonare**, sarà oggetto di ampia trattazione nei capitoli successivi.

9.12 SEQUENZA DI AMINOACIDI “MAMTHKSQMKPRQMAP”

In questo paragrafo sono mostrati alcuni risultati della ricerca BLAST effettuata sulla sequenza di aminoacidi MAMTHKSQMKPRQMAP, la sequenza già presa in considerazione nel *paragrafo precedente* e a cui sono stati aggiunti gli aminoacidi **Prolina (P)**, **Arginina (R)**, **Glutamina (Q)**, **Metionina (M)**, **Alanina (A)** e **Prolina (P)**, determinati rispettivamente dai *codoni* CCT, CGT, CAA, ATG, GCA e CCA.

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 mldqdgmfela lalrrkkstv lesqpmdiec vsllhdihss nisnynyrgy tilgqndkyr
61 eikmyhalgp krslallmwh altgcdptls frgrskkmaf nawkknpsv teailqimek
121 pfmeidmgsy elkhldsymi ilysatvgdk inglrrkify hkyqnpeimp ptqdcldlqh
181 kravfqasiw amShkskmkp rnpskygwnv kenilfelsy kleemyhalg pkrslallmw
241 haltgcdptl sfrgrskkma fnawkknpsd vteailqime kpfmeidmgs yelkhldsym
301 iilysatvgd kinglrrkif yhkyqnpeim pptqdcldlqh ckrafvqasi wamShkskmk
361 prnpskygwn vkenilfels yklee
```

//

unnamed protein product, partial [Timema tahoe]Sequence ID: [CAD7463592.1](#) Length: 385 Number of Matches: 2Range 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_011927984.1Range 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

Proteina contenente il dominio di legame FAD (Dominio C-terminale della flavoproteina CO deidrogenasi)

Le **proteine** contenenti questo dominio formano complessi strutturali con altre *famiglie note*, come *Aldeide ossidasi/xantina deidrogenasi*, *superfamiglia del dominio di legame della molibdotterina* ([IPR008274](#)) e *Dominio di legame ferro-zolfo di tipo ferredossina 2Fe-2S* ([IPR001041](#)). Il *monossido di carbonio* (CO) *deidrogenasi* di *Oligotropha carbosidovorans* è un complesso eterotrimerico composto da un'**apoflavoproteina**, una *molibdotterina* e una proteina ferro-zolfo. Può essere dissociato con *sodio dodecilsolfato*. La CO deidrogenasi catalizza l'ossidazione della CO secondo la seguente equazione: $CO + H_2O = CO_2 + 2e + 2H^+$. La **subunità S** rappresenta la *proteina ferro-zolfo della CO deidrogenasi* ed è chiaramente divisa in un **dominio C** e uno **N-terminale**, ciascuno dei quali lega un cluster [2Fe-2S]. La struttura del **dominio C-terminale** ha una piega **beta-alfa** disposta in un **sandwich alfa + beta** (<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

Coniferil-aldeide deidrogenasi

Una **coniferil-aldeide deidrogenasi** è un **enzima** che catalizza una reazione chimica. I **4 substrates** di questo enzima sono **coniferil-aldeide**, **H₂O**, **NAD⁺**, e **NADP⁺**, mentre i suoi **4 prodotti** sono **ferulati**, **NADH**, **NADPH** e **H⁺**. Questo enzima appartiene alla famiglia delle *ossidoreduttasi*, in particolare quelle che agiscono sul *gruppo aldeidico* o *oxo* del donatore con *NAD +* o *NADP + come accettore*. Il nome sistematico di questa classe di enzimi è **coniferyl aldeide: NAD (P) + ossido reduttasi** (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 è una **proteina** nell'uomo codificata dal gene *DSP*. Desmoplakin è un componente critico delle *strutture desmosomiche* nel *muscolo cardiaco* e nelle *cellule epidermiche*, che **funzionano per mantenere l'integrità strutturale ai contatti cellulari adiacenti**. Nel **muscolo cardiaco**, il desmoplakin è localizzato in *dischi intercalati* che accoppiano meccanicamente le cellule cardiache per funzionare in una struttura sinciziale coordinata. È stato dimostrato che le **mutazioni in desmoplakin** svolgono un ruolo nella *cardiomiopatia dilatativa*, *cardiomiopatia aritmogena del ventricolo destro*, *cheratoderma palmo-plantare striato*, *sindrome di Carvajal* e *pemfigo paraneoplastico* (<https://en.wikipedia.org/wiki/Desmoplakin>).

helix-loop-helix DNA-binding domain-containing protein [Coniochaeta sp. 2T2.1]

Sequence ID: [KAB5551371.1](#) Length: 395 Number 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
```

Dominio di tipo Myc, basic helix-loop-helix (bHLH)

Un certo numero di **proteine eucariotiche**, che probabilmente sono **proteine leganti il DNA specifiche della sequenza che agiscono come fattori di trascrizione**, condividono un dominio conservato di 40-50 residui di amminoacidi. È stato proposto che questo dominio è formato da due eliche anfipatiche unite da una regione di collegamento di lunghezza variabile che potrebbe formare un anello. Questo dominio **“helix-loop-helix” (HLH)** media la **dimerizzazione di alcune proteine**. La maggior parte di queste proteine (si consulti questo sito web: <https://www.ebi.ac.uk/interpro/entry/InterPro/IPR011598/>) ha una *regione extra basica* di circa 15 residui di amminoacidi *adiacente al dominio HLH e che si lega specificamente al DNA*. Sono indicate come *proteine di base elica-anello-elica (bHLH)* e sono classificate in due gruppi: *classe A (ubiquitaria)* e *classe B (tessuto-specifica)*. I membri della famiglia bHLH associano variazioni alla sequenza principale “CANNTG”, nota anche come **motivo E-box**. L'omo o l'eterodimerizzazione mediata dal dominio HLH è indipendente, ma necessaria per il legame al DNA, poiché per l'attività di legame al DNA sono necessarie due regioni di base. Le proteine HLH prive del dominio di base (Emc, Id) funzionano come **regolatori negativi**, poiché formano eterodimeri, **ma non riescono a legare il DNA**. Anche le proteine pelose (hairy, E (spl), deadpan) **reprimono la trascrizione sebbene possano legare il DNA**. Le proteine di questa sottofamiglia agiscono insieme alle *proteine co-repressive*, come il *groucho*, attraverso il loro **motivo terminale 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
```

Glicosil idrolasi famiglia 3, proteina contenente il dominio C-terminale

La **Famiglia 3 di glicosidi idrolasi** è una famiglia di **glicosidi idrolasi** (Si veda quanto discusso riguardo le idrolasi nei paragrafi precedenti e il link <http://pfam.xfam.org/family/PF01915.22>). La Famiglia 3 di glicosidi idrolasi (**CAZY GH_3**) comprende *enzimi con una serie di attività note; beta-glucosidasi; beta-xilosidasi; N-acetil beta-glucosaminidasi; glucano beta-1,3-glucosidasi; cellodestrinasi; eso-1,3-1,4-glucanasi*. Questi enzimi sono *proteine globulari a due domini che sono N-glicosilate in tre siti* (<http://pfam.xfam.org/family/PF00933>; per importanti informazioni, consultare anche il 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](#)

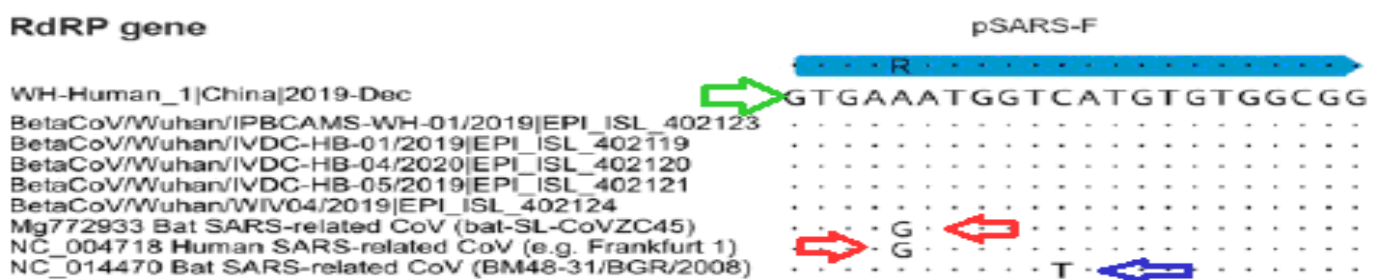
Score	Expect	Identities	Positives	Gaps
28.6 bits(60)	435	7/8(88%)	8/8(100%)	0/8(0%)

```
Query 7 SQMKPRQM 14
      S+MKPRQM
Sbjct 299 SHMKPRQM 306
```

MRPP3 ribonuclease

PRORP (RNasi P solo proteico) sono una classe di *enzimi di elaborazione dell'RNA* che catalizzano la maturazione dell'estremità 5' dei tRNA precursori negli eucarioti. *Arabidopsis thaliana* contiene gli enzimi PRORP (PRORP1, PRORP2 e PRORP3) dove PRORP1 si localizza nei mitocondri e nei cloroplasti, mentre PRORP2 e PRORP3 si trovano nel nucleo. Nell'uomo e nella maggior parte degli altri metazoi, mt-RNasi P è composta da **tre subunità proteiche (proteine mitocondriali RNasi P 1-3; MRPP1-3)**, omologhe dell'*Arabidopsis thaliana* PRORP1-3. Questo dominio corrisponde al dominio *metallonucleasi* dei PRORP. PRORP1 ha un'identità di sequenza del 22% rispetto all'omologo umano **MRPP3**. La struttura cristallina PRORP1 mostra una struttura tripartita a forma di V con un dominio metallonucleasi C-terminale della famiglia NYN (N4BL1, nucleasi simile a YacP), con un tipico e funzionale sito catalitico a due ioni metallici che ha conservato residui di aspartato (<https://pfam.xfam.org/family/PRORP#tabview=tab0>). *L. Reinhard, S. Sridhara e B. M. Hällberg* ci spiegano che “**L'RNA polimerasi mitocondriale produce precursori policistronici lunghi che contengono gli mRNA, gli rRNA e i tRNA necessari per la traduzione mitocondriale.** La RNasi P mitocondriale (mt-RNasi P) avvia la maturazione dei precursori scindendo alle estremità 5' dei tRNA. La mt-RNasi P umana è attiva solo come complesso tripartito (proteine mitocondriali RNasi P 1-3; MRPP1-3), mentre la pianta e la RNasi Ps tripanosomiale (PRORP), benché omologhe a MRPP3, sono attive come singole proteine. La ragione di questa discrepanza è rimasta finora oscura....**la struttura cristallina di MRPP3 umano, che presenta un sito attivo notevolmente distorto e quindi non produttivo...(passa)... a uno stato completamente produttivo solo dopo l'associazione con MRPP1, MRPP2 e il substrato pre-tRNA** (<https://pubmed.ncbi.nlm.nih.gov/25953853/>)”.

I **primi due allineamenti significativi** identificati dalla ricerca BLAST sulla sequenza **MAMTHKSQMKPRQMAP** (sequenza oggetto di approfondimento di questo paragrafo) si riferiscono ad un *prodotto proteico parziale e non classificato* dell'insetto **Timema tahoe**. **Timema** è un genere di *insetti stecco* relativamente corti e robusti originari degli *Stati Uniti occidentali* ed è l'unico membro esistente della *famiglia Timematidae* (<https://en.wikipedia.org/wiki/Timema>). Secondo Tanja Schwander, “i *timema* sono davvero gli *insetti più antichi* per i quali ci sono buone prove che siano stati asessuali per lunghi periodi di tempo”. Tanja Schwander è a capo di un team di ricercatori che hanno scoperto che cinque specie *Timema* (*T. douglasi*, *T. monikense*, *T. shepardi*, ***T. tahoe*** e *T. genevieveae*) hanno usato solo la riproduzione asessuata per più di 500.000 anni, con ***T. tahoe*** e *T. genevieveae* che *si riproducono asessualmente per oltre un milione di anni* (<https://en.wikipedia.org/wiki/Timema>). Ai fini della nostra ricerca, l'importanza di questo allineamento è giustificata dal fatto che (come annunciato nel Capitolo II e poi mostato nel Capitolo V) ***Timema tahoe*** è tra i pochissimi “organismi” identificati dalla ricerca BLAST sulla sequenza “**GTGAGATGGTCATGTGTGGCGG**”. Questa sequenza è importante perché essa è quasi totalmente identica alla sequenza “**GTGAAATGGTCATGTGTGGCGG**”, la sequenza di RNA oggetto di studio di questo documento. Inoltre, come già visto nel Capitolo III in Fig. 3.1 (della quale qui di seguito riporto il quadrante interessato), la sequenza “**GTGAGATGGTCATGTGTGGCGG**” (nel quadrante, sequenza indicata dalle frecce rosse) è una delle due sequenze del **gene RdRp** (*RNA polimerasi RNA dipendente*, il motore del processo di **trascrizione** del SARS-CoV-2) i cui allineamenti significativi delle ricerche BLAST identificano varianti della specie del **Pipistrello** [*Rhinolophus Sinicus* (chiamato *Ferro di Cavallo*), *Rhinolophus Pusillus* e *Rhinolophus Monoceros*].



La presenza dell'insetto **Timema tahoe** tra gli allineamenti significativi (mostrati in questo e nei capitoli successivi) non può certamente considerarsi casuale, ma determinata dall'azione del “*Trend Non Manifesto*” della **Sequenza n° 1/3** e di altre **sequenze modificate** (che saranno presentate nei successivi capitoli). Inoltre, è importante sottolineare che la **predizione dei codoni costitutivi del segmento** della **Sequenza n° 1/3** (e del segmento successivo di basi azotate identificato- si veda pagina 257) è corrispondente all'**ipotetica** sequenza di aminoacidi “**MAMTHKSQMKPRQMAP**” determinata dagli allineamenti significativi di una sequenza dell'**HIV-1**, la sequenza oggetto di studio approfondito di questo capitolo.

L'allineamento significativo con una sequenza della “*Proteina contenente il dominio di legame FAD (Dominio C-terminale della flavoproteina CO deidrogenasi)*” ripropone con insistenza il ruolo delle *deidrogenasi* e dei *legami metallici* (in questo caso, *legami di ferro-zolfo*).

Il ruolo delle *deidrogenasi* è ribadito dall'allineamento significativo con una sequenza il cui prodotto è l'enzima “*Coniferil-aldeide deidrogenasi*”.

Il prodotto “*Glicosil idrolasi famiglia 3, proteina contenente il dominio C-terminale*” dell'allineamento significativo di un'altra sequenza sottolinea invece il ruolo dell'*idrolasi*.

Meno chiaro è l'allineamento con una sequenza il cui prodotto è una proteina simile a *Desmoplakin* (componente critico delle *strutture desmosomiche* nel *muscolo cardiaco* e nelle *cellule epidermiche*), proteina che *funziona per mantenere l'integrità strutturale ai contatti cellulari adiacenti*. È stato dimostrato che le **mutazioni in desmoplakin** svolgono, tra l'altro, un ruolo nel *pemfigo paraneoplastico* (<https://en.wikipedia.org/wiki/Desmoplakin>). Il **pemfigo paraneoplastico (PNP)** è una *malattia autoimmune* derivante da un *tumore sottostante*. Si ipotizza che gli *antigeni associati al tumore* inneschino una *risposta immunitaria* con conseguente formazione di *vesciche sulla pelle e sulle mucose*. Circa 1/3 dei **decessi per PNP** derivano da **insufficienza polmonare causata dall'azione del PNP sulla mucosa respiratoria**. Si manifesta come *dispnea* e progredisce in *bronchiolite obliterante* (malattia polmonare ostruttiva non reversibile) attraverso un meccanismo sconosciuto (https://en.wikipedia.org/wiki/Paraneoplastic_pemphigus).

L'allineamento significativo relativo al “**Dominio di tipo Myc, basic helix-loop-helix (bHLH)**” riguarda *proteine leganti il DNA* (proteine già discusse nei paragrafi precedenti) *specifiche della sequenza che agiscono come fattori di trascrizione*.

Concludo questo paragrafo e l'intero capitolo, sottolineando l'identificazione di una sequenza che si riferisce alla “**MRPP3 ribonucleasi**” (della classe delle *proteine mitocondriali RNasi P 1-3; MRPP1-3*), la quale riprende in considerazione l'importanza degli **enzimi di elaborazione dell'RNA polimerasi mitocondriale**.

CAPITOLO X

**RICERCHE BLAST
E DISCUSSIONE DEI
RISULTATI OTTENUTI
DAGLI ALLINEAMENTI
DELLA SEQUENZA N° 3/1**

OBIETTIVI DEL CAPITOLO X

In questo capitolo sono riportati i risultati degli *allineamenti significativi* della ricerca BLAST effettuata sulla **Sequenza n° 3/1**. Questa sequenza è una delle tante generate dalla **Sequenza Originaria** [la sequenza “GTGAAATGGTCATGTGTGGCGG” del Gene RdRp (*RNA-dependent RNA polymerase*, RdRp_SARSr-F)] e dalla quale differisce per l’**81,82 %** delle basi.

I risultati degli allineamenti significativi della ricerca BLAST (**Hitlist size 100**) hanno riguardato un solo “**organismo**” comune (**Chanos chanos**) agli “**organismi**” identificati dalle ricerche BLAST effettuate sulle sequenze a cui è stata **sostituita una base alla sequenza originaria** (si vedano i Capitoli V e VI).

Estendendo la ricerca BLAST (**Hitlist size 5000**) sono stati individuati molti “**organismi**” **comuni** (distinguibili in diversi colori nelle pagine seguenti) e, in particolare, **Manis pentadactyla** (*Pangolino Cinese*).

La stessa ricerca BLAST (**Hitlist size 5000**) ha identificato sei sequenze di **Pipistrellus pipistrellus**, quattro sequenze del **Virus dell’Influenza A** e, soprattutto, un numero considerevole di sequenze dei **Bacilli Paranthracis** e **Anthraxis**.

Quest’ultimi due “**organismi**” non erano sicuramente previsti tra i risultati attesi, ancora meno di quanto potesse esserlo il *Virus dell’HIV-1* identificato dalla ricerca BLAST effettuata sulla **Sequenza n° 1/3**. Tuttavia, stranamente, come mostrato nel capitolo precedente, la ricerca BLAST effettuata sulla sequenza di aminoacidi **M A M T H K S Q M K** [determinata a partire dalla **Sequenza n° 1/3** (per approfondimenti si veda il capitolo precedente)] ha identificato numerosissime sequenze del **Bacillus Anthracis** [i cui “prodotti” corrispondono a *GH18* (*glicosil idrolasi, famiglia 18*) *tipo II*].

Il **Bacillus Anthracis**, che sarà identificato anche da altre sequenze ottenute dalle modificazioni della **sequenza originaria**, per la sua **letalità** e per il suo massiccio coinvolgimento nel **danno polmonare**, sarà oggetto di ampia trattazione in questo e nei capitoli successivi.

I **Bacilli Paranthracis** e **Anthraxis** fanno parte del **gruppo del Bacillo Cereus** (<http://lifemap-ncbi.univ-lyon1.fr/?tid=2026186>), gruppo che include anche l’agente patogeno di origine alimentare *B. Cereus* e il biopesticida *B. Thuringiensis* (https://en.wikipedia.org/wiki/Bacillus_anthraxis). La ricerca BLAST effettuata sulla **Sequenza n° 3/1** ha identificato numerose sequenze di tutti questi **Bacilli** (<https://mbio.asm.org/content/11/1/e00034-20>). “Il **gruppo del Bacillo Cereus** comprende 21 specie strettamente correlate di *batteri Gram-positivi* che formano **spore ubiquitarie**. Alcune di queste specie sono **patogeni opportunisti**. A causa dei diversi stili di vita e delle differenze nel contenuto genico, alcuni ceppi del gruppo *B. Cereus* sono considerati **probiotici**” (<https://mra.asm.org/content/8/44/e00662-19>).

Come già discusso nella **Premessa** di questo documento (e alla quale rimando per altre importanti informazioni), **Bacillus Anthracis** è un batterio gram-positivo infettivo, agente eziologico del **carbonchio ematico** nell’animale e dell’**antrace** nell’uomo (https://it.wikipedia.org/wiki/Bacillus_anthraxis).

“Il **carbonchio** è una patologia degli animali, soprattutto erbivori, causata dall’ingestione o dall’inalazione di bacilli vivi o sporificati. Gli animali affetti presentano febbre alta e vanno incontro rapidamente a morte, presentando all’esame autoptico una milza nerastra, fragile e ricchissima di bacilli, e sangue dall’aspetto scuro” (https://it.wikipedia.org/wiki/Bacillus_anthraxis).

L’antrace è l’equivalente umano del carbonchio. Il contagio uomo-animale è difficile, e ne sono esposti soltanto i pazienti che per lavoro sono a contatto prolungato con gli animali ospiti del batterio o con i loro prodotti, perciò può essere considerata una tipica zoonosi.

Il contagio interumano è impossibile” (https://it.wikipedia.org/wiki/Bacillus_anthraxis).

L’infezione da **antrace** si manifesta in *tre forme*:

- la **forma cutanea**, di gran lunga più comune, con una mortalità del 20% nei casi non trattati. Si manifesta con una lesione cutanea tipica. La malattia diventa letale quando i bacilli superano la barriera cutanea e si disseminano nel sangue, causando una grave setticemia;

- la **forma gastrointestinale**, rara, contratta attraverso l’ingestione di carne poco cotta proveniente da animali infetti. Si manifesta con sintomi generici gastrointestinali (nausea, inappetenza, diarrea sanguinolenta), seguiti da febbre alta e dolore addominale intenso;

- la **forma polmonare** (rarissima, presente soltanto in categorie a rischio), contratta per inalazione, **è notevolmente più grave** ed è stata associata alla paura diffusa di una guerra biologica (https://it.wikipedia.org/wiki/Bacillus_anthraxis)”.

La **forma polmonare** inizia con i sintomi tipici di un’influenza (mal di gola, febbre, mialgia); a seguire tosse, dolore al petto e difficoltà di respirazione. Questa forma è, dal punto di vista anatomo-patologico, una **linfadenite-mediastinite emorragica**, piuttosto che una **polmonite vera e propria**, e **non comporta presenza di bacilli nell’espettorato** (https://db0nus869y26v.cloudfront.net/it/Bacillus_anthraxis).

I **sintomi** in tutte e tre le forme possono manifestarsi **entro 7 giorni dal contatto con il batterio**, con l’eccezione della **forma polmonare che può presentarli a distanza di 7-42 giorni**. **L’infezione di antrace non è identificabile in maniera agevole**, perché “i sintomi della malattia (tranne che per la forma cutanea, più specifica) difficilmente consentono una diagnosi sicura. Un metodo più preciso è rappresentato dalla coltura del batterio nel sangue.” (https://it.wikipedia.org/wiki/Bacillus_anthraxis).

Il vaccino del **carbonchio** creato da Pasteur ha ridotto notevolmente il pericolo di infezioni. Più complesso è il problema dell’**antrace umana**. Il vaccino esistente contro la malattia non è disponibile al pubblico, a causa della **presenza di effetti collaterali**; è previsto invece il suo utilizzo per coloro che possono potenzialmente venire a contatto con il batterio, tra cui **alcune forze armate statunitensi** e il **personale di laboratori in cui sono presenti i bacilli**. L’antrace è considerata nella categoria A (cioè di massima priorità) degli agenti infettivi associati a rischio potenziale d’uso come **arma biologica**. (https://it.wikipedia.org/wiki/Bacillus_anthraxis).

SEQUENZA N° 3/1

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

10.1 SEQUENZA N° 3/1 (T=2; C=0; G=3; A=1) (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° 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

Query: **TATCCATTACATGATATTATT**

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_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
	Bacillus anthracis strain FDAARGOS_702 chromosome							
CP054800.1	Bacillus anthracis strain FDAARGOS_703 chromosome	Bacillus anthracis	38.2	38.2	86%	2.2	100.00%	5272559
	Bacillus anthracis strain FDAARGOS_703 chromosome							
CP054797.1	Bacillus thuringiensis strain FDAARGOS_791 chromosome, complete genome	Bacillus thuringiensis	38.2	38.2	86%	2.2	100.00%	5261520
CP054568.1	Bacillus thuringiensis strain HER1410 chromosome, complete genome	Bacillus thuringiensis	38.2	38.2	86%	2.2	100.00%	5281841
CP050183.1	Bacillus thuringiensis strain FDAARGOS_792 chromosome, complete genome	Bacillus thuringiensis	38.2	72.4	100%	2.2	100.00%	5585577
CP053938.1	Bacillus thuringiensis strain FDAARGOS_794 chromosome, complete genome	Bacillus thuringiensis	38.2	70.4	86%	2.2	100.00%	5251676
CP053934.1	Bacillus cereus strain FDAARGOS_797 chromosome, complete genome	Bacillus thuringiensis	38.2	70.4	86%	2.2	100.00%	5214223
CP053931.1	Bacillus cereus strain FDAARGOS_780 chromosome, complete genome	Bacillus cereus	38.2	70.4	86%	2.2	100.00%	5413450
CP053997.1	Bacillus cereus strain FDAARGOS_781 chromosome, complete genome	Bacillus cereus	38.2	70.4	86%	2.2	100.00%	5271040
CP053991.1	Bacillus thuringiensis strain FDAARGOS_795 chromosome,	Bacillus thuringiensis	38.2	70.4	86%	2.2	100.00%	5271029
CP053980.1								

	complete genome								
CP053954.1	Bacillus cereus strain FDAARGOS_798 chromosome, complete genome	Bacillus cereus	38.2	38.2	86%	2.2	100.00%	5381410	
CP053965.1	Bacillus cereus strain FDAARGOS_802 chromosome, complete genome	Bacillus cereus	38.2	38.2	86%	2.2	100.00%	5342923	
CP053955.1	Bacillus tropicus strain FDAARGOS_782 chromosome, complete genome	Bacillus tropicus	38.2	38.2	86%	2.2	100.00%	5263142	
CP053951.1	Bacillus cereus strain FDAARGOS_799 chromosome, complete genome	Bacillus cereus	38.2	38.2	86%	2.2	100.00%	5381556	
CP053981.1	Bacillus thuringiensis strain FDAARGOS_793 chromosome, complete genome	Bacillus thuringiensis	38.2	38.2	86%	2.2	100.00%	5256259	
CP053289.1	Bacillus cereus strain WPySW2 chromosome, complete genome	Bacillus cereus	38.2	72.4	100%	2.2	100.00%	5328644	
CP049978.1 1/2 8/1	Bacillus sp. RZ2MS9 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 genome	Bacillus anthracis	38.2	38.2	86%	2.2	100.00%	5227419
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
CP026607.1/2	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
	Bacillus anthracis strain HDZK-BYSB7 chromosome, complete genome							
CP026608.1	Bacillus wiedmannii bv. thuringiensis strain FCC41 chromosome, complete genome	Bacillus anthracis Bacillus wiedmannii bv. thuringiensis	38.2	70.4	86%	2.2	100.00%	5294779
CP024684.1			38.2	38.2	86%	2.2	100.00%	5513449
CP033795.1/2 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 strain MCCC 1A02161 chromosome, complete genome							
CP031642.1	Bacillus anthracis strain MCCC 1A01412 chromosome, complete genome	Bacillus anthracis	38.2	70.4	86%	2.2	100.00%	5231857
CP031643.1								
CP034516.1	Eukaryotic synthetic construct chromosome 13	Bacillus anthracis eukaryotic synthetic construct	38.2	70.4	86%	2.2	100.00%	5291783
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%	96089878
CP034548.1	Bacillus albus strain PFYN01 chromosome, complete genome	Bacillus albus	38.2	38.2	86%	2.2	100.00%	5637049
	Bacillus anthracis strain London_499 chromosome, complete genome							
CP029805.1	Bacillus cereus strain FORC60 chromosome, complete genome	Bacillus anthracis	38.2	38.2	86%	2.2	100.00%	5228599
CP020383.1	Bacillus cereus strain TG1-6 chromosome, complete genome	Bacillus cereus	38.2	72.4	100%	2.2	100.00%	5361178
CP026678.1			38.2	72.4	100%	2.2	100.00%	5417196
	Bacillus anthracis strain FDAARGOS_341 chromosome, complete genome							
CP022044.2	Bacillus cereus ATCC 10987 chromosome	Bacillus anthracis Bacillus cereus ATCC 10987	38.2	38.2	86%	2.2	100.00%	5228806
CP026375.1	Bacillus cereus strain G9241 chromosome		38.2	104	100%	2.2	100.00%	4871269
CP026376.1			38.2	104	100%	2.2	100.00%	4880708
CP025122.1/2 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

VISTA DI ALLINEAMENTO VERTICALE AGGANCIATO A QUERY CON LETTERE D'IDENTITA'

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	CCCATACATGATATT	587176
CP053934.1	2468315	TATCCCATTACATGATATT	2468333
CP053934.1	3602104	CCCATACATGATATT	3602089
CP053931.1	4336123	TATCCCATTACATGATATT	4336141
CP053931.1	216870	CCCATACATGATATT	216855
CP053997.1	615231	TATCCCATTACATGATATT	615249
CP053997.1	1757114	CCCATACATGATATT	1757099
CP053991.1	2062698	TATCCCATTACATGATATT	2062680
CP053991.1	920826	CCCATACATGATATT	920841
CP053980.1	4574420	TATCCCATTACATGATATT	4574402
CP053980.1	3420689	CCCATACATGATATT	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

<u>CP053289.1</u>	540310	CATTACATGATATTATT	540326
<u>CP049978.1</u>	2581345	TATCCCATTACATGATATT	2581327
<u>CP050973.1</u>	1083080	TATCCCATTACATGATATT	1083098
<u>CP050972.1</u>	3047578	TATCCCATTACATGATATT	3047596
<u>CP050971.1</u>	61005	TATCCCATTACATGATATT	60987
<u>CP050970.1</u>	4915760	TATCCCATTACATGATATT	4915742
<u>CP049019.1</u>	1935112	TATCCCATTACATGATATT	1935094
<u>CP049019.1</u>	791722	CCCATTACATGATATT	791737
<u>CP048687.1</u>	5037812	TATCCCATTACATGATATT	5037830
<u>CP048687.1</u>	430393	CCCATTACATGATATT	430378
<u>AP022643.1</u>	68533	TATCCCATTACATGATATT	68515
<u>CP028009.1</u>	5140082	TATCCCATTACATGATATT	5140100
<u>CP027920.1</u>	2880954	TATCCCATTACATGATATT	2880936
<u>CP041981.1</u>	2708055	TATCCCATTACATGATATT	2708037
<u>CP041981.1</u>	1645460	CCCATTACATGATATT	1645475
<u>CP041750.1</u>	2498278	TATCCCATTACATGATATT	2498260
<u>CP041750.1</u>	1427940	CCCATTACATGATATT	1427955
<u>CP047131.1</u>	2458699	TATCCCATTACATGATATT	2458681
<u>CP047111.1</u>	400614	TATCCCATTACATGATATT	400632
<u>CP047107.1</u>	4514812	TATCCCATTACATGATATT	4514830
<u>CP047104.1</u>	231066	TATCCCATTACATGATATT	231048
<u>CP047099.1</u>	2849926	TATCCCATTACATGATATT	2849908
<u>CP047098.1</u>	5146998	TATCCCATTACATGATATT	5147016
<u>CP047097.1</u>	1622506	TATCCCATTACATGATATT	1622488
<u>CP047085.1</u>	1792017	TATCCCATTACATGATATT	1791999
<u>CP047085.1</u>	717567	CCCATTACATGATATT	717582
<u>CP046398.1</u>	2475271	TATCCCATTACATGATATT	2475253
<u>CP046398.1</u>	4811602	CATTACATGATATTATT	4811586
<u>CP040344.1</u>	4155946	TATCCCATTACATGATATT	4155928
<u>CP040342.1</u>	3933233	TATCCCATTACATGATATT	3933251
<u>CP040342.1</u>	1632121	CATTACATGATATTATT	1632137
<u>CP040340.1</u>	3334375	TATCCCATTACATGATATT	3334357
<u>CP040340.1</u>	122423	CATTACATGATATTATT	122407
<u>CP040334.1</u>	3724392	TATCCCATTACATGATATT	3724410
<u>CP040334.1</u>	1469776	CATTACATGATATTATT	1469792
<u>CP045777.1</u>	3106170	TATCCCATTACATGATATT	3106188
<u>CP045777.1</u>	4177604	CCCATTACATGATATT	4177589
<u>CP045606.2</u>	2670105	TATCCCATTACATGATATT	2670087
<u>CP045606.2</u>	4981301	CATTACATGATATTATT	4981285
<u>CP045537.1</u>	2495274	TATCCCATTACATGATATT	2495256
<u>CP045537.1</u>	1459815	CCCATTACATGATATT	1459830
<u>CP045533.1</u>	2467973	TATCCCATTACATGATATT	2467955
<u>CP045533.1</u>	1432466	CCCATTACATGATATT	1432481
<u>CP042270.1</u>	3697495	TATCCCATTACATGATATT	3697477
<u>CP042270.1</u>	726299	CATTACATGATATTATT	726283
<u>CP044978.1</u>	3359847	TATCCCATTACATGATATT	3359829
<u>CP044978.1</u>	313889	CATTACATGATATTATT	313873
<u>CP032365.1</u>	2623129	TATCCCATTACATGATATT	2623111
<u>XM 030889606.1</u>	1055	CCCATTACATGATATTATT	1037
<u>XM 030911861.1</u>	1422	CCCATTACATGATATTATT	1404
<u>CP031068.1</u>	2497032	TATCCCATTACATGATATT	2497014
<u>CP031065.1</u>	2530420	TATCCCATTACATGATATT	2530402
<u>CP031065.1</u>	4747176	CATTACATGATATTATT	4747160
<u>CP031062.1</u>	2529186	TATCCCATTACATGATATT	2529168
<u>CP031062.1</u>	4746547	CATTACATGATATTATT	4746531
<u>CP042929.1</u>	2716926	TATCCCATTACATGATATT	2716908
<u>CP042929.1</u>	5050716	CATTACATGATATTATT	5050700
<u>CP042874.1</u>	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

ALLINEAMENTI SIGNIFICATIVI CON I BACILLI DELL'ANTRACE E CON QUELLI DEI BACILLI DEL PARANTRACE (CDS FEATURE)

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="MKTQDQIDNIVKETYRHRIDFSGVILVKEKKGIVYEEVFGYANRN
ECINNTLQTRFGIASGCKIFTAIGICQLVENGVITFQTKLKECLSVNFPNFNEDITIH
HLLTHSSGIPDYFDESIMDNFEDLWKQTPMYLLKSLKDFLPLFQNRDMMYAPGSKFHY
NNAGFIILGLIEEQTGLTFTEYIEKNIFNPIGMNHSYFSLDRLPRQTALGYIKDEI
SQTWRTNAYSIIPIKGGSDGGAFITASDMLKFWREALFNYEIISQEYTKILLTPHIQVNN
NQSYGYGIWIETRENKIFKYHVMGYDPGVSFSAVYPDLGITLVIPSNKGAGAELMI
EIEGSF"

```

ORIGIN

1 **tatccatta 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; ncrRNA;
                             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 givyeefgy anrnecinnt lqtrfgiasg
61 ckiftaigic qlvengvitf qtklkeclsv nfpnfedit ihhllthssg ipdyfdesim
121 dnfedlwkqt pmyllkslkd flplfqnrmd myapgskfhy nnagfiilgl iieeqtgltf
181 teyieknifn pigmnhsgyf sldrlprqta lgyikdeisq twrtnaysip ikggsdggaf
241 itasdmkfw ealfnyeiis qeytkilltp hiqvnngsy gygiwietre nkifkyhvmg
301 Ydpgvsvfrsa vpydlgitlv 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](#)

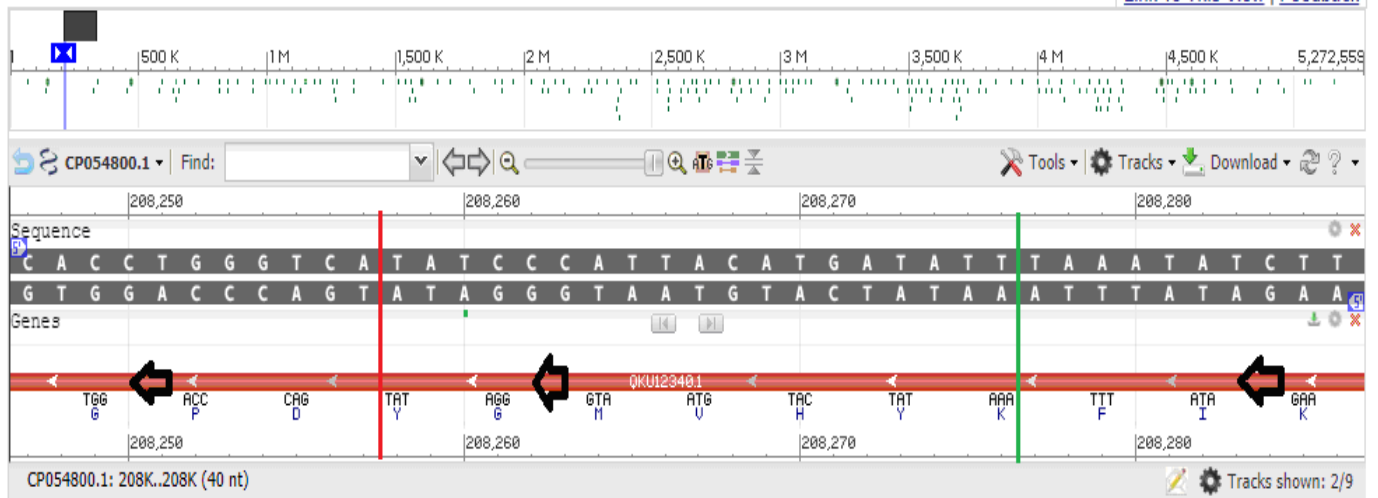
Score	Expect	Identities	Gaps	Strand
38.2 bits(19)	2.2	19/19(100%)	0/19(0%)	Plus/Plus
CDS: Putative 1	5	G M V H Y		
Query	1	TATCC CATTAC ATGATATT	19	
Sbjct	208258	TATCC CATTAC ATGATATT	208276	
CDS: serine hydrolase	301	Y G M V H Y K		

Bacillus anthracis strain FDAARGOS_702 chromosome

GenBank: CP054800.1

[GenBank](#) [FASTA](#)

[Link To This View](#) | [Feedback](#)



Osservando il grafico appena mostrato, si notano **due aminoacidi** comuni rispettivamente alle sequenze (già mostrate nei capitoli precedenti e che ripropongo qui di seguito) del “**Bat SARS-like coronavirus**” e della “**Severe acute respiratory syndrome coronavirus 2**”.

Gli aminoacidi in questione sono la **Valina (V)** e la **Metionina (M)**, codificati rispettivamente dai codoni (letti secondo il senso stabilito dalla freccia nera) **GTA** e **ATG** [nel grafico presenti nella sequenza denominata “*Genes*” e corrispondenti a **TAC** e **CAT** nella sequenza denominata “*Sequence*”].

La differenza che si nota nella sequenza del **Bacillus anthracis** rispetto alle sequenze del “**Bat SARS-like coronavirus**” e della “**Severe acute respiratory syndrome coronavirus 2**” è nel codone **GTA** della **Valina (V)**, in cui la base “**A**” sostituisce la base “**C**” nel codone corrispondente **GTC** (si vedano le due sequenze riportate qui di seguito).

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	GTGAG ATGGT CATGTGTGGCGG 22		
Sbjct	15418	GTGAG ATGGT CATGTGTGGCGG 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	GTGAG ATGGTC ATGTGTGGCGG 22		
Sbjct	15377	GTGAR ATGGTC ATGTGTGGCGG 15398		
CDS:ORF1ab polyprote	5056	S E M V M C G G		

ATTENZIONE: R = A o G

Alla luce di quanto già discusso, soprattutto nel capitolo precedente, la corrispondenza di questi due aminoacidi (tra le tre sequenze sopra considerate) acquista una certa rilevanza, anche perché il “**prodotto**” della sequenza del **Bacillus anthracis** identificato dalla ricerca BLAST è la **Serina** (le **Idrolasi di Serina**), già ampiamente presa in considerazione nel Capitolo IX.

Le idrolasi di serina

Le **idrolasi di serina** sono una delle più grandi classi di enzimi conosciute che comprendono circa ~ 200 enzimi o l'1% dei geni nel proteoma umano. Una caratteristica distintiva di questi enzimi è la presenza di una **serina nucleofila** nel loro *sito attivo*, che viene utilizzata per l'**idrolisi dei substrati**. La *catalisi* procede con la formazione di un intermedio enzimatico acilico attraverso questa **serina**, seguita dalla saponificazione dell'intermedio indotta da acqua / idrossido e dalla rigenerazione dell'enzima. A differenza di altre *serine non catalitiche*, la **serina nucleofila** di *queste idrolasi* è tipicamente attivata da un **relè protonico** che coinvolge una **triade catalitica** costituita dalla **serina**, un residuo acido (es. *aspartato* o *glutammato*) e un **residuo basico** (solitamente *istidina*), sebbene esistano variazioni su questo meccanismo. Le superfamiglie di **serina idrolasi** includono: *Proteasi della serina*, *Extracellulari lipasi*, *Intracellulari lipasi*, *Colinesterasi*, *Tioesterasi*, alcune *Fosfolipasi*, *Proteine e Glicani idrolasi*, alcune *Amidasi*, alcune *Peptidasi* (https://en.wikipedia.org/wiki/Serine_hydrolase). Come mostrato a pagina 282, il nome della **regione** (region_name) dell'**Idrolasi di Serina** (protein_id="QKU12340.1") è “**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>). Le **Beta lattamasi** (β -lattamasi, β L), sono una grande famiglia di idrolasi che comprende più di 850 membri identificati espressi in batteri Gram-positivi e Gram-negativi. I β L possono essere classificati in base alla loro specificità del substrato o dell'inibitore. *Questi enzimi sono in grado di idrolizzare quattro anelli atomici noti come β -lattamici* (https://www.abcam.com/beta-lactamase-activity-assay-kit-colorimetric-ab197008.html?gclid=aw.ds&gclid=EAIaIQobChMIiLTdx6Oo7wIVAuqyCh2G6AsAEAYASAAEgJrPD_BwE; si veda anche https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi?seqinput=NP_845386.1). **COG1680** è

membro della **superfamiglia transpeptidasi cl21491**, **dominio transpeptidasi della proteina legante la penicillina**. Il **sito attivo della serina** è conservato in tutti i membri di questa famiglia (<https://www.ncbi.nlm.nih.gov/Structure/cdd/cl21491>). Questa **codifica** identifica un **ampio gruppo di proteine**, che includono (<https://www.ebi.ac.uk/interpro/entry/InterPro/IPR001460/>):

- **Precursori della beta-lattamasi, penicillinasi** (un gruppo di enzimi di varia specificità che idrolizzano i β -lattamici; alcuni agiscono più rapidamente sulle *penicilline*, altri più rapidamente sulle *cefalosporine*. Questi ultimi erano precedentemente elencati come CE 3.5.2.8, *cefalosporinasi*) (<https://www.ebi.ac.uk/intenz/query?cmd=SearchEC&ec=3.5.2.6>);

- **Peptidoglicano sintetasi *ftsI*** (*peptidoglicano glicosiltransferasi 3*) è un enzima che funziona anche quando il residuo di *lisina* viene sostituito dal meso-2,6-diaminoeptandioato (meso-2,6-diamminopimelato, A2pm) combinato con residui adiacenti attraverso il suo centro L, poiché è in organismi *Gram-negativi* e alcuni *Gram-positivi*. L'*undecaprenolo* coinvolto è *ditrans, octacis-undecaprenol*. Coinvolto nella sintesi del *peptidoglicano della parete cellulare*] (<https://www.ebi.ac.uk/intenz/query?cmd=SearchEC&ec=2.4.1.129>).

- **Resistenza alla meticillina proteina *mecR1*** (si veda sotto);

- **Proteina *blaR1* resistenza alla meticillina** (si veda sotto).

Il gran numero di **proteine leganti la penicillina**, che sono rappresentate in questo gruppo di sequenze, sono **responsabili degli stadi finali della biosintesi del peptidoglicano per la formazione della parete cellulare**. Le proteine sintetizzano il *peptidoglicano reticolato* da intermedi lipidici e contengono un **dominio carbossi-terminale transpeptidasi sensibile alla penicillina**. Il **sito attivo della serina (residuo 337 in P14677)** è conservato in tutti i membri di questa famiglia.

MecR1 e **BlaR1** sono *metallopeptidasi* appartenenti alla famiglia di *peptidasi MEROPS M56, clan M-*. *BlaR1* e *MecR1* scindono i loro *repressori trascrizionali affini BlaI* e *MecI*, rispettivamente, attivando la *sintesi di MecA*.

MecR1 è presente in *Staphylococcus aureus* e *Staphylococcus sciuri*, mentre **BlaR1** (noto anche come *BlaR*, *PenR1* o *PenJ*) è stato trovato in *Bacillus licheniformis*, *Staphylococcus epidermidis*, *Staphylococcus haemolyticus* e diversi ceppi di *S. aureus*. Queste proteine sono codificate da **plasmidi**, cromosomiche o mediate da trasposoni. Le proteine **MecR1/BlaR1** sono costituite da *domini metallopeptidasi transmembrana omologhi di 330 residui N-terminali* collegati a *porzioni di sensori di penicillina simili a PBP omologhe 260 residui extracellulari* (<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
                        NNAGFIILGLIEEQTGLTFTEYIEKNIFNPIGMNHSGYFSLDRLPRQTALGYIKDEI
                        SQTWRTNAYSIPKGGSDGGAFITASDMLKFWEALFNYEIISQEYTKILLTPHIQVNN
                        NQSYGYGIWIETRENKIFKYHVMGYDPGVSFSAVYPDLGITLVIPSNKGAGAELMI
                        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	TATCC	CATTACATGATATT		19
Sbjct	65999	TATCC	CATTACATGATATT		65981
CDS: serine hydrolase	301	Y	G M V	H Y K	

ATTENZIONE

La sequenza “**aatatcatgt aatgggata**” mostrata sopra è la *sequenza contraria e speculare* della sequenza “**TATCCATTACATGATATT**” (sequenza parziale della **Sequenza n° 3/1**, “**TATCCATTACATGATATTATT**”):

ATAGGGTAATGTACTATAA

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

TATCCATTACATGATATT

Bacillus anthracis strain FDAARGOS_703 chromosome

GenBank: CP054797.1
[GenBank](#) [FASTA](#)



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 micRbial 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="MKTKDQIDNIVKETYRHHDFSGVILVKEKKGIVYEEVFGYANRN
ECINNTLQTRFGIASGCKIFTAIGICQLVENGVITFQTKLKECLSVNFPNFNEDITIH
HLLTHSSGIPDYFDESIMDNFEDLWKQTPMYLLKSLKDFLPLFQNRDMMYAPGSKFHY
NNAGFIILGLIEEQTGLTFTEYIEKNIFNPIGMNHSYFSLDRLPRQTALGYIKDEI
SQTWRTNAYSIPKGGSDGGAFITASDMLKFWREALFNYEIISQEYTKILLTPHIQVNN
NQSYPYGIWIETRENKIFKYHVMGYDPGVSFSAVYPDLGITLVIPSNKGAGAEKLMIEIEGSF"

```

ORIGIN

1 **tatccatta 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	TATCCATTACATGATATT	19	
Sbjct	1083080	TATCCATTACATGATATT	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
                     SQTWR TNAYSIPKGGSDGGAFITASDMLKFWEALFNYEIISQEYTKILLTPHIQVNN
                     NQSYGYGIWIETRENKIFKYHVMGYDPGVSFSAVYPDLGITLVIPSNKGAGAEKLMIEIEGSF"

ORIGIN
1  tatccatta 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		TATCCCATACATGATATT 19	
Sbjct	3047578		TATCCCATACATGATATT 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="MKTQDQIDNIVKETYRHIDFSGVILVKEKKGIVYEEVFGYANRN
ECINNTLQTRFGIASGCKIFTAIGICQLVENGVITFQTKLKECLSVNFPNFNEDITIH
HLLTHSSGIPDYFDESIMDNFEDLWKQTPMYLLKSLKDFLPLFQNRDMMYAPGSKFHY
NNAGFIILGLIEEQTGLTFTEYIEKNIFNPIGMNHSYFSLDRLPRQTALGYIKDEI
SQTWRTNAYSIPKGGSDGGAFITASDMLKFWREALFNYEIIISQEYTKILLTPHIQVNN
NQSYYGIWIETRENKIFKYHVMGYDPGVSFSAVYPDLGITLVIPSNKGAGAELMI
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	TATCCCATACATGATATT	19	
Sbjct	61005	TATCCCATACATGATATT	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="MKTQDQIDNIVKETYRHIDFSGVILVKEKKGIVYEEVFGYANRN
ECINNTLQTRFGIAGSCKIFTAIGICQLVENGVITFQTKLKECLSVNFPNFNEDITIH
HLLTHSSGIPDYFDESIMDNFEDLWKQTPMYLLKSLKDFLPLFQNRDMMYAPGSKFHY
NNAGFIILGLIIEEQTGLTFTEYIEKNIFNPIGMNHSYFSLDRLPRQTALGYIKDEI
SQTWRTNAYSIPKGGSDGGAFITASDMLKFWEALFNYEIISQYETKILLTPHIQVNN
NQS YGYGIWIETRENKIFKYHVMGYDPGVSFSAVYPDLGITLVIPSNKGAGAEKLMI
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	TATCCCATACATGATATT	19	
Sbjct	4915760	TATCCCATACATGATATT	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 **tatccatta 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		TATCCATTACATGATATT	19	
Sbjct 5037812		TATCCATTACATGATATT	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
GVEGMVLLHLINQVNP SAKVVF LDTNVHFQETYELIQKVRERFPSLNII EKQPKLTLD
EQAKLHGNKLWESNPNLCKIRKISPLEESLANEKAWISGLRREQSETRKHTKFINQD
HRFQSIKVCPLIHWTWKEVWRYVYKHS LPYNPLHDIGYPSIGCEKCTLPVGE GGD SRD
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

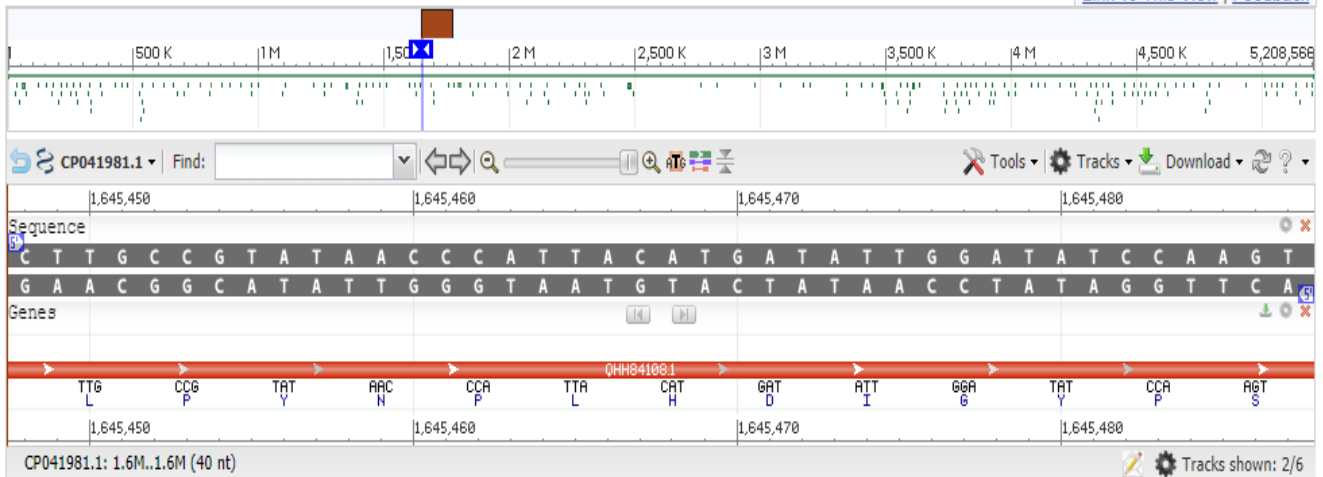
CDS: Putative 1	1	P L H D I	
Query	4	CCCATTACATGATATT	19
Sbjct	1645460	CCCATTACATGATATT	1645475
CDS: phosphoadenylyl-	192	P L H D I	

Bacillus paranthracis strain NCCP 15910 chromosome, complete genome

GenBank: CP041981.1

[GenBank](#) [FASTA](#)

[Link To This View](#) | [Feedback](#)



Fosfoadenilil-solfato reduttasi (tioredossina)

Fosfoadenilil-solfato reduttasi (tioredossina) è un enzima ([EC 1.8.4.8](#)) che catalizza la reazione chimica
$$\text{adenosina } 3', 5\text{-bisfosfato} + \text{solfito} + \text{tioredossina disolfuro} \rightleftharpoons 3\text{-fosfoadenil solfato} + \text{tioredossina}$$

I 3 substrati di questo enzima sono adenosina 3', 5'-bisfosfato, solfito e tioredossina disolfuro, mentre i suoi due prodotti sono **3'-fosfoadenil solfato** e **tioredossina**. Questo enzima appartiene alla famiglia delle *ossidoreduttasi*, in particolare quelle che agiscono su un gruppo di *donatori di zolfo* con un disolfuro come accettore. Questo enzima partecipa al **metabolismo dello zolfo** ([https://en.wikipedia.org/wiki/Phosphoadenylyl-sulfate_reductase_\(thioredoxin\)](https://en.wikipedia.org/wiki/Phosphoadenylyl-sulfate_reductase_(thioredoxin))).

Il **3'-fosfoadenil solfato** o **3'-fosfoadenosina-5'-fosfosolfato (PAPS)** è un derivato dell'*adenosina monofosfato* che è fosforilato nella posizione 3' e ha un gruppo solfato attaccato al fosfato 5'. È il coenzima più comune nelle *reazioni sulfotransferasi*. Viene sintetizzato endogenamente dagli organismi tramite la fosforilazione dell'*adenosina 5'-fosfosolfato (APS)*, un metabolita intermedio. Nell'uomo tale reazione viene eseguita da *3'-fosfoadenosina 5'-fosfosolfato sintasi bifunzionale (PAPSS1 e PAPSS2)* utilizzando ATP come donatore di fosfato (<https://en.wikipedia.org/wiki/3%27-Phosphoadenosine-5%27-phosphosulfate>).

La **Tioredossina (Trx)** è una classe di *piccole proteine redox* note per essere presenti in tutti gli organismi. La tioredossina svolge un ruolo in molti importanti processi biologici, inclusa la *segnalazione redox*. Recentemente è stato scoperto che queste proteine svolgono un ruolo anche nella comunicazione cellula-cellula (<https://en.wikipedia.org/wiki/Thioredoxin>). Le tioredossine sono mantenute allo stato ridotto dalla *flavoenzima tioredossina reduttasi*, in una *reazione dipendente dal NADPH*. La funzione primaria della tioredossina (Trx) è la *riduzione dei residui di cisteina ossidata* e la *scissione dei legami disolfuro*.

La **Tioredossina reduttasi (TR, TrxR)** ([EC 1.8.1.9](#)) sono gli unici enzimi conosciuti per ridurre la **tioredossina (Trx)**. Sono state identificate *due classi di tioredossina reduttasi*: una classe in *batteri* e alcuni *eucarioti* e una in *animali*. Nei **batteri TrxR** catalizza anche la *riduzione della glutaredossina come proteine note come NrdH*. Entrambe le classi sono **flavoproteine** che funzionano come *omodimeri*. Ogni monomero contiene un *gruppo protesico FAD*, un *dominio di legame NADPH* e un *sito attivo* contenente un *legame disolfuro attivo redox* (https://en.wikipedia.org/wiki/Thioredoxin_reductase). L'attività di questo enzima è essenziale per la crescita e la sopravvivenza delle cellule, è un buon bersaglio per la **terapia antitumorale**. Inoltre, **l'enzima è sovraregolato in diversi tipi di cancro**. Recentemente sono state condotte alcune ricerche per dimostrare che la *tioredossina reduttasi a basso peso molecolare* potrebbe essere un bersaglio per nuovi antibiotici (come *auranofin* o *Ebselen*) e potrebbe essere utilizzato per **batteri resistenti agli antibiotici**.

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
HLLTHSSGIPDYFDESILDNFEDLWKETPMYLLKLNKDFLPLFQNRDMMFAPGSKFHY
NNAGFIILGLIIEEQTGLKFTEYIETNIFKFLIGMNSGYFSLDRLPKQTALGYIKDEI
SQTWRTNAYSIPKGGSDGGAFITAPDMMKFWEALFNYEIISQEYTKILLTPHIQVNS
YQSYGYGIWIEIRENKIFKYHVMGYDPGVSFSAVYPDLGITLVIPSNKGAGPEKLM
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		TATCCATTACATGATATT 19	
Sbjct	2498278		TATCCATTACATGATATT 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
      GVEGMVLLHLLINQVNPVSAKVVFLDTNVHFQETYELIQVRRERFPSLNII EKQPKLTLD
      EQAKLHGNKLWESNPNLCKIRKILPLEESLANEKAWISGLRREQSETRKHTKFINQD
      HRFQSIKVCPLIHWTWKEVWRVYVKHSLPYNPLHDIGYPSIGCEKCTLPVGE GDSRD
      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="MKTQDQIDNIVKETYRHHDFSGVILVKEKKGIVYEEVFGYANRN
ECINNTLQTRFGIASGCKIFTAIGICQLVENGVITFQTKLKECLSVNFPNFNEDITIH
HLLTHSSGIPDYFDESIMDNFEDLWKQTPMYLLKSLKDFLPLFQNRDMMYAPGSKFHY
NNAGFIILGLIIEEQTGLTFTEYIEKNIFNPIGMNHSYFSLDRLPRQTALGYIKDEI
SQTWRNTNAYSIPKGGSDGGAFITASDMLKFWEALFNYEIISQEYTKILLTPHIQVNN
NQSYYGIGIWIETRENKIFKYHVMGYDPGVSFSAVYVYDGLGITLVIPSNKGAGAELMI
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		TATCCATTACATGATATT	19
Sbjct	2458699		TATCCATTACATGATATT	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
                     NNAGFIILGLIEEQTGLTFTEYIEKNIFNPIGMNHSGYFSLDRLPRQTALGYIKDEI
                     SQTWRTNAYSIPKGGSDGGAFITASDMLKFWEALFNYEIISQEYTKILLTPHIQVNN
                     NQSYGYGIWIETRENKIFKYHVMGYDPGVSFSAVYPDLGITLVI PSNKGAGA EKLMI
                     EIEGSF"

```

ORIGIN

1 **tatccatta 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="MKTQDQIDNIVKETYRHIDFSGVILVKEKKGIVYEEVFGYANRN
ECINNTLQTRFGIASGCKIFTAIGICQLVENGVITFQTKLKECLSVNFPNFNEDITIH
HLLTHSSGIPDYFDESIMDNFEDLWKQTPMYLLKSLKDFLPLFQNRDMMYAPGSKFHY
NNAGFIILGLIIEEQTGLTFTEYIEKNIFNPIGMNHSQYFSLDRLPRQALGYIKDEI
SQTWRTNAYSIPKGGSDGGAFITASDMLKFWEALFNYEIISQEYTKILLTPHIQVNN
NQS YGYGIWIETRENKIF **KYHVMGY** DPGVSFRSAVYPDLGITLVIPSNKGAGAEKLMI
EIEGSF"

ORIGIN

1 **tatccatta catgatatt**

//

Bacillus anthracis strain FDAARGOS_705 chromosome, complete genome

Sequence ID: [CP047107.1](#) Length: 5227452 Number 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		TATCCATTACATGATATT 19	
Sbjct	4514812		TATCCATTACATGATATT 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="MKTQDQIDNIVKETYRHIDFSGVILVKEKKGIVYEEVFGYANRN
ECINNTLQTRFGIASGCKIFTAIGICQLVENGVITFQTKLKECLSVNFPNFNEDITIH
HLLTHSSGIPDYFDESIMDNFEDLWKQTPMYLLKSLKDFLPLFQNRDMMYAPGSKFHY
NNAGFIILGLIIEEQTGLTFTEYIEKNIFNPIGMNHSYFSLDRLPRQTALGYIKDEI
SQTWRTNAYSIIPIKGGSDGGAFITASDMLKFWEALFNYEIISQYETKILLTPHIQVNN
NQSYYGIGIWIETRENKIFKYHVMGYDPGVSFSAVYVYDLGITLVIPSNKGAGAELMI
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="MKTQDIDNIVKETYRHIDFSGVILVKEKKGIVYEEVFGYANRN
                     ECINNTLQTRFGIASGCKIFTAIGICQLVENGVITFQTKLKECLSVNFPNFNEDITIH
                     HLLTHSSGIPDYFDESIMDNFEDLWKQTPMYLLKSLKDFLPLFQNRDMMYAPGSKFHY
                     NNAGFIILGLIEEQTGLTFTEYIEKNIFNPIGMNHSGYFSLDRLPRQTALGYIKDEI
                     SQTWRTNAYSIPKGGSDGGAFITASDMLKFWEALFNYEIISQYETKILLTPHIQVNN
                     NQSYGYGIWIETRENKIFKYHVMGYDPGVSFSAVYPDLGITLVI PSNKGAGA EKLM I
                     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="MKTQDQIDNIVKETYRHFDFSGVILVKEKKGIVYEEVFGYANRN
ECINNTLQTRFGIASGCKIFTAIGICQLVENGVITFQTKLKECLSVNFPNPNEDITIH
HLLTHSSGIPDYFDESIMDNFEDLWKQTPMYLLKSLKDFLPLFQNRDMMYAPGSKFHY
NNAGFIILGLIEEQTGLTFTEYIEKNIFNPIGMNHSYFSLDRLPRQTALGYIKDEI
SQTWRNTNAYSIPKGGSDGGAFITASDMLKFWEALFNYEIISQEYTKILLTPHIQVNN
NQSRYGYGIWIETRENKIFKYHVMGYDPGVSFSAVYPDLGITLVIPSNKGAGAEKLMIEIEGSF"

```

ORIGIN

1 **tatccatta 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	TATCCATTACATGATATT	19	
Sbjct	5146998	TATCCATTACATGATATT	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="MKTQDQIDNIVKETYRHIDFSGVILVKEKKGIVYEEVFGYANRN
ECINNTLQTRFGIASGCKIFTAIGICQLVENGVITFQTKLKECLSVNFPNFNEDITIH
HLLTHSSGIPDYFDESIMDNFEDLWKQTPMYLLKSLKDFLPLFQNRDMMYAPGSKFHY
NNAGFIILGLIIEEQTGLTFTEYIEKNIFNPIGMNHSYFSLDRLPRQTALGYIKDEI
SQTWRTNAYSIPKGGSDGGAFITASDMLKFWEALFNYEIISQYTKILLTPHIQVNN
NQSYYGYGIWIETRENKIFKYHVMGYDPGVSFSAVYVYDLGITLVIPSNKGAGAEKLMIEIEGSF"

```

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	TATCCCATACATGATATT	19	
Sbjct	1622506	TATCCCATACATGATATT	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
                        EYINNTLQTRFGIASGCKIFTAIGICQLVERGVITFHHTKLKDCLSIKFSNFNEDITIH
                        HLLTHSSGIPDYFDESILDNFEDLWKETPMYLLKNLKDPLPLFQNRDMMFAPGSKFHY
                        NNAGFIILGLIIEEQTGLKFTEYIETNIFKLI GMNDSGYFSLDRLPKQTALGYIKDEI
                        SQTWR TNAYS IPIKGGSDGGAFITAPDMMKFWEALFNYEIISQ EYTKILLTPHIQVNS
                        YQSYGYGIWIEIRENKIFKYHVMGYDPGV SFRSAVYPDLGITLVIPSNKGAGPEKLM I
                        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	TATCC	CATTAC	ATGATATT	19	
Sbjct	1792017	TATCC	CATTAC	ATGATATT	1791999	
CDS: serine hydrolase	301	Y	G	M V	H	Y
						K

Bacillus paranthracis strain BC307 chromosome, complete genome

GenBank: CP047085.1

[GenBank](#) [FASTA](#)

[Link To This View](#) | [Feedback](#)



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
GVEGMVLLHLINQVNP SAKVVF LDTNVHFQETYELIQVRRERF PSLNII EKQPKLTL D
EQAKLHGK NLWESNPNL CCKIRKIL PLEESLANEKAWISGLRREQSETRKHTKFINQD
HRFQSIKVCPLIHW TWKEVWRYVYKHS LPYN PLHDIGYPSIGCEKCTLPV GEGGDSRD
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 BC307 chromosome, complete genome

GenBank: CP047085.1
[GenBank](#) [FASTA](#)



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
                        EYINNTLQTRFGIASGCKIFTAIGICQLVERGVITFHTKLDCLSIKFSNFNEDITIH
                        HLLTHSSGIPDYFDESILDNFEDLWKETPMYLLKNLKDPLPLFQNRDMMFAPGSKFHY
                        NNAGFIILGLIIEEQTGLKFTEYIETNIFKLGIMNDSGYFSLDRLPKQTALGYIKDEI
                        SQTWRTNAYSIPKGGSDGGAFITAPDMMKFWEALFNYEIISQEYTKILLTPHIQVNS
                        YQSYGYGIWIEIRENKIFKYHVMGYDPGVSFSAVYPDLGITLVIPSNKGAGPEKLMIE
                        IEIIEGAF"
ORIGIN
     1  tatccatta 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		TATCCCATACATGATATT 19	
Sbjct	3106170		TATCCCATACATGATATT 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"
```

```

gene      /collected_by="FDA"
          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
GVEGMVLLHLINQVNPSTAKVVFVLDTNVHFQETYELIQKVRERFPSLNIIEKQPKLTL
EQAKLHGKWLWESNPNLCKIRKILPLEESLANEKAWISGLRREQSETRKHHTKFINQD
HRFQSIKVCPLIHWTKVWRYVYKHSPLYNPLHDIGYPSIGCEKCTLPVGGGDSRD
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](#)

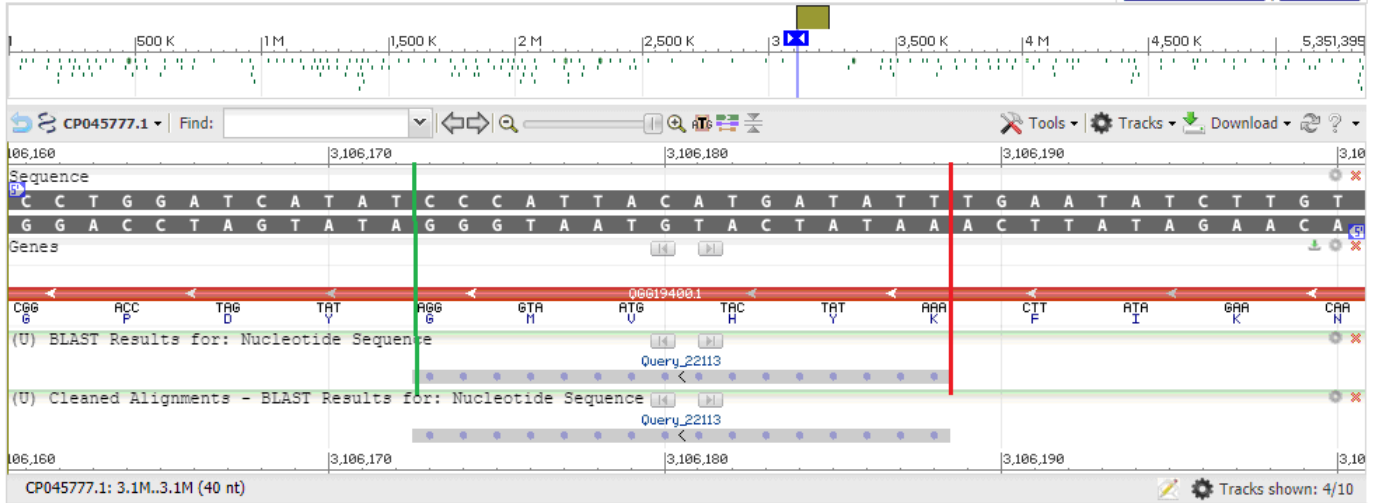
Score	Expect	Identities	Gaps	Strand
32.2 bits(16)	143	16/16(100%)	0/16(0%)	Plus/Minus
CDS: Putative 1	1		Y H V M G	
Query	1		AATATCATGTAATGGG	16
Sbjct	3106188		AATATCATGTAATGGG	3106173
CDS:serine hydrolase	295		K Y H V M G	

Bacillus paranthracis strain CFSAN068816 chromosome, complete genome

GenBank: CP045777.1

[GenBank](#) [FASTA](#)

[Link To This View](#) | [Feedback](#)



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

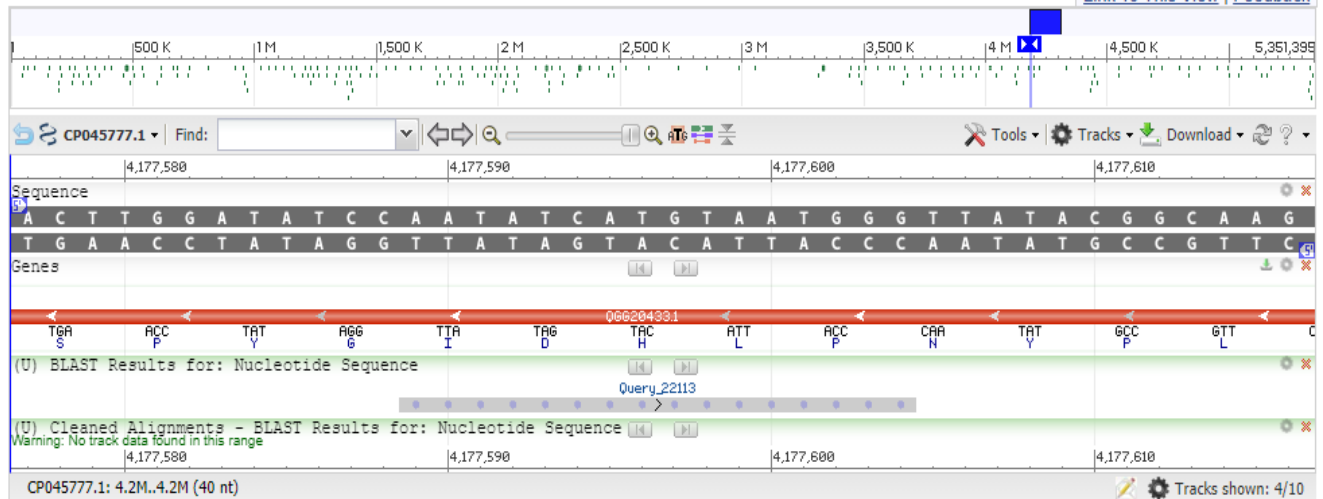
CDS: Putative 1 1 P L H D I
 Query 4 CCCATTACATGATATT 19
 |||
 Sbjct 4177604 CCCATTACATGATATT 4177589
 CDS: **phosphoadenylyl-** 192 P L H D I

Bacillus paranthracis strain CFSAN068816 chromosome, complete genome

GenBank: CP045777.1

[GenBank](#) [FASTA](#)

[Link To This View](#) | [Feedback](#)



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="MKTQDQIDNIVKETYRHHDFSGVILVKEKKGIVYEEVFGYANRN
ECINNTLQTRFGIASGCKIFTAIGICQLVENGVITFQTKLKECLSVNFPNFNEDITIH
HLLTHSSGIPDYFDESIMDNFEDLWKQTPMYLLKSLKDFLPLFQNRDMMYAPGSKFHY
NNAGFIILGLIIEEQTGLTFTEYIEKNIFNPIGMNHSYFSLDRLRQ TALGYIKDEI
SQTWR TNAYS IPIKGGSDGGAFITASDMLKFWEALFNYEIISQ EYTKILLTPHIQVNN
NQSYGYGIWIETRENKIF KYHVMGYDPGV SFRSAVYPDLGITLVIPSNKGAGAEKLMI
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		TATCCATTACATGATATT 19	
Sbjct	1368346		TATCCATTACATGATATT 1368328	
CDS: penicillin-bindi	301		Y G M V H Y K	

Penicillin-binding proteina 4

Si veda quanto discusso a pagina 286 sulla *Idrolasi di Serina*.

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, Laengasstrasse 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="MKTKDQIDNIVKETYRHFIDFSGVILVKEKKGIVYEEVFGYANRN
                     ECINNTLQTRFGIASGCKIFTAIGICQLVENGVITFQTKLKECLSVNFPNFNEDITIH
                     HLLTHSSGIPDYFDESIMDNFEDLWKQTPMYLLKSLKDFLPLFQNRDMMYAPGSKFHY
                     NNAGFIILGLIIEEQTGLTFTEYIEKNIFNPIGMNHSYFSLDRLPRQTALGYIKDEI
                     SQTWRTNAYSIPKGGSDGGAFITASDMLKFWEALFNYEIISQEYTKILLTPHIQVNN
                     NQSYGYGIWIETRENKIFKYHVMGYDPGVSFSAVYPDLGITLVIPSNKGAGAELMI
                     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		TATCCCATACATGATATT	19
Sbjct	2458998		TATCCCATACATGATATT	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
HLLTHSSGIPDYFDESIMDNFEDLWKQTPMYLLKSLKDFLPLFQNRMYMYAPGSKFHY
NNAGFIILGLIIEEQTGLKFTEYIEKNIFNPIGMNHSYFSLDRLPRQTALGYIKDEI
SQTWRTNAYSIPKGGSDGGAFITASDMLKFWREALFNYEIISQEYTKILLTPHIQVNN
NQSYGYGIWIETRENKVFKYHVMGYDPGVSFSAVYPDLGITLVIPSNKGAGPEKLMIE
IEGAF"

```

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 Rec J penicillin-binding protein				
CDS: Putative 1	5		G M V H Y	
Query	1		TATCCATTACATGATATT	19
Sbjct	3594133		TATCCATTACATGATATT	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-BYSE7
 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="MLTYETWEENNVSFSEEDETKGALSVLSWAYKEYEYEGEIVYACSF
                     GVEGMVLLHLINQVNPSTAKVVFVLDTNVHFQETYELIQKVRERFPSLNIIIEKQPKLTLTLD
                     EQAKLHGNKLWESNPNLCKIRKILPLEESLANEKAWISGLRREQSETRKHTKFINQD
                     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: SAM00002586
               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
SQTWRTNAYSIPKGGSDGGAFITASDMLKFWREALFNYEIISQEYTKILLTPHIQVNN
NQSYGYGIWIETRENKIFKYHVMGYDPGVSFSAVYPDLGITLVIPSNKGAGAEKLMIEIEGSF"

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		TATCCCATACATGATATT	19
Sbjct	2463024		TATCCCATACATGATATT	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="AX098692.1"
          /translation="MKTQDQIDNIVKETYRHFDFSGVILVKEKKGIVYEEAFGYANRN
          ECINNTLQTRFGIASGCKIFTAIGICQLVENGVITFQTKLKECLSVNFPNFNEDITIH
          HLLTHSSGIPDYDESIMDNFEDLWKQTPMYLLKSLKDFLPLFQNRDMMYAPGSKFHY
          NNAGFIILGLIIEEQTGLTFTEYIEKNIFNPIGMNHSYFSLDRLRQRTALGYIKDEI
          TQTWRTNAYSIPKGGSDGGAFITASDMLKFWEALFNYEIISQYTKILLTPHIQVNN
          NQSYGYGIWIETRENKIFKYHVMGYDPGVSFSAVYPDLGITLVIPSNKEAGPEKLMIE
          IEIGDF"

```

```

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	TATCCCATACATGATATT	19	
Sbjct	2575655	TATCCCATACATGATATT	2575637	
CDS: class C beta-lac	301	Y G M V H Y K		

Class C beta-lactamase-related serine hydrolase

Si veda quanto discusso a pagina 286 sull'Idrolasi di Serina.

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
                        GVEGMVLLHLINQVNPQAKVVFVLDTNVHFQETYELIQKVRERFPSLNIIEKQPKLTL
                        EQDKLHGDKLWESNPNLCKIRKILPLEESLANEKAWISGLRREQSETRKHTKFINQD

```

HRFQSIKVCPLIHWTWKEVWRVYVKHSLPYN**PLHDI**GYPSIGCEKCTLPVGE^{GG}DSRD
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="MKTQIDNIVKETYRHHDFSGVILVKEKKGIVYEEAFGYANRN
ECINNTLQTRFGIASGCKIFTAIGICQLVENVITFQTKLKECLSVNFPNFNEDITIH
HLLTHSSGIPDYFDESIMDNFEDLWKQTPMYLLKSLKDFLPLFQNRDMMYAPGSKFHY
NNAGFIILGLIIEEQTGLKFTEYIEKNIFNPIGMNHSGYFSLDRLPRQTALGYIKDEI
SQTWRTNAYSIPKGGSDGGAFITAPDMLKFWEALFNYEIISQEYTKILLTPHIQVNN
NQSYYGIWIETRENKVF**KYHVMGY**DPGVSFSAVYPDLGITLVIPSNKGAGPEKLMI
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		TATCCCATACATGATATT 19	
Sbjct	2546571		TATCCCATACATGATATT 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="MLTYETWEENNVSFSEEDETKGALSVLSWAYKEYEIGEIVYACSF
                     GVEGMVLLHLINQVNPQAKVVFVLDTNVHFQETYELIQKVRERFPSLNIIEKQPKLTL
                     EQAKLHGNKLWESNPNLCKIRKILPLEESLANEKAWISGLRREQSETRKHTKFINQD

```


HRFQSIKVCPLIHWTWKEVWRYVYKHS LPYN**PLHDI**GYPSIGCEKCTLPVGE GDSRD
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="MKTQIDNIVKETYRHIDFSGVILVKEKKGIVYEEVFGYANRN
ECINNTLQTRFGIASGCKIFTAIGICQLVENGVITFQTKLKECLSVNFPNFNEDITIH
HLLTHSSGIPDYFDESIMDNFEDLWKQTPMYLLKSLKDFLPLFQNRDMMYAPGSKFHY
NNAGFIILGLIIEEQTGLTFTEYIEKNIFNPIGMNHSYFSLDRLPRQTALGYIKDEI
SQTWRTNAYSIPKGGSDGGAFITASDMLKFWEALFNYEIISQEYTKILLTPHIQVNN
NQSYGYGIWIETRENKIF**KYHVMGY**DPGVSFSAVYPDLGITLVIIPSNKGAGAEKLMI
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

```

```

CDS
    /locus_tag="CEQ19_00235"
    <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="MKTQDQIDNIVKETYRHIDFSGVILVKEKKGIVYEEVFGYANRN
    ECINNTLQTRFGIASGCKIFTAIGICQLVENGVI TFQTKLKECLSVNFPNFNEDITIH
    HLLTHSSGIPDYFDESIMDNFEDLWKQTPMYLLKSLKDFLPLFQNRDMMYAPGSKFHY
    NNAGFIILGLIIEEQTGLTFTEYIEKNIFNPIGMNHSYFSLDRLPRQTALGYIKDEI
    SQTWRTNAYSIIPIKGGSDGGAFITASDMLKFWEALFNYEIISQEYTKILLTPHIQVNN
    NQSYGYGIWIETRENKIFKYHVMGYDPGVSFSAVYPDLGITLVIPSNKGAGAEKLMI
    EIEGSF"

```

ORIGIN

1 **aatatcatgt aatggata**

//

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	TATCCCATACATGATATT	19	
Sbjct	43833	TATCCCATACATGATATT	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	

ALTRI PRODOTTI DEGLI ALLINEAMENTI SIGNIFICATIVI (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		TATCCATTACATGATATT	19
Sbjct	4982888		TATCCATTACATGATATT	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		TATCCATTACATGATATT	19
Sbjct	2588912		TATCCATTACATGATATT	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		TATCCATTACATGATATT	19
Sbjct	1731233		TATCCATTACATGATATT	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      TATCCCATACATGATATT 19
                  |||
Sbjct             2468315 TATCCCATACATGATATT 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      TATCCCATACATGATATT 19
                  |||
Sbjct             4336123 TATCCCATACATGATATT 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		TATCCATTACATGATATT	19
Sbjct	4574420		TATCCATTACATGATATT	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		TATCCATTACATGATATT	19
Sbjct	5056629		TATCCATTACATGATATT	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: 5342923 Number 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		TATCCCATACATGATATT 19	
Sbjct	3056974		TATCCCATACATGATATT 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		TATCCCATACATGATATT 19	
Sbjct	3671059		TATCCCATACATGATATT 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		TATCCCATACATGATATT 19	
Sbjct	1001141		TATCCCATACATGATATT 1001159	
CDS: serine hydrolase	302		Y G M V H Y K	

Bacillus thuringiensis strain FDAARGOS_793 chromosome, complete genome

Sequence ID: [CP053281.1](#) 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 TATCCATTACATGATATT 19
 |||||
 Sbjct 3437422 TATCCATTACATGATATT 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 TATCCATTACATGATATT 19
 |||||
 Sbjct 2826804 TATCCATTACATGATATT 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 TATCCATTACATGATATT 19
 |||||
 Sbjct 2581345 TATCCATTACATGATATT 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		TATCCATTACATGATATT 19	
Sbjct	2880954		TATCCATTACATGATATT 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		TATCCATTACATGATATT 19	
Sbjct	2475271		TATCCATTACATGATATT 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		TATCCATTACATGATATT 19	
Sbjct	2670105		TATCCATTACATGATATT 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		TATCCATTACATGATATT 19	
Sbjct	3697495		TATCCATTACATGATATT 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		TATCCCATACATGATATT 19	
Sbjct	3359847		TATCCCATACATGATATT 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		TATCCCATACATGATATT 19	
Sbjct	2497032		TATCCCATACATGATATT 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		TATCCCATACATGATATT	19
Sbjct	2530420		TATCCCATACATGATATT	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		TATCCCATACATGATATT	19
Sbjct	2529186		TATCCCATACATGATATT	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		TATCCCATACATGATATT	19
Sbjct	2565787		TATCCCATACATGATATT	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		TATCCCATACATGATATT	19
Sbjct	2451680		TATCCCATACATGATATT	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		

Proteina di resistenza alla bacitracina BacA (PRK00281)

PRK00281 è un membro della superfamiglia [cl00858](#). La **proteina di resistenza alla bacitracina (BacA)** è una *nondecaprenolo chinasi putativa*. BacA conferisce resistenza alla bacitracina, probabilmente per fosforilazione dell'undecaprenolo. Studi più recenti mostrano che BacA ha un'attività *undecaprenil pirofosfato fosfatasi*. Il *difosfato di undecaprenile* è un composto poliprenolo difosfato con undici unità prenili con stereochimica indefinita sui doppi legami. Ha un ruolo come metabolita *Saccharomyces cerevisiae*. È un *poliprenile difosfato e un fosfato undecaprenile* (<https://pubchem.ncbi.nlm.nih.gov/compound/Undecaprenyl-pyrophosphate>; <https://www.ebi.ac.uk/chebi/searchId.do?chebiId=CHEBI:53042>). Il **fosfato di undecaprenil** è un intermedio lipidico chiave coinvolto nella sintesi di vari polimeri della parete cellulare batterica. La bacitracina, una miscela di antibiotici polipeptidi ciclici correlati, viene utilizzata per trattare le infezioni dei tessuti superficiali. Il suo principale modo di azione è l'inibizione della sintesi della parete cellulare batterica attraverso il sequestro del pirofosfato lipidico essenziale non decaprenile, C55-PP, con conseguente perdita di integrità cellulare e lisi. Il caratteristico motivo a sequenza fosfatasi in questa famiglia è probabilmente il PGxSRSGG, rispetto al PSGH della famiglia pap delle fosfatasi (<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	TATCCATTACATGATATT	19	
Sbjct	1106485	TATCCATTACATGATATT	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		TATCCCATACATGATATT 19	
Sbjct	2546892		TATCCCATACATGATATT 2546874	
CDS: Penicillin-bindin	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		TATCCCATACATGATATT 19	
Sbjct	2649612		TATCCCATACATGATATT 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		TATCCCATACATGATATT 19	
Sbjct	2587934		TATCCCATACATGATATT 2587916	
CDS: penicillin-bindin	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		TATCCCATACATGATATT 19	
Sbjct	2524526		TATCCCATACATGATATT 2524508	
CDS: penicillin-bindin	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	TATCCCATACATGATATT 19
Sbjct	2600026	TATCCCATACATGATATT 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	TATCCCATACATGATATT 19
Sbjct	2707607	TATCCCATACATGATATT 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 1774691 [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	TATCCCATACATGATATT 19
Sbjct	1774691	TATCCCATACATGATATT 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		TATCCCATACATGATATT 19	
Sbjct	2730671		TATCCCATACATGATATT 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		TATCCCATACATGATATT 19	
Sbjct	2518436		TATCCCATACATGATATT 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		TATCCCATACATGATATT 19	
Sbjct	4967235		TATCCCATACATGATATT 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		TATCCATTACATGATATT 19	
Sbjct	5166699		TATCCATTACATGATATT 5166681	
CDS: penicillin-bindin	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		TATCCATTACATGATATT 19	
Sbjct	2770639		TATCCATTACATGATATT 2770657	
CDS: penicillin-bindin	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		TATCCATTACATGATATT 19	
Sbjct	2405971		TATCCATTACATGATATT 2405953	
CDS: penicillin-bindin	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	TATCCCATACATGATATT 19
Sbjct	2498158	TATCCCATACATGATATT 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	TATCCCATACATGATATT 19
Sbjct	4159070	TATCCCATACATGATATT 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

Proteina della famiglia TIGR02687

I membri di questa famiglia sono proteine insolitamente distribuite in batteri e archei, circa 880 amminoacidi di lunghezza. Questa proteina si trova ripetutamente a monte di un'altra proteina insolita di circa 470 amminoacidi di lunghezza, modellata da TIGR02688. TIGR02687 è un membro della superfamiglia 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      TATCCCATACATGATATT 19
                    |||
Sbjct                2700310 TATCCCATACATGATATT 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      TATCCCATACATGATATT 19
                    |||
Sbjct                2709891 TATCCCATACATGATATT 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	TATCCCATACATGATATT	19	
Sbjct	237616	TATCCCATACATGATATT	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	TATCCCATACATGATATT	19	
Sbjct	990178	TATCCCATACATGATATT	990196	
CDS: penicillin-	277	D W * M I N		

Penicillin-binding protein

Si veda quanto discusso a pagina 286 sulla *Idrolasi di Serina*.

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 4846937 [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	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 I PRINCIPALI “PRODOTTI” IDENTIFICATI DALLE RICERCHE BLAST EFFETTUATE SULLA **SEQUENZA N° 3/1** (Hitlist size: 100) E IL RUOLO FONDAMENTALE SVOLTO DALLE IDROLASI DI SERINA

I “prodotti” delle sequenze dei **Bacilli Paranthracis e Anthracis** identificati dalla ricerca BLAST sono le **Idrolasi di Serina** [in particolare, la regione “AmpC” (*CubicO group peptidase, beta-lactamase class C family; COG1680*, membro della superfamiglia *transpeptidasi cl21491, dominio transpeptidasi della proteina legante la penicillina*)] e la **Fosfoadenil-solfato reduttasi (tiodrossina)**.

Prima di tutto è necessario approfondire il ruolo svolto dalle **Idrolasi di Serina**, più precisamente dalla **Serina Proteasi Transmembrana 2** (TMPRSS2, *Transmembrane Serine Protease 2*) e **4** (TMPRSS4, *Transmembrane Serine Protease 4*), nella COVID-19. L’espressione di TMPRSS2 e TMPRSS4 (**serina proteasi mucosa-specifiche**) facilita l’attività fusogena della **Spike di SARS-CoV-2** e promuove l’ingresso del virus nelle cellule ospiti (<https://www.uniprot.org/citations/32404436>).

Le **proteasi della serina** sono coinvolte nel processo di infezione innescando la **proteina spike del coronavirus**.

In un ottimo articolo di diversi autori italiani (<https://sibbm.zanichelli.it/italiano/2020/08/23/infezione-biologia-strutturale/#proteasi-principale>) è spiegato che: “L’ingresso del coronavirus in cellule suscettibili è un processo complesso che richiede l’azione concertata del legame al recettore e del processamento proteolitico della **proteina S** per promuovere la fusione tra virus e cellula. Una delle **proteasi cellulari** coinvolta nell’ingresso della CoV-2 nelle cellule ospiti è la **TMPRSS2**, che è richiesta anche per l’infezione da SARS-CoV. Infatti, l’inibizione di **TMPRSS2** blocca l’ingresso di SARS-CoV e CoV-2 nelle cellule”.

E’ noto che “...il recettore dell’enzima di conversione dell’angiotensina 2 (ACE2) può legarsi alla **proteina spike di SARS-CoV-2**. La *proteina spike di SARS-CoV-2* può essere scissa e attivata dalla **proteasi transmembrana**, la **serina 2** (TMPRSS2) delle cellule ospiti per l’infezione da SARS-CoV-2. Pertanto, ACE2 e TMPRSS2 sono potenziali bersagli antivirali per il trattamento della prevenzione dell’infezione da SARS-CoV-2” (https://www.sciencedirect.com/science/article/pii/S075333222031009X?fbclid=IwAR0Y7RMgCMj_d2RdrGasy5g05QranXJi9dxKZtCnCG-zYJWVrVg0tEZgf-0). Nello studio appena citato (si veda il link tra parentesi), gli autori hanno scoperto che “....GB-2 può inibire l’espressione dell’*mRNA* di ACE2 ed l’espressione delle *proteine* ACE2 e TMPRSS2 in cellule T HepG2 e 293 senza citotossicità. Il **trattamento con GB-2** potrebbe ridurre il livello di espressione di ACE2 e TMPRSS2 del tessuto polmonare e del tessuto renale senza effetti avversi....Inoltre, nelle composizioni di GB-2, gli autori hanno scoperto la **teaflavina** potrebbe *inibire l’espressione proteica di ACE2 e TMPRSS2 e l’espressione dell’*mRNA* di ACE2*. In conclusione,....GB-2 e **teaflavina** potrebbero agire come *potenziali composti per gli inibitori di ACE2 e TMPRSS2*”.

In uno studio che impiega *cellule epiteliali polmonari infettate con SARS-CoV-2* (<https://www.nature.com/articles/s41598-020-78402-2>), gli autori hanno identificato i geni co-espressi con l'enzima 2 (**ACE2**) (noto come *enzima di conversione dell'angiotensina 1*), il **recettore di ingresso delle cellule umane di SARS-CoV-2**, e hanno scoperto diversi geni correlati o inversamente correlati con alta significatività. Tra questi geni è stato identificato **TMPRSS4**.

TMPRSS4 è una nuova **Serina Proteasi Transmembrana di tipo II** altamente espressa sulla superficie cellulare nei tessuti del pancreas, della tiroide, del colon e di altri tumori, è ancorata alla membrana plasmatica e induce direttamente l'elaborazione di **pro-uPA/PLAU** [*plasminogeno di tipo urochinasi (uPA)*; (PubMed: 24434139)] nella forma attiva attraverso l'attività proteolitica (<https://www.uniprot.org/uniprot/Q9NRS4>). **TMPRSS4** media l'invasione delle cellule tumorali (sovraregolando l'espressione genica di **uPA**), la transizione epiteliale-mesenchimale e le metastasi. L'aumento dell'espressione di **TMPRSS4** è anche correlato alla progressione del cancro del colon-retto. In sostanza, **TMPRSS4** potrebbe **aumentare l'invasione mediata da pro-uPA in modo dipendente dall'attività proteolitica della serina**. Queste osservazioni suggeriscono che **TMPRSS4** è un regolatore a monte dell'attivazione di pro-uPA (PubMed: 24434139).

TMPRSS4 sembra essere in grado di attivare **ENaC** [per somiglianza (<https://www.uniprot.org/uniprot/Q8VCA5>)]. L'**equilibrio del sodio** è mantenuto dalla regolazione precisa dell'**attività del canale epiteliale del sodio (ENaC) nel rene**. In uno studio (<https://pubmed.ncbi.nlm.nih.gov/12149280/>) è stata segnalata un'**attivazione extracellulare del trasporto di sodio mediato da ENaC (I (Na)) da una serina proteasi ancorata a GPI (mouse channel-activating protein, mCAP1)** che è stata isolata da una linea cellulare del condotto di raccolta corticale derivata dal rene del topo. In questo studio, gli autori hanno identificato **due ulteriori serina proteasi legate alla membrana (mCAP2 e mCAP3)** che sono espresse nella stessa linea cellulare. Ciascuna di queste **proteasi** è in grado di **aumentare I (Na) 6-10 volte** nel sistema di espressione degli *ovociti Xenopus*. Gli autori provano che **mCAP1** aumenta *I (Na)* di 10 volte ma il numero dei canali (N) è rimasto invariato, indicando che **mCAP1 regola l'attività ENaC aumentando la sua probabilità media di apertura dell'intera cellula**. Risultati simili sono stati osservati per **mCAP2** e **mCAP3**. Il sinergismo tra **CAP** e **Sgk1** (*chinasi sierica regolata da glucocorticoidi* coinvolta nella cascata del segnale dipendente dall'*aldosterone*) su **I (Na)** è sempre stato più che additivo, indicando un vero potenziamento (<https://pubmed.ncbi.nlm.nih.gov/12149280/>).

Infine, **TMPRSS4**, nelle *cellule epiteliali intestinali*, **facilita l'infezione da coronavirus umano SARS-CoV-2** attraverso, almeno, la **scissione delle glicoproteine spike del coronavirus, scissione che attiva la glicoproteina per l'ingresso nella cellula ospite** (<https://www.uniprot.org/citations/32404436>). I risultati di questo studio (<https://pubmed.ncbi.nlm.nih.gov/32404436/>) evidenziano l'intestino come un potenziale sito di replicazione della SARS-CoV-2, che può contribuire alla malattia locale e sistemica e alla progressione generale della malattia.

Un altro “**prodotto**” delle sequenze dei **Bacilli Paranthracis e Anthracis** identificato dalla ricerca BAST è la **Fosfoadenil-solfato reduttasi (tioredossina)**. Come riportato nell’articolo seguente (il link è mostrato nella seconda frase successiva), “Il **genoma del coronavirus (SARS-CoV-2)** recentemente identificato *codifica per varie proteine*, come la proteina dell’involucro, della membrana, del nucleocapside e dello spike. Tra queste, *la proteina necessaria al virus per entrare nella cellula ospite è la **proteina spike***”. In questa ricerca (<https://link.springer.com/article/10.1007/s11030-020-10171-6>), “...brevi *molecole peptidiche sintetiche* che si legano a una *specifica proteina bersaglio* con un’elevata affinità per esercitare la sua funzione, note come **aptameri peptidici**, sono state progettate e mirate al **dominio di legame del recettore spike di SARS-CoV-2**. L’**aptamero peptidico** è stato progettato utilizzando la **Tioredossina A batterica** come *proteina scaffold* e un *peptide lungo 18 residui*. La struttura terziaria dell’*aptamero peptidico* è modellata e ancorata al *dominio di legame del recettore spike della SARS CoV2*. L’*aptamero* si lega al *dominio di legame del recettore spike di SARS-CoV-2* in un **modello simile a quello di ACE2**. Si prevede che l’**aptamero** sia non antigenico, non allergenico, non emolitico, non infiammatorio, solubile in acqua con elevata affinità verso l’**ACE2** rispetto all’*albumina sierica*. Pertanto, il **peptide aptamer** può essere un nuovo approccio per il trattamento terapeutico per SARS-CoV-2”.

In aggiunta ai *prodotti* sopra menzionati, altri due “**prodotti**” delle sequenze del **Bacillus Cereus** identificati dalla ricerca BAST sono l’**Undecaprenil-difosfato fosfatasi (PRK00281)**, membro della *Superfamiglia BacA* [**Proteina di resistenza alla Bacitracina (BacA)**, una *nondecaprenolo putativa chinasi*] e la **Proteina della famiglia TIGR02687**.

In un recente studio (<https://pubmed.ncbi.nlm.nih.gov/33352458/>) sono stati identificati dei **farmaci** approvati dalla FDA e analoghi **nucleosidi** (in particolare **NA1**, un *composto di piombo*) come **potenziali inibitori del dominio SARS-CoV-2 A1pp**. I *coronavirus* possiedono un *macro dominio virale A1pp altamente conservato con siti di legame dell’**adenosina difosfato (ADP)-ribosio** e siti di attività della **fosfatasi***. **A1pp inibisce la ribosilazione dell’adenosina difosfato (ADP) nell’ospite e promuove l’infezione virale e la patogenesi**. Nel loro insieme, i risultati di questo studio propongono **NA1** come **inibitore principale del dominio SARS-CoV-2 A1pp per futuri test e per lo sviluppo come terapia contro il coronavirus umano**.

Con riferimento specifico al sopra menzionato **Undecaprenil-difosfato fosfatasi**, alcuni autori, studiando dei composti capaci di interferire con la **biosintesi della parete cellulare batterica**, hanno scoperto che i composti più attivi inibiscono due enzimi coinvolti nella biosintesi: **undecaprenil difosfato sintasi (UPPS)** e **undecaprenil difosfato fosfatasi (UPPP)**, ma non *farnesil difosfato sintasi*, e c’erano buone correlazioni tra l’**inibizione della crescita cellulare batterica** e l’**inibizione UPPS** e di **UPPP** (<https://chemistry-europe.onlinelibrary.wiley.com/doi/abs/10.1002/cmdc.201600342>).

Infine, la ricerca su **Pubmed** non ha evidenziato relazioni tra **SARS-CoV-2** e la **Proteina della famiglia TIGR02687**.

10.3 SEQUENZA 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 genome	Bacillus paranthracis	38.2	70.4	86%	2.3	100.00%	5351395
.....
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.	38.2	38.2	86%	2.3	100.00%	5210966

SVA11

CP066168.1	Bacillus anthracis strain ST11 chromosome, complete genome	Bacillus anthracis	38.2	38.2	86%	2.3	100.00%	5229731
OD002082.1	1_Tps_b3v08 Bacillus paranthracis strain Gxun-30 chromosome, complete genome	Timema poppensis	38.2	38.2	86%	2.3	100.00%	94204
CP065149.1	1_Tdi_b3v08	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 shepardii	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%	9723243 1
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: 21	Salmo trutta	34.2	64.4	81%	37	100.00%	52728040
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 shepardii	34.2	34.2	77%	37	100.00%	316
OC001206.1	2_Tsi_b3v08	Timema shepardii	34.2	34.2	77%	37	100.00%	182350
OC000708.1	2_Tsi_b3v08	Timema shepardii	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	Homo sapiens	34.2	34.2	77%	40	100.00%	169669

on chromosome 6, complete sequence

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

chromosome: 1

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
	PREDICTED: Myripristis murdjan acidic mammalian chitinase-like (LOC115359376), mRNA	pinecone soldierfish	32.2	32.2	72%	162	100.00%	1538
XM_030051815.1								
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%	2980965
CP025662.1	Ipomoea triloba cultivar NCNSP0323 chromosome 3	trilobed morning glory	32.2	32.2	72%	162	100.00%	3344819
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%	26089173
CP025646.1	Ipomoea trifida cultivar NCNSP0306 chromosome 3	NA	32.2	32.2	72%	162	100.00%	28619729
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%	21381654
CP012615.1	Oryza sativa Indica Group cultivar RP Bio-226 chromosome 7 sequence	long-grained rice	32.2	62.4	72%	162	100.00%	26660614
CP012614.1	Oryza sativa Indica Group cultivar RP Bio-226 chromosome 6 sequence	long-grained rice	32.2	32.2	72%	162	100.00%	30903862
CP012611.1	Oryza sativa Indica Group cultivar RP Bio-226 chromosome 3 sequence	long-grained rice	32.2	62.4	77%	162	100.00%	39504267
CP012609.1	Oryza sativa Indica Group cultivar RP Bio-226 chromosome 1 sequence	long-grained rice	32.2	94.6	72%	162	100.00%	44526634
HG975449.1	Solanum pennellii chromosome ch10, complete genome	NA	32.2	153	90%	162	100.00%	82529941
HG975448.1	Solanum pennellii chromosome ch09, complete genome	NA	32.2	306	100%	162	100.00%	84057508
HG975446.1	Solanum pennellii chromosome ch07, complete genome	NA	32.2	153	77%	162	100.00%	79292169
HG975444.1	Solanum pennellii chromosome ch05, complete genome	NA	32.2	157	95%	162	100.00%	77991103
HG975443.1	Solanum pennellii chromosome ch04, complete genome	NA	32.2	183	95%	162	100.00%	77197300
HG975442.1	Solanum pennellii chromosome ch03, complete genome	NA	32.2	122	100%	162	100.00%	75414019
CP056063.1	Oryza sativa Indica Group cultivar Zhenshan 97 chromosome 12	Oryza sativa Indica Group	32.2	32.2	72%	144	100.00%	27232431
CP056062.1	Oryza sativa Indica Group cultivar Zhenshan 97 chromosome 11	Oryza sativa Indica Group	32.2	92.7	95%	144	100.00%	32624309
CP056058.1	Oryza sativa Indica Group cultivar Zhenshan 97 chromosome 7	Oryza sativa Indica Group	32.2	94.6	72%	144	100.00%	30840550
CP056057.1	Oryza sativa Indica Group cultivar Zhenshan 97 chromosome 6	Oryza sativa Indica Group	32.2	32.2	72%	144	100.00%	32119910
CP056054.1	Oryza sativa Indica Group cultivar Zhenshan 97 chromosome 3	Oryza sativa Indica Group	32.2	62.4	77%	144	100.00%	39636162
CP054687.1	Oryza sativa Indica Group cultivar Minghui 63 chromosome 12	Oryza sativa Indica Group	32.2	32.2	72%	144	100.00%	26942889
CP054686.1	Oryza sativa Indica Group cultivar Minghui 63 chromosome 11	Oryza sativa Indica Group	32.2	92.7	95%	144	100.00%	34100580
CP054683.1	Oryza sativa Indica Group cultivar Minghui 63 chromosome 8	Oryza sativa Indica Group	32.2	32.2	72%	144	100.00%	30492302
CP054682.1	Oryza sativa Indica Group cultivar Minghui 63 chromosome 7	Oryza sativa Indica Group	32.2	94.6	72%	144	100.00%	30877072
CP054681.1	Oryza sativa Indica Group cultivar Minghui 63 chromosome 6	Oryza sativa Indica Group	32.2	32.2	72%	144	100.00%	31921180
CP054678.1	Oryza sativa Indica Group cultivar Minghui 63 chromosome 3	Oryza sativa Indica Group	32.2	62.4	77%	144	100.00%	39893253
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 monikensis	32.2	32.2	72%	144	100.00%	7147
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 shepardii	32.2	32.2	72%	144	100.00%	30778
OC004557.1	2_Tsi_b3v08	Timema shepardii	32.2	32.2	72%	144	100.00%	63513
OC004547.1	2_Tsi_b3v08	Timema shepardii	32.2	32.2	72%	144	100.00%	63626
OC003968.1	2_Tsi_b3v08	Timema shepardii	32.2	32.2	72%	144	100.00%	74580
OC002709.1	2_Tsi_b3v08	Timema shepardii	32.2	32.2	72%	144	100.00%	106344
OC000605.1	2_Tsi_b3v08	Timema shepardii	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 (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	32.2	32.2	72%	144	100.00%	2174
MW109287.1	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="MKTKDQIDNIVKETYRHHDFSGVILVKEKKGIVYEEVFGYANRN
ECINNTLQTRFGIASGCKIFTAIGICQLVENGVITFQTKLKECLSVNFPNFNEDITIH
HLLTHSSGIPDYFDESIMDNFEDLWKQTPMYLLKSLKDFLPLFQNRDMMYAPGSKFHY
NNAGFIILGLIIEEQTGLTFTEYIEKNIFNPIGMNHSYFSLDRLPRQTALGYIKDEI
SQTWRTNAYSIPKGGSDGGAFITASDMLKFWEALFNYEIISQYTKILLTPHIQVNN
NQSYYGYGIWIETRENKIFKYHVMGYDPGVSFSAVYPDLGITLVIPSNKGAGAELMI
EIEGSF"

```

ORIGIN

1 **tatccatta catgatatt**

//

Bacillus anthracis strain FDAARGOS_702 chromosome

Sequence ID: [CP054800.1](#) Length: 5272559 Number of Matches: 1

Range 1: 208258 to 208276 [GenBank](#) [Graphics](#)

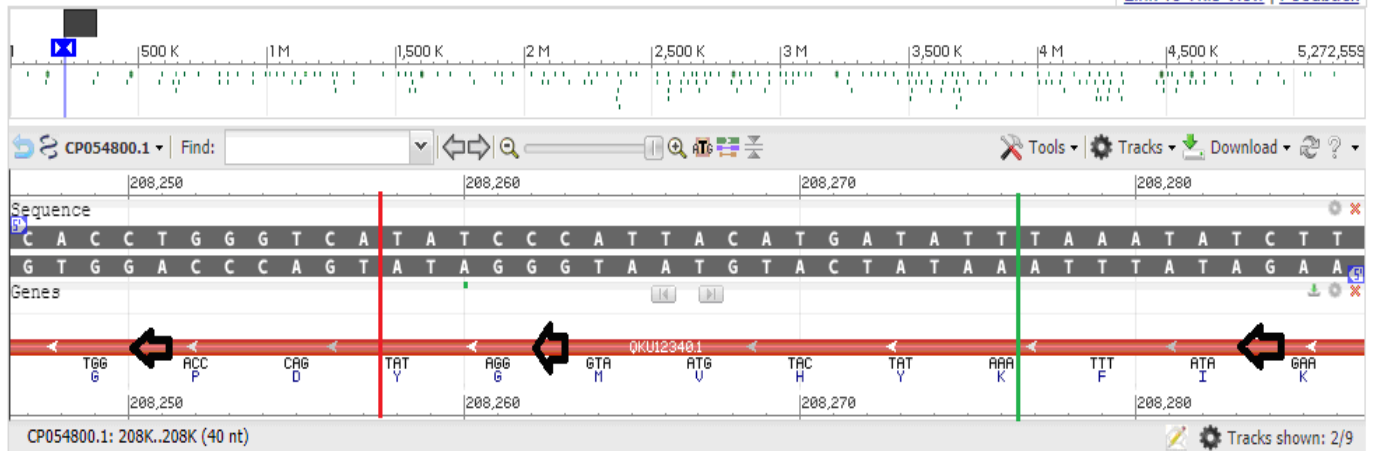
Score	Expect	Identities	Gaps	Strand
38.2 bits(19)	2.2	19/19(100%)	0/19(0%)	Plus/Plus
CDS: Putative 1	5	G M V H Y		
Query	1	TATCC CATTAC ATGATATT	19	
Sbjct	208258	TATCC CATTAC ATGATATT	208276	
CDS: serine hydrolase	301	Y G M V H Y K		

Bacillus anthracis strain FDAARGOS_702 chromosome

GenBank: CP054800.1

[GenBank](#) [FASTA](#)

[Link To This View](#) | [Feedback](#)



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
NNAGFIILGLIIIEEQTGLTFTEYIEKNIFNPIGMNHSYFSLDRLRQ TALGYIKDEI
SQTWR TNAYS IPIKGGSDGGAFITASDMLKFWEALFNYEIISQ EYTKILLTPHIQVNN
NQS YGYGIWIETRENKIFKYHVMGYDPGV SFRSAVY PDLGITLVIPSNKGAGAEK LMI
EIEGSF"

```

ORIGIN

1 **aatatcatgt aatgggata**

//

ATTENTION:

The sequence “**aatatcatgt aatgggata**” shown above is the *reverse and mirror sequence* of the sequence “**TATCCATTACATGATATT**” (partial sequence of **Sequence n° 3/1**, “**TATCCATTACATGATATTATT**”):

ATAGGGTAATGTACTATAA

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

TATCCATTACATGATATT

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

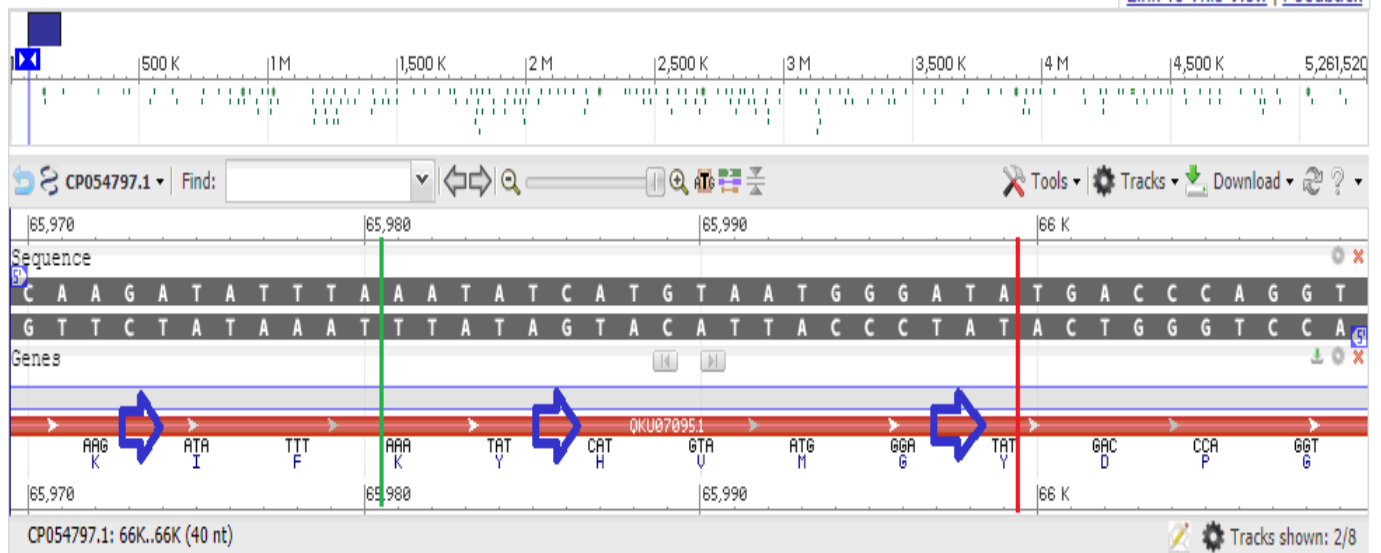
CDS: Putative 1	5	G	M	V	H	Y		
Query	1	TATCC	CATTAC	ATGATATT	19			
Sbjct	65999	TATCC	CATTAC	ATGATATT	65981			
CDS: serine hydrolase	301	Y	G	M	V	H	Y	K

Bacillus anthracis strain FDAARGOS_703 chromosome

GenBank: CP054797.1

[GenBank](#) [FASTA](#)

[Link To This View](#) | [Feedback](#)



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	TATCCATTACATGATATT	19	
Sbjct	2708055	TATCCATTACATGATATT	2708037	
CDS: serine hydr	301	Y G M V H Y K		

Bacillus paranthracis strain NCCP 15910 chromosome, complete genome

GenBank: CP041981.1

[GenBank](#) [FASTA](#)

[Link To This View](#) | [Feedback](#)



Range 2: 1645460 to 1645475 [GenBank](#) [Graphics](#)

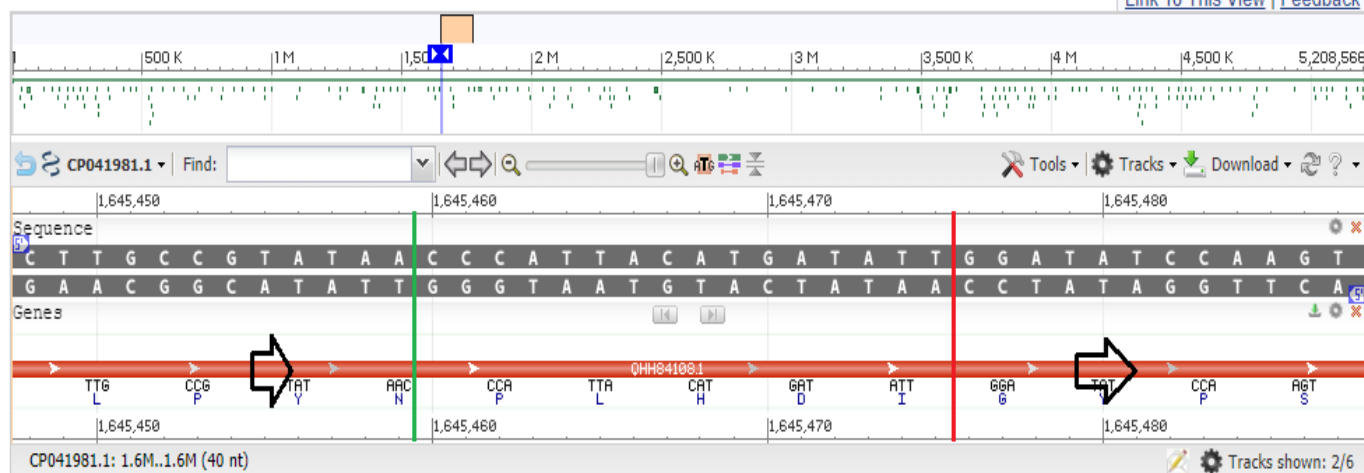
Score	Expect	Identities	Gaps	Strand
32.2 bits(16)	144	16/16(100%)	0/16(0%)	Plus/Plus
CDS: Putative 1	1	P L H D I		
Query	4	C CCATTACAT GAT AATT	19	
Sbjct	1645460	C CCATTACAT GAT AATT	1645475	
CDS: phosphoadenylyl-	192	P L H D I		

Bacillus paranthracis strain NCCP 15910 chromosome, complete genome

GenBank: CP041981.1

[GenBank](#) [FASTA](#)

[Link To This View](#) | [Feedback](#)



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
SVWEQLIDLPPETVLAVLKKSTEEAARPSEARPLVEAERASGITDFPKIFCETSQKL
LSVEAFALAAKYYSLQNLGVCTPLEQLLVALRGNQTQKTGENVKPDEFMNLKKSHEVQ
AMSELISSIADYYGIKQVIDLGSGKGYLSSFLSLKYGLKVYIGIDSSNTNTHGAQERNR
KLKHKHWKICHTQSKSDVNLGALKMAKGKKMQGEIPYKANTEEVCNNSPTSQEMMSTSE
FLPDFSVSVISNIRKQMENLHVCSHREENLCFANAFSLIDLPLIDAIPTSSSRMPKR
KMLEANKEERRKTTKSTESNIYSPLTSFVTADSELHDIKDLLEDCLMVGLHTCGDLAP
NTLRIFTSKSEIKGVCSVGCYHLLSEEFEDPHKECTQEKWGFPMCHYLKEERWCCGR
NARMSACLALERVAVGRGLPTESLFYRAVLQDIKDCYGITKCDRHVKGKIYSKSSSFL
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	A TTACAT GAT ATT ATT	22	
Sbjct	1140	A TTACAT GAT ATT ATT	1155	
CDS:methyltransferas	369	L H D I I		

Metiltransferasi like 25

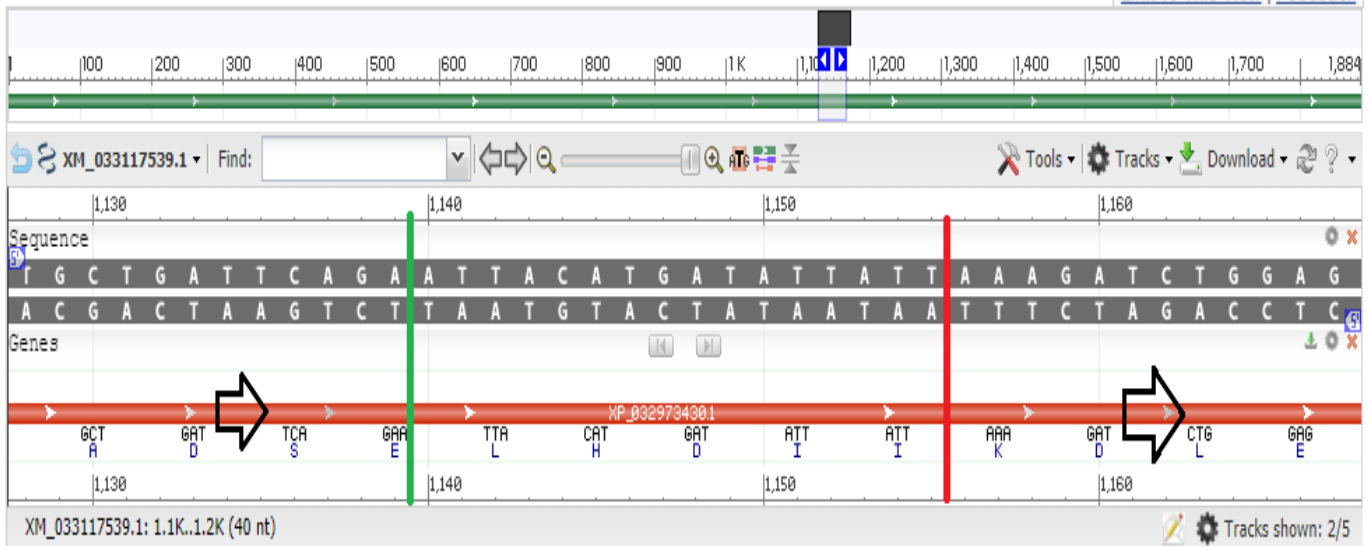
Una **Metiltransferasi** è un enzima appartenente alla classe delle *transferasi*, in grado di trasferire un *gruppo metile* da una *molecola donatrice* ad un *accettore*. **Metiltransferasi like 25 (METTL25)** è un *Gene Codificante Proteine*. Le malattie associate al METTL25 includono [Optic Atrophy 7 With Or Without Auditory Neuropathy](#). Le annotazioni di *Ontologia Genica (GO)* relative a questo gene includono *l'attività metiltransferasi*. Un importante paralogo di questo gene è **RRNAD1** ([https://www.genecards.org/cgi-bin/carddisp.pl?gene=METTL25#:~:text=METTL25%20\(Methyltransferase%20Like%20is,of%20this%20gene%20is%20RRNAD1\)](https://www.genecards.org/cgi-bin/carddisp.pl?gene=METTL25#:~:text=METTL25%20(Methyltransferase%20Like%20is,of%20this%20gene%20is%20RRNAD1))). **RRNAD1 (Ribosomal RNA Adenine Dimethylase Domain Containing 1)** è un *Gene Codificante Proteine*. Le annotazioni di *GO* relative a questo gene includono *rRNA (adenine-N6,N6)-dimethyltransferase activity* (<https://www.genecards.org/cgi-bin/carddisp.pl?gene=RRNAD1>). Vedi anche quanto discusso alle pagine 208 e 433.

PREDICTED: *Rhinolophus ferrumequinum* methyltransferase like 25 (METTL25), transcript variant X2, mRNA

NCBI Reference Sequence: XM_033117539.1

[GenBank](#) [FASTA](#)

[Link To This View](#) | [Feedback](#)



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
SVWEQLIDLPPETVLAVLKKSTEEAARPSEARPLVEAERASGITDFPKIFCETSQKL
LSVEAFALAAKYYSLQNLGVCTPLEQLLVALRGNQTQKTGENVKPDEFMNLKKSHEVQ
AMSELISSIADYYGIKQVIDLGSGKGYLSSFLSLKYGLKVYGIDSSNTNTHGAQERNR
KLKKHWKICHTQSKSDVNLGALKMAKGKKMQGEIPYKANTEEVCNNSPTSQEMMSTSE
FLPDFSVSVISNIRKQMENLHVCSHREENLCFANAFSLIDLPIAIEPTSSSRMPKR
KMLEANKEERRKTTSKSTESNIYSPLTSFVTADSELHDI IKDLEDCLMVGLHTCGDLAP
NTLRIFTSKSEIKGVCSVGCCYHLLSEEFEDPHKECTQEKWGFPMCHYLKEERWCCGR
NARMSACLALERVAVGRGLPTESLFYRAVLQDIIKDCYGITKCDRHVGKIYSKSSSFL
DYVRKSLRKLGLDESKLPEKIMDYEYKYPKMNELEAFNMLKVVLAPCIETLILLDR
LCYLKEQEDIAWSALVKLFDVPKSPRCYAVIALKQQQ"

```

ORIGIN

1 attacatgat attatt

//

PREDICTED: Rhinolophus ferrumequinum methyltransferase like 25 (METTL25), transcript variant X1, mRNA
Sequence ID: [XM_033117538.1](#) Length: 1994 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		

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 ttttggtttg ctccacatc agttctccat
61 tcctcgctgt cctgggctct gaggccgact gtacagacca aaccacacg ggcccctcac
121 cctttggttt cccacacact cagccaaag gaggctggaa ggggatgtgg gggctgagga
181 gggagaggcg gggcttgttc cctccccacc tgggtgcaac tcggcagggg accctccaac
241 cttggccaca gctgccgagc cacaccacgc cgttccctct ccggtcccct caggcctgga
301 ggggactatc ccaaggttcc caccgccctt tgtttctgta accctgcaa ccccttcagt
361 gaactttctt taattatcac tcagtttagt ttaatgtaa cagtcccttc attaaactaa
421 ttaccacttt ttgttttcca tttctttcct ccaggatgc taacagacac acctgcctcc
481 catctaccta tcaaatcttg tgtggaactc tcaactcaa gctgaaccct tgctgtgaga
541 tttgctaagt gtaccagccc atcctgagct gtcctggaga gaattcaaca gtttactaa
601 aagagataga ctttcccttg ggtatgtagt gccacctgct ggacatccag agaaacagcg
661 ccatctagct ttgaagtccc tgatgaagaa agcccacttg gaaggtagca atcactcttc
721 aatgaattg acttatgaag aaaaaaaca cagcatgaaa tagagaactt taaatgagga
781 ccagaaaaga aggaaaggag aatgttaaaa agtataagag tacaaggcga gtctccttct
841 tccagagata aagaaatgga gatattcaga tttcttaaaa gggcccttcc tgtcctgcca
901 catatgctgt attatgatta tgttgttcat ttcctcattt tcattacttc tgtcctcttg
961 gttaagtagc ctgattctaa agacctaagg gctagaacta acactgacaa gatggaaaaa
1021 aagctcttga tgtgcactta gctgctccaa atgaacatgc aaacaggatc tatggttctc
1081 aaagcccact gatgagataa ttcattcaat atcaaaaatc ctttggaat cttttgggta
1141 gaattctgcc ctactatta ttatacaata ggcacacccc tacagtcgtt tcatgtgtaa
1201 ttctaactct tcataaaaatc tcttaataat ttacaaatga ggaacaaaag gctcaaaaga
1261 gttaacaac ttocctcaagc ttctagtac ataagtggtg gagctgggct ttgaatccag
1321 gtctctgtga tacaaaaaaa aaaacccaaa aactatattt ttctatatg ttaagctgcc
1381 tttcagaacc ctactactt cagatcaaga gcttccaata tgcaaggcat agggctagat
1441 gcattacctc atttagtcac agagtaatgc atttgagttc attatcatgg tcatttgata
1501 ttgacattgc tagcaagatt taaaaccag taattaaata ggttttgagc aatggggaaa
1561 ggattcaatc actgaaaact agtaaggctt tgctatgagt aatcatgcc tgcctatctt
1621 cacttctttt taaatttggc caacagatga tagagaaatg gtggagttat ttagacaggg
1681 atttgaaaaa gtctttctta tgatcctctt ttatgaaca cctgcactac atagaataat
1741 aatatcatgt aatccagtga gctccacca aagagctttg actgatgtac tgacttctca
1801 gatcagctct ttttgattta ttaatttgc aggatgagga catggatgat atcaactctc
1861 ctgatcctct gaagctgaga ggaatagctg atatatacga atgaacctag tccaagacat

```

```

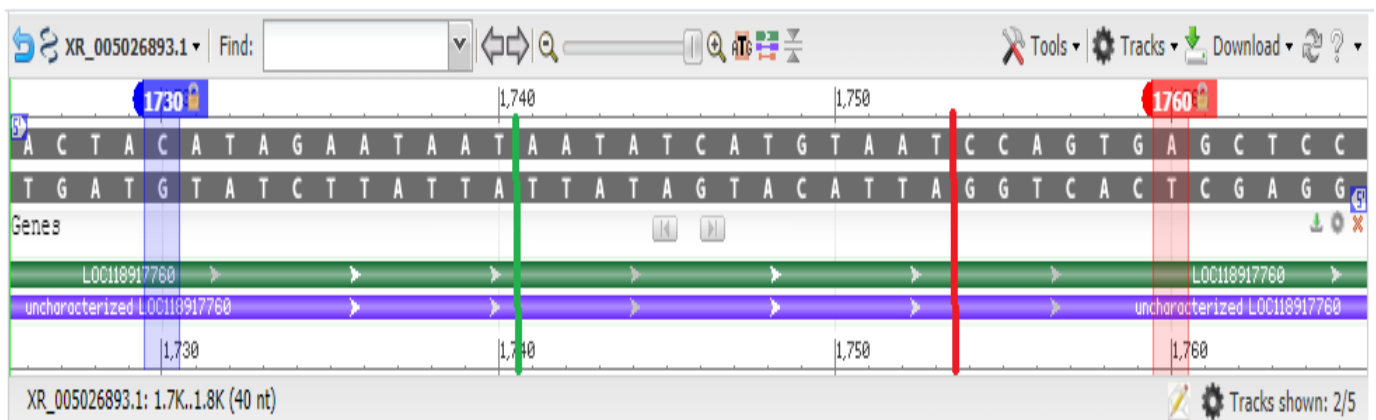
1921 tgtcaacaca cgggatgat gagcagatcc taatgagatg aagtttaaca gaactgaaag
1981 ggaggtoctc cactatgtcc caggaaaaag tgctcaaggg agacgaacaa atcatttata
2041 tgaaaaggct agcaatggca gaaagaatga actacagtgt agtcagaatc tgaaggaggc
2101 ggttcatggc ccacttctc cacttgttca caggggagct cctgctgggc tttcaagggg
2161 actcacggga cacatgctaa 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 gtggagggga aattctggca ttcagtggag gcttgatcgt attactgcta aaacccttc
2581 taagccagag actctacaat actctgcaaa tgaaactgtg catcttctaa gtataagatg
2641 caggacata ctaaaagaag agaagaagaa agtgatgagt ccagtcagat gaaaaaagtg
2701 ttgtgctggc tgagcgaaga tgacataacc ctctgcaagc ttgctggaca cagttacggt
2761 ggttttggac tcggcttgca gacagatatg gtggccatca aacccccatt gatggacggt
2821 gctgtaaagt cacattgctc actacttggg aaagccttca tgttttcaa aagagacaaa
2881 ctttcaatc ctttaatta tattattagg tgggctctta atggcagcca gcttggcct
2941 tgcatttgc tgcaccctt tattccagac ttaagccct tcagtcactg aaacttgagc
3001 cacttatggt acaatcaagc cacaaagaga acaggggaaa atggatgac ttctttact
3061 ttgaagagat gacttcatca ccagtgggta 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="MEDFVRQCFNPMIVELAETMKEYGESPKIETNKFAAICTHLEV
CFMYSDFHFINERGESIIAESGDPNALLKHRFEIIEGRDRTMAWTVVNSICNTTGVDK
PKFLPDLYDYKENRFIEIGVTRREVVHIYYLEKANKIKSERTHIHIFSFTGEEMATKAD
YTLDEESRARIKTRLFTIRQEMASRGLWDSFRQSERGEXTIEERFEITGTMRLADQS
LPPNFSSLENFRAYVDGFEPNGCIEGKLSQMSKEVNARIEPFXKTPRPLRLPDGPPC
SQRSKFLMDALKLSIEDPSHEGEGIPLHDAIKCMKTFFGWKEPNIVKPHDKGINPNY
LLAWKQVLAELQDIENEDKIPKTKNMKTSQLKWALGENMAPEKVDVFEDCKDVSDLKQ
YDSDEPEPRSLSSWVQSEFNKACELTDSSWIELDEIGEDVAPIEHIASMRRNYFTAEV
SHCRATEYIMKGVYINTALLNASCAAMDDFQLIPMISKCRTREGRKTNLYGFIIKGR
SHLRNDTDVVNFVSMESLTDPRLEPHKWKVCVLEIGDMLLRTAIGQVSRPMPFLYVR
TNGTSKIKMKWGMEMRRCLLQSLQQIESMIEAESSVKEKDMTKEFFETKSETWPIGES
PKGVEESSIGKVCRTLLAKSVFNLSYASPQLEGFSAESRLLLLIVQALRDNLEPGTFD
LGGLYDAIEECLINDPWVLLNASWFNSFLTHALK"

```

ORIGIN

1 **tatccatta catgat**

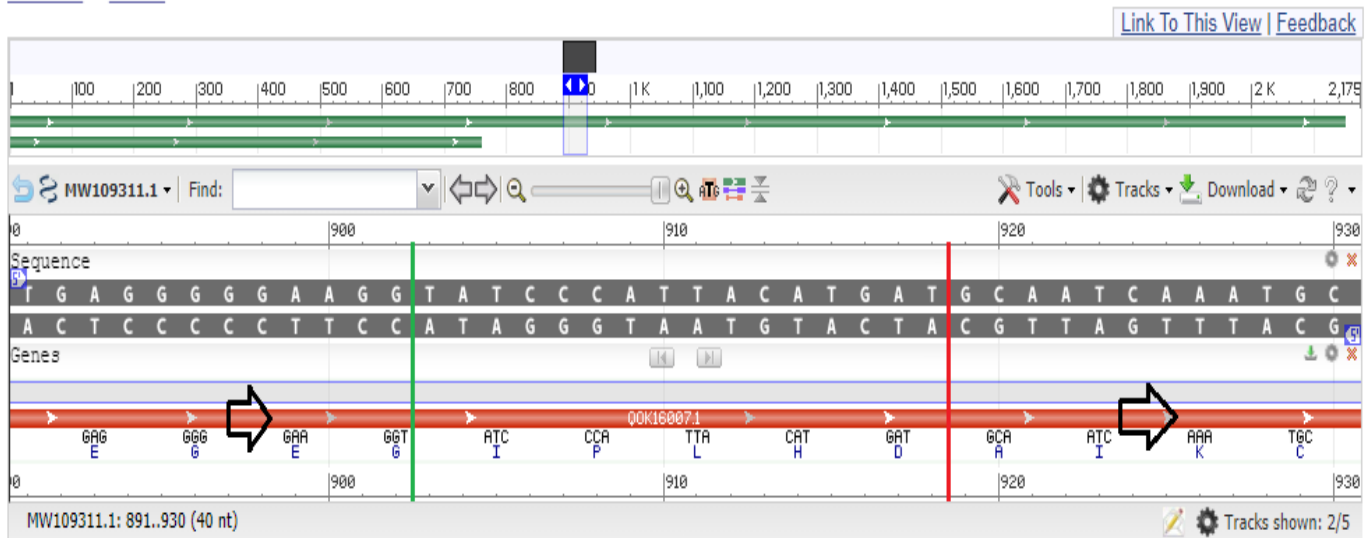
//

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	TATC CCATTACAT GAT	16	
Sbjct	903	TATC CCATTACAT GAT	918	
CDS: polymerase PA [I	302	I P L H D		

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](#)
[GenBank](#) [FASTA](#)



Polimerasi PA (PA) - Proteina Acida della Polimerasi

La **Proteina Acida della Polimerasi** (o **RNA-directed RNA polymerase subunit P2**) del **gene PA** svolge un ruolo essenziale nella *trascrizione* e nella *replicazione dell'RNA virale* formando il *complesso eterotrimerico della polimerasi insieme alle subunità PB1 e PB2*. Il *complesso trascrive gli mRNAs virali* utilizzando un meccanismo unico chiamato **cap-snatching**. Consiste nel dirottamento e nel clivaggio di *pre-mRNA* ricoperti dall'ospite. Questi RNA con cappuccio corto vengono quindi utilizzati come **primer per mRNA virali**. La **subunità PB2** è responsabile del legame del cappuccio 5' dei *pre-mRNA* cellulari che vengono successivamente scissi dopo 10-13 nucleotidi dalla subunità **PA** che trasporta l'attività endonucleasica (<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	TATCCCATACATGAT 16		
Sbjct	903	TATCCCATACATGAT 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	TATCCCATACATGAT 16		
Sbjct	903	TATCCCATACATGAT 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	TATCCCATACATGAT 16		
Sbjct	903	TATCCCATACATGAT 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 I PRINCIPALI “PRODOTTI” IDENTIFICATI DALLE RICERCHE BLAST EFFETTUATE SULLA **SEQUENZA N° 3/1** (**Hitlist size: 5000**)

Oltre i “prodotti” delle sequenze dei **Bacilli Paranthracis, Anthracis e Cereus** identificati dalle precedenti ricerche BLAST, altri due “prodotti” sono stati considerati molto interessanti ai fini di questo lavoro: “**Metiltransferasi like 25**” e “**Polimerasi PA (PA) - Proteina Acida della Polimerasi**”.

In uno studio recente (<https://pubmed.ncbi.nlm.nih.gov/33685335/>), gli autori mirano a identificare un potenziale *nuovo inibitore per SARS COV-2 2'-O-metiltransferasi (nsp16)*, che è uno degli obiettivi più attraenti nel ciclo di vita del virus, responsabile della *protezione dell'RNA virale* attraverso un processo di formazione del cappuccio. I risultati hanno rivelato il *composto 11* come il *miglior potenziale inibitore nsp16*, in quanto ha mostrato una migliore stabilità e l'energia media libera di legame per il complesso enzima-ligando rispetto a *Sinefungin*.

In un altro studio di identificazione in silico e convalida di *composti antivirali naturali* come *potenziali inibitori per la metiltransferasi del SARS-CoV-2* (<https://pubmed.ncbi.nlm.nih.gov/33583328/>), gli autori considerano i **prodotti botanici** una risorsa importante per diversi **agenti antivirali efficaci** ed evidenziano che i *composti naturali* stanno guadagnando un'attenzione significativa per il *trattamento della COVID-19*. Gli autori hanno trovato quattro *composti antivirali naturali* **Amentoflavone, Baicalin, Daidzin e Luteoloside** come *potenti inibitori per la metiltransferasi di SARS-CoV-2*. La loro analisi suggerisce che esiste un'alta probabilità che questi composti possano avere un buon potenziale per inibire la metiltransferasi (MTasi) della SARS-CoV-2 e per essere utilizzati nel trattamento del COVID-19.

In uno studio approfondito (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7217786/>), alcuni ricercatori discutono delle strutture, delle caratteristiche di trasmissione, dei sintomi clinici, della diagnosi, dei cambiamenti patologici, del trattamento e della prevenzione di **due tipi di virus, CoV (SARS e MERS) e IAV (Virus dell'Influenza A)**. “La malattia respiratoria che ha causato la COVID-19 ha sintomi clinici simili ad altre **due malattie CoV**, la *sindrome respiratoria acuta grave* e la *sindrome respiratoria mediorientale (SARS e MERS)*, di cui i virus causali sono la SARS-CoV e MERS-CoV, rispettivamente. Queste tre malattie che ne derivano condividono anche molti sintomi clinici con altre malattie respiratorie causate da IAV”. Poiché sia i CoV che gli IAV sono patogeni generali responsabili del freddo stagionale, durante il cambio delle stagioni, i medici e la salute pubblica potrebbero dover distinguere la **polmonite COVID-19** da **altri tipi di polmonite virale** (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7217786/>). A tal proposito, è fondamentale anche una farmacoterapia idonea. Attualmente, per il *trattamento clinico dell'infezione influenzale*, sono disponibili solo **tre classi di farmaci antivirali approvati dalla FDA (Food and Drug Administration)**, inclusi gli **inibitori dei canali ionici M2** (amantadina e rimantadina), gli **inibitori della neuraminidasi** (oseltamivir, zanamivir, peramivir e laninamivir) e i più recenti **inibitori della RNA**

polimerasi (*balxavir marboxil*). “**Balxavir marboxil**, con un marchio di Xofloza, può mirare alla **proteina acida della polimerasi virale (PA)**, bloccarne la funzione endonucleasica, portando all’inibizione della trascrizione dell’*mRNA* del virus e **prevenire efficacemente l’infezione da virus dell’influenza A**. Lo spettro antivirale di Xofloza include **ceppi influenzali stagionali e ceppi influenzali resistenti a “oseltamivir”** [*un inibitore della neuraminidasi (NAI)*], riducendo significativamente la carica virale rispetto a *oseltamivir*. Anche **T-705**, un nuovo tipo di inibitore della RNA polimerasi, può inibire un’ampia gamma di virus entrando direttamente nella catena dell’RNA virale o legandosi direttamente al dominio dell’RNA polimerasi virale per bloccare la replicazione e la trascrizione della catena dell’RNA virale” (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7217786/>).

CAPITOLO XI

**RICERCHE BLAST
E DISCUSSIONE DEI
RISULTATI OTTENUTI
DAGLI ALLINEAMENTI
DELLA SEQUENZA N° 8/1**

OBIETTIVI DEL CAPITOLO XI

In questo capitolo sono riportati i risultati degli *allineamenti significativi* della ricerca BLAST effettuata sulla **Sequenza n° 8/1**. Questa sequenza è una delle tante generate dalla **Sequenza Originaria** [la sequenza “GTGAAATGGTCATGTGTGGCGG” del Gene RdRp (*RNA-dependent RNA polymerase*, RdRp_SARSr-F)] e dalla quale differisce per il **90,90 %** delle basi.

I risultati degli allineamenti significativi della ricerca BLAST (**Hitlist size 100**) hanno riguardato due specie di pipistrelli (*Rhinolophus ferrumequinum* e *Pipistrellus pipistrellus*) e molti “**organismi**” comuni (distinguibili dai colori differenti nelle pagine successive) agli “**organismi**” identificati dalle ricerche BLAST effettuate sulle sequenze a cui è stata **sostituita una base alla sequenza originaria** (si vedano i Capitoli V e VI). Gli “organismi” comuni identificati sono: **Chanos chanos**; **Ipomoea triloba**; **Ipomoea Trifida**; **Coregonus sp. 'balchen'** e **Salmo trutta**.

ATTENZIONE: L'enorme mole di lavoro che è stato fino ad adesso necessario per analizzare e selezionare tutte le sequenze significative delle diverse ricerche BLAST mostrate nei capitoli precedenti rende difficile poter proseguire velocemente con tale meticolosità. La necessità di pubblicare al più presto i risultati di questo ampio documento, mi costringe, da questa pagina in poi, a lasciare allo scienziato interessato l'arduo compito (che consiglio assolutamente di eseguire) di approfondire l'analisi e la selezione delle sequenze ritenute rilevanti secondo le indicazioni proposte nei capitoli precedenti.

Per i motivi appena esposti, le ricerche BLAST di approfondimento avranno un'estensione limitata.

Estendendo la ricerca BLAST (**Hitlist size 250**) è stata individuata un'altra sequenza del *Pipistrellus pipistrellus* e un numero considerevole di sequenze dei **Bacilli Paranthracis** e **Anthraxis**.

SEQUENZA N° 8/1

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

11.1 SEQUENZA N° 8/1 (T=1; C=0; G=3; A=2) – (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° 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
Sequenza 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: **ACATTTCTGGATGCACACAATAA**

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

XM_034397081.1	homocysteine S-methyltransferase-like (LOC106090333), mRNA 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 PREDICTED: Mastacembelus armatus endoplasmic reticulum membrane sensor NFE2L1-like (LOC113126748), mRNA	river trout	34.2	34.2	77%	35	100.00%	44893998
XM_026300864.1	PREDICTED: Drosophila willistoni chromosome-associated kinesin KIF4A (LOC6648886), mRNA	zig-zag eel	34.2	34.2	77%	35	100.00%	3081
XM_002071777.3	Drosophila melanogaster strain sitter (fors) chromosome 2R	Drosophila willistoni	34.2	34.2	77%	35	100.00%	4044
CP023332.1	PREDICTED: Cyprinus carpio uncharacterized LOC109080287 (LOC109080287), transcript variant X2, mRNA	fruit fly	34.2	34.2	77%	35	100.00%	21146707
XM_019095370.1	PREDICTED: Cyprinus carpio uncharacterized LOC109080287 (LOC109080287), transcript variant X1, mRNA	common carp	34.2	34.2	95%	35	95.24%	1325
XM_019095369.1	PREDICTED: Drosophila kikkawai uncharacterized LOC108071571 (LOC108071571), mRNA	common carp	34.2	34.2	95%	35	95.24%	1325
XM_017162352.1	Spirometra erinaceieuropaei genome assembly S_erinaceieuropaei, scaffold SPER_scaffold0029177	Drosophila kikkawai	34.2	34.2	77%	35	100.00%	1341
LN029746.1	Drosophila melanogaster chromosome 2R	Spirometra erinaceieuropaei	34.2	34.2	77%	35	100.00%	7894
AE013599.5	Phaseolus vulgaris hypothetical protein (PHAVU_009G082600g) mRNA, complete cds	fruit fly	34.2	34.2	77%	35	100.00%	25286936
XM_007136828.1		Phaseolus vulgaris Notodromas monacha	34.2	34.2	77%	35	100.00%	5319
OA888277.1	Notodromas monacha PREDICTED: Nematolebias whitei nuclear factor, erythroid 2-like 3 (nfe2l3), mRNA		34.2	34.2	77%	35	100.00%	13315
XM_037692747.1	Danio rerio genome assembly, chromosome: 17	Rio pearlfish	34.2	34.2	77%	35	100.00%	2051
LR812054.1	Danio rerio strain Cooch Behar (CB) genome assembly, chromosome: 17	zebrafish	34.2	34.2	77%	35	100.00%	54645068
LR812560.1	Carnobacterium sp. 17-4, complete genome	zebrafish	34.2	34.2	77%	35	100.00%	50025757
CP002563.1	Medicago truncatula clone mth2-133o10, complete sequence	Carnobacterium sp. 17-4	34.2	34.2	77%	35	100.00%	2635294
AC146941.8	Danio rerio genome assembly, chromosome: 19	barrel medic	34.2	34.2	77%	35	100.00%	134901
LR812081.1	Danio rerio genome assembly, chromosome: 9	zebrafish	34.2	34.2	77%	35	100.00%	49536214
LR812071.1		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 PREDICTED: Salmo trutta coiled-coil domain-containing protein 85C-A-like (LOC115172326), transcript variant X3, mRNA	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 X2, 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 X1, 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		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
	Salmo trutta genome assembly,	river trout	34.2	34.2	77%	35	100.00%	45589842

LR584423.1	chromosome: 31								
1/2 2/2 3/3 3/4 3/6									
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	Solanum tuberosum cultivar P8 chromosome 2	potato	30.2	60.5	68%	547	100.00%	48611948	
1/2 2/2 3/2 3/3 3/5 6/1 9/1 10/1									

VISTA DI ALLINEAMENTO VERTICALE AGGANCIATO A QUERY CON LETTERE D'IDENTITA'

Query range 1: 1 to 22

Query	1	ACATTTTCGGATGCACACAATAA	22
LR884471.1	5872729	ACATTTTCGGATGCACACAAT	5872748
XM 030786843.1	1749	ACATTTTCGGATGCACACAA	1731
LR697115.1	19324111	ACATTTTCGGATGCACACAA	19324093
LR812078.1	6416688	CATTTTCGGATGCACACAA	6416671
XM 017289635.2	165	ACATTTTCGGATGCACACA	148
XM 001355143.4	171	ACATTTTCGGATGCACACA	154
XM 031276121.1	386	ATTTTCGGATGCACACAAT	369
XM 002022755.2	130	ACATTTTCGGATGCACACA	113
CP025660.1	13048241	ATTTTCGGATGCACACAAT	13048224
CP032234.1	29284203	ACATTTTCGGATGCACACA	29284186
CP032234.1	30390915	TTTCGGATGCACACA	30390901
CP032595.1	11460181	CATTTTCGGATGCACACAA	11460164
CP027278.1	9140189	CATTTTCGGATGCACACAA	9140172
CP026254.1	7772702	ACATTGCGGATGCACACAATAA	7772681
XM 019320759.1	469	ATTTTCGGATGCACACAAT	452
XM 019320758.1	449	ATTTTCGGATGCACACAAT	432
XM 019320757.1	453	ATTTTCGGATGCACACAAT	436
XM 013256488.1	851	ATTTTCGGATGCACACAAT	834
XM 034397081.1	826	ACATTTTCGGATGCACACA	843
XM 034397072.1	826	ACATTTTCGGATGCACACA	843
XM 034397054.1	847	ACATTTTCGGATGCACACA	864
LR738627.1	9776344	ACATTTTCGGATGCACACA	9776361
LR738627.1	43382214	GGATGCACACAATAA	43382200
LR738607.1	12873654	ACATTTTCGGATGCACACA	12873671
LR738607.1	65672302	GGATGCACACAATAA	65672316
CP025644.1	8402783	ATTTTCGGATGCACACAAT	8402800
LT594792.1	27997305	ATTTTCGGATGCACACAAT	27997322
LT594792.1	7456216	GGATGCACACAATAA	7456202
LR812514.1	16160607	ACATTTTCGGATGCACACA	16160624
JN949980.1	27127	TTTCGGATGCACACAATAA	27144
JN949979.1	27177	TTTCGGATGCACACAATAA	27194
JN949026.1	11815	TTTCGGATGCACACAATAA	11832
JN949025.1	11815	TTTCGGATGCACACAATAA	11832
AF345929.1	185	TTTCGGATGCACACAATAA	202
AL662812.12	58729	TTTCGGATGCACACAATAA	58746
LR792833.1	3481679	TTTCGGATGCACACAATA	3481663
LR761618.1	3481679	TTTCGGATGCACACAATA	3481663
LR792548.1	27700846	CATTTTCGGAAGCACACAATAA	27700826
XR 004425352.1	869	CATTTTCGGATGCACACA	853
XM 033131157.1	886	CATTTTCGGATGCACACA	870
XM 033131156.1	955	CATTTTCGGATGCACACA	939
XM 033131154.1	869	CATTTTCGGATGCACACA	853
LR778290.1	30972461	CATTTTCGGATGCACACA	30972445
LR778290.1	31640047	CATTTTCGGATGCACACA	31640031
LR778290.1	31724491	CATTTTCGGATGCACACA	31724507
LR778290.1	25939133	CATTTTCGGATGCACA	25939147
LR778289.1	37891225	CATTTGGGATGCACACAATA	37891244
LR778289.1	38138938	CATTTTCGGATGCACACA	38138922
LR778289.1	1178049	GGATGCACACAATAA	1178063
LR584434.1	4737613	CATTTTCGGATGCACACA	4737597

<u>LR584426.1</u>	40716693	CATTTCCGGATGCACACTATAA	40716673
<u>LR584426.1</u>	44568833	CATTTCCGGATGCACACA	44568849
<u>LR584421.1</u>	6926956	CATTTGGGATGCACACAATAA	6926936
<u>LR584414.1</u>	8526326	CATTTCCGGATGCACACA	8526310
<u>XM 026300864.1</u>	583	ACATTTCCGGATGCACAC	567
<u>XM 002071777.3</u>	408	TTTCCGGATGCACACAAT	392
<u>CP023332.1</u>	8950321	TTTCCGGATGCACACAAT	8950305
<u>XM 019095370.1</u>	861	ACATTTCCGGATGCTCACAATA	841
<u>XM 019095369.1</u>	861	ACATTTCCGGATGCTCACAATA	841
<u>XM 017162352.1</u>	914	ACATTTCCGGATGCACAC	898
<u>LN029746.1</u>	3786	ACATTTCCGGATGCACAC	3770
<u>AE013599.5</u>	13062817	TTTCCGGATGCACACAAT	13062801
<u>XM 007136828.1</u>	1139	TTTCCGGATGCACACAAT	1123
<u>OA888277.1</u>	4987	ATTTCCGGATGCACACAA	4971
<u>XM 037692747.1</u>	308	ACATTTCCGGATGCACAC	292
<u>LR812054.1</u>	43062162	ACATTTCCGGATGCACAC	43062146
<u>LR812560.1</u>	42722772	ACATTTCCGGATGCACAC	42722756
<u>CP002563.1</u>	1553457	TTCCGGATGCACACAATA	1553441
<u>AC146941.8</u>	83883	CATTTCCGGATGCACACA	83867
<u>LR812081.1</u>	42184890	ACATTTCCGGATGCACAC	42184906
<u>LR812071.1</u>	20700303	TTTCCGGATGCACACAAT	20700319
<u>LR812071.1</u>	7146821	TTCCGGATGCACACAA	7146835
<u>LR778269.1</u>	2268914	CATTTCCGGATGCAGACAAT	2268896
<u>LR778269.1</u>	12469841	CATTTGGGATGCACACAAT	12469823
<u>LR778269.1</u>	40386131	CATTTCCGGATGCACACA	40386147
<u>LR778269.1</u>	21275431	GGATGCACACAATAA	21275417
<u>LR778267.1</u>	2833714	CATTTGGGATGCACACAATAA	2833734
<u>LR778267.1</u>	42529115	GGATGCACACAATAA	42529101
<u>LR778267.1</u>	12460471	GGATGCACACAATAA	12460485
<u>LR778260.1</u>	16977017	CATTTGGGATGCACACAAT	16976999
<u>LR778260.1</u>	43845202	CATTTGGGATGCACACAAT	43845220
<u>LR778260.1</u>	18485245	CATTTCCGGATGCACACA	18485261
<u>LR778260.1</u>	13433975	ATTTCCGGATGCACAC	13433961
<u>XM 029729683.1</u>	3338	CATTTCCGGATGCACACA	3354
<u>XM 029729682.1</u>	3442	CATTTCCGGATGCACACA	3458
<u>XM 029729681.1</u>	3445	CATTTCCGGATGCACACA	3461
<u>LR584435.1</u>	4663289	CATTTAGGATGCACACAATAA	4663309
<u>LR584435.1</u>	56328727	CATTTGGGATGCACACAAT	56328709
<u>LR584435.1</u>	50726731	CATTTCCGGATGCACA	50726717
<u>LR584433.1</u>	2698528	CATTTAGGATGCACACAATAA	2698548
<u>LR584423.1</u>	23636461	CATTTCCGGATGCACACA	23636477
<u>LR535843.1</u>	9318085	ACATTTCCGGATGCACAC	9318101
<u>LR131966.1</u>	1611448	TCGGATGCACACAATAA	1611464
<u>CP025673.1</u>	18984285	CATTTCCGGATGCACAGAATAA	18984305
<u>LT594791.1</u>	27812647	TTTCCGGATGCACACAAT	27812663
<u>LT594790.1</u>	5296331	TTCCGGATGCACACAATA	5296347
<u>CP014500.1</u>	975022	ACATTTCCGGATGCACAC	975038
<u>AP015036.1</u>	43118478	ACATTTCGAATGCACACAATA	43118498
<u>XM 014194909.1</u>	3809	CATTTCCGGATGCACACA	3825
<u>XM 014194899.1</u>	3914	CATTTCCGGATGCACACA	3930
<u>XM 014194890.1</u>	3915	CATTTCCGGATGCACACA	3931
<u>CP012024.1</u>	588716	TTCCGGATGCACACAATA	588732
<u>LK066269.1</u>	301368	TCGGATGCACACAATAA	301384
<u>LK066269.1</u>	313490	TCGGATGCACACAATAA	313506
<u>LK066269.1</u>	325800	TCGGATGCACACAATAA	325816
<u>LN013793.1</u>	6993	CATTTCCGGATGCACACA	7009
<u>LN595178.1</u>	201332	ACATTTCCGGATGCTCACAATA	201352
<u>LR899427.1</u>	14480733	ACATTTCCGGATGCACAC	14480749

LR862152.1	6253400	TTTCGGATGCACACAAT	6253416
LR812056.1	42149186	ACATTTTCGGATGCACAC	42149202
LR812046.1	20731743	TTTCGGATGCACACAAT	20731759
HE601631.2	58795411	CATTTTCGTATGCACACAAT	58795393
HE601631.2	55394958	TTTCGGATGCACACAAT	55394974
HE601631.2	17951684	CATTTTCGGATGCACAC	17951669
LR812612.1	40748804	ACATTTTCGGATGCACAC	40748820
LR812602.1	20725446	TTTCGGATGCACACAAT	20725462
LR812587.1	41935364	ACATTTTCGGATGCACAC	41935380
LR812585.1	43351023	ACATTTTCGGATGCACAC	43351039
LR812577.1	21923491	TTTCGGATGCACACAAT	21923507
LR812577.1	7565339	TTCGGATGCACACAA	7565353
LR812552.1	20433661	TTTCGGATGCACACAAT	20433677
LR812552.1	5522106	TTCGGATGCACACAA	5522120
AC136507.19	74014	CATTTTCGGATGCACACA	74030
CR391919.5	156569	TTTCGGATGCACACAAT	156585
BX510640.4	83940	ACATTTTCGGATGCACAC	83956
AC006247.12	139248	TTTCGGATGCACACAAT	139264
CP046680.1	41617088	GGATGCACACAATAA	41617102
CP046680.1	41620099	GGATGCACACAATAA	41620113

PRODOTTI DEGLI ALLINEAMENTI SIGNIFICATIVI (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 SEQUENZA N° 8/1 (T=1; C=0; G=3; A=2) (Hitlist size 250)

Query: **ACATTTTCGGATGCACACAATAA**

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
.....

Accession	Assembly Name	Reference	Length (bp)	Contigs	GC (%)	Size (bp)	Identity (%)	Size (bp)
LR862364.1	Pipistrellus pipistrellus genome assembly, chromosome: 8	common pipistrelle	32.2	32.2	90%	138	95.00%	83522093
.....								
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

CONSRM 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	ATTTTCGGATGCACACAATAA	22	
Sbjct	52733536	ATTTTCAGATGCACACAATAA	52733517	

PRODOTTI DEGLI ALLINEAMENTI SIGNIFICATIVI (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="MSVAGELDFLKRVDKALHDPQMRSAPFTQDRLRMGREQATHEL
GNMEEWREAASE IRMHTI ENLDSYLEQLATNVRKNGGHVHFAPKAEDAVSYVSEIVRK
KGARSVIKSKSMVSEEIHLNKNLEEMGVVVEYDLGEYIIQLAGETPSHLIAPAIHKT
KEQVAELFSKEAGRKIPVDTNQLLQFAREKLRKEFLEADIGISGCNFAVAESGSVLI
SNEGNAARLTTLTPKVHIAIMGEMERLVPTWKDLVDVVISMLTRSATGQKITSYVTGINGA
RLHEDMDGPEEFHLIILDNGRSDILGTAYEEVLKCIKRCGACVNVCPVYRNIGGHAYGS
VYSGPIGAVLAPLLEGEYELKELPFASSLCGACTDVCPVKIPLHDLLEHRKDVVEQG
FITKFEQLAFKGFTFMVSHPFVYEKAIKMAHTGLSIFAKDGYIANGPGLKGWTESRD
FPKPAKQSFDRDWWKKETRDS"

```

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
GNMEEWREAASEIRMHTIENLDSYLEQLATNVRKNGGHVHFAFKAEDAVSYVSEIVRK
KGARSVIKSKSMVSEEIHLNKNLEEMGVKVEYDLGEYIIQLAGETPSHLIAPAIHKT
KEQVAELFSKEAGRKIPVDTNQLLQFAREKLRKEFLEADIGISGCNFAVAESGSVVL
SNEG NARLTTTLPKVHIAIMGMRVLVPTWKDLDVVISMLTRSATGQKITSYVTGINGA
RLHEDMDGPEEFHLIILDNGRSDILGTAYEEVLKCIKRCGACVNVCPVYRNIGGHAYGS
VYSGPIGAVLAPLLEGEELKELPFASSLCGACTDVCVPIPLHDLLIEHRKDVVEQG
FITKFEQLAFKGFTFMVSHPFVYEKAIKMAHTGLSIFAKDGYIANGPGPLKGWTESRD
FPKPAKQSF RDWKKETRDS"
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
PAYNSGYQEDAKLAAKQMIKAFEHSEYIVTPSGSCASMVHHYKEMFKGDSEWYEKAV
HLADRTYELTDFVFNILGKNDWKS LVEKAVFHQSCHMSRALGIKKEEPLKLLSQVEGL
DIKELPYCQDCCGFGGTFAVKMSSISETMVDEKIKHIEATEANLLIGADMGLMNIGG
RLRRENKNIQVLHVAEVLAKGLNK"

```

ORIGIN

```
1 ttatgaatcc ctcgtttcct 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 gagaaggcca tgcaatggaa ttttcacagg
301 acaaacatct gtacatgtct cgcacaaaact 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 aatgactac
661 atcaagatct ttccatgtcg gaacaaggcg ttccatcccc ataatcgcaa tatgaacctt
721 tggcaatgtc gtagtaaggc gagcatttcc ttcattactt ataagcacia ctgatccaga
781 ttccgctacc gcaaagttag aaccggaaat accaatatca gcctctaaaa actctttacg
841 caatttctcc cttgcaaatt gtagtaattg attcgtatca actgggatct ttcttctgcg
901 ttccttagaa aaaagctctg ctacttggtc cttcgtctta tgaatggccg gggcgattaa
961 atgggatggc gtttctccag ctaattgaat aatatactcc ccaagatcat actctactac
1021 ttttaccccc atctcttcta agtttttatt aagatgtatt tcttctgaaa ccatggattt
1081 tgacttaatt actgatctcg ctctttttt tcttacgatt tcagaaacat atgatacagc
1141 atcttccgct ttgaaagcga aatgcacatg cccccattt ttccgaacat tagtagctaa
1201 ttgttccaaa taactatcta aattttcgat tgtgtgcatc cgaatttcag atgctgcctc
1261 gcgccattct tccatgttcc ctaattcatg tgtagcctgt tcccttcca 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 krvdkaahdp qmrsavpftq drlrmgreqa thelgnmeew reease irmh
61 tienldsyle qlatnvrkng ghvhfakae davsylvseiv 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 circgacvnn cpvyrniggh aygsvysgpi gavlaplleg yeelkelpfa
361 sslcgactdv cpvkiplhdl liehrkdvve qgfitkfeql afkgftfmvs hpfvyekaik
421 mahtglsifa kdgyiangpg plkgwtesrd fpkpakqsfr dwkketrds
```

//

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](#)

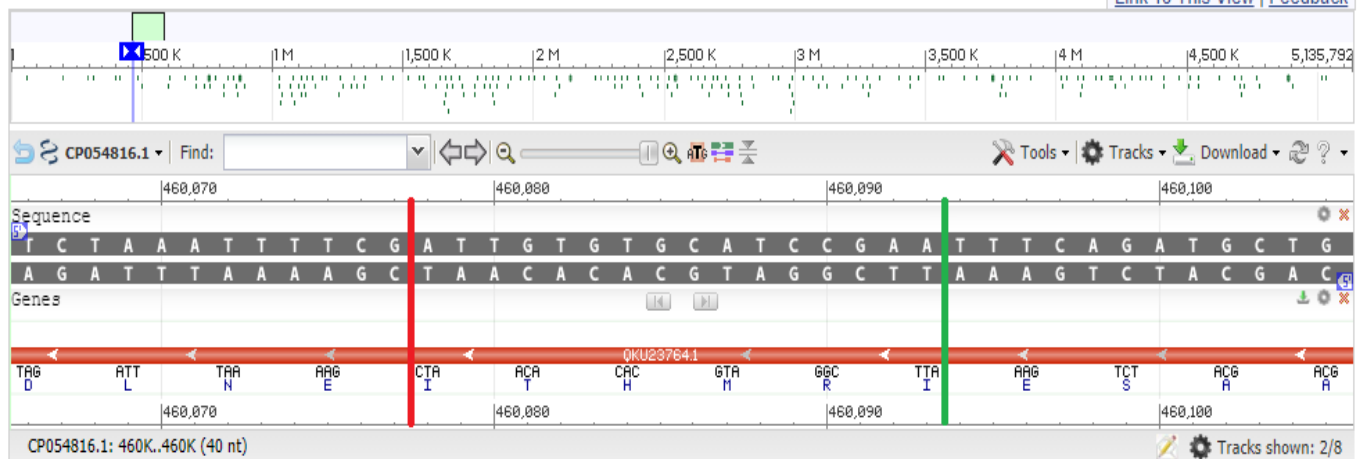
CDS: Putative 1	1		R M H T	
Query	5		TTCGGATGCACACAAT	20
Sbjct	460093		TTCGGATGCACACAAT	460078
CDS: iron-sulfur clus	57		I R M H T I	

Bacillus anthracis strain FDAARGOS_695 chromosome

GenBank: CP054816.1

[GenBank](#) [FASTA](#)

[Link To This View](#) | [Feedback](#)



PROTEINE LEGANTI I CLUSTER FERRO-ZOLFO

JB Broderick, in *Comprehensive Coordination Chemistry II*, 2003 (<https://doi.org/10.1016/B0-08-043748-6/08214-1>) così scrive: “I **Cluster Ferro-Zolfo** sono tra le strutture contenenti metalli più diffuse in biologia e sono noti per il loro ruolo nel trasferimento di elettroni, tra cui in particolare nelle *ferredossine*, nella *catena di trasporto degli elettroni mitocondriali* e nella *fotosintesi*....i cluster ferro-zolfo funzionano (anche) in ruoli regolatori, attivano o disattivano l’espressione genica in risposta ai livelli di ferro, ossigeno o superossido....Le prove indicano anche un ruolo *strutturale* essenziale per i cluster ferro-zolfo in diversi enzimi, inclusi gli enzimi di riparazione del DNA endonucleasi III e MutY. I cluster ferro-zolfo possono anche essere usati direttamente nella catalisi della chimica redox su piccole molecole” (<https://www.sciencedirect.com/topics/biochemistry-genetics-and-molecular-biology/iron-sulfur-cluster>). La proteina identificata dalla sequenza **CP054816.1** del **Bacillus Anthracis** è la **proteina di utilizzo del L-lattato LutB, che contiene un dominio di tipo ferredossina** [Produzione e conversione di energia] <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	

CAPITOLO XII

**RICERCHE BLAST DI ALTRE
SEQUENZE MODIFICATE DELLA
SEQUENZA DI RNA
“GTGAAATGGTCATGTGTGGCGG”
del Gene RdRP (RdRp_SARSr-F)**

OBIETTIVI DEL CAPITOLO XII

In questo capitolo sono riportati i risultati delle Ricerche BLAST di altre Sequenze Modificate della Sequenza di RNA “GTGAAATGGTCATGTGTGGCGG” del Gene RdRp (RdRp_SARSr-F). Le sequenze modificate sono le sequenze 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 e la 10/1.

ATTENZIONE: Per le ragioni già esposte a pagina 437 (negli “Obiettivi del capitolo XI”) sono costretto a lasciare allo scienziato interessato l’arduo compito (che consiglio assolutamente di eseguire) di approfondire l’analisi e la selezione delle sequenze ritenute rilevanti secondo le indicazioni proposte nei capitoli precedenti.

SEQUENZA N° 1/2

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

SEQUENZA N° 1/2 (T=0; C=1; G=2; A=3)

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° 1/2	1 0 1 2 2 2 1 3 3 0 2 3 1 3 1 3 1 3 3 0 1 1
Sequenza 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	Chanos chanos genome assembly, chromosome: 1	milkfish	38.2	100	86%	2.2	100.00%	63121898
LR633955.1 2/2 3/1 3/3 3/4 8/1	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 willdenowii 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	Bacillus subtilis	34.2	34.2	77%	35	100.00%	4197841

	chromosome, complete genome									
CP051306.1	Bacillus subtilis strain RI4914 chromosome, complete genome	Bacillus subtilis	34.2	34.2	77%	35	100.00%	4100930		
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 PREDICTED: Piliocolobus tephrosceles adaptor related protein complex 3 subunit mu 2 (AP3M2), transcript variant X1, mRNA	Bacillus subtilis	34.2	34.2	77%	35	100.00%	4119195		
XM_023214463.2 9/1	Ugandan red Colobus	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 PREDICTED: Sphaerama orbicularis Alport syndrome, mental retardation, midface hypoplasia and elliptocytosis chromosomal region gene 1 (ammecr1), mRNA	Bacillus subtilis	34.2	34.2	77%	35	100.00%	4126713		
XM_030153818.1 2/2 3/1 3/5 3/6	orbiculate cardinalfish	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 3/2 3/3 3/5 9/1 LR597471.1	Aquila chrysaetos chrysaetos genome assembly, chromosome: 16	Aquila chrysaetos chrysaetos	34.2	34.2	95%	35	95.24%	30610376
LR597461.1 2/2 3/1 3/5 3/6	Sphaeramia orbicularis genome assembly, chromosome: 4	Sphaeramia orbicularis orbiculate cardinalfish	34.2	34.2	77%	35	100.00%	55997211
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 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	Bacillus subtilis	34.2	34.2	77%	35	100.00%	4142132

chromosome

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 PREDICTED: Drosophila virilis octopamine receptor Oamb (LOC6629914), transcript variant X2, mRNA	Coregonus sp. 'balchen'	34.2	64.4	81%	35	100.00%	92224161
XM_032439116.1	PREDICTED: Drosophila virilis octopamine receptor Oamb (LOC6629914), transcript variant X1, 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 LR697108.1 2/2 3/1 3/3 3/4 8/1	PREDICTED: Quercus lobata F-box protein At3g07870-like (LOC115983915), mRNA	Quercus lobata	34.2	34.2	77%	35	100.00%	1990
LR606189.1 3/2 3/3 3/5 9/1	Chanos chanos genome assembly, chromosome: 3	milkfish	34.2	64.4	86%	35	100.00%	58651890
CP021123.1	Aquila chrysaetos chrysaetos genome assembly, chromosome: 9	Aquila chrysaetos chrysaetos	34.2	34.2	77%	35	100.00%	45241448
CP025987.1 LR584417.1 2/2 3/3 3/4 3/6 8/1	Bacillus subtilis strain SEM-9 chromosome, complete genome	Bacillus subtilis	34.2	34.2	77%	35	100.00%	4121982
LR535845.1	Marinomonas sp. FW-1 chromosome, complete genome	Marinomonas sp. FW-1	34.2	34.2	77%	35	100.00%	3964876
LR877217.1	Salmotrutta genome assembly, chromosome: 37	river trout	34.2	64.4	77%	35	100.00%	35205783
LR862365.1 2/2 3/2 3/3 3/5 6/1 7/2 8/1 9/1 10/1	Mastacembelus armatus genome assembly, chromosome: 13	zig-zag eel	34.2	64.4	77%	35	100.00%	27498257
LR828303.1	Acomys russatus genome assembly, chromosome: 6	golden spiny mouse	34.2	64.4	77%	35	100.00%	89442464
CP028812.1	Pipistrellus pipistrellus genome assembly, chromosome: 9	common pipistrelle	34.2	68.4	77%	35	100.00%	72542072
	Ananas comosus genome assembly, chromosome: 23	pineapple	34.2	34.2	77%	35	100.00%	7979763
	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

CONSRM 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	

VISTA DI ALLINEAMENTO VERTICALE AGGANCIATO A QUERY CON LETTERE D'IDENTITA'

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	TCTGGGCAATGACACACAA	30835063
LR131992.1	2784243	GGCAATGACACACAA	2784257
CP027092.1	45563484	CGGGCAATGACACACAAT	45563501
CP027092.1	57838314	AATGACACACAATCC	57838328
AL445218.12	9837	GGCAATGACACACAATCC	9854
LR812132.1	2930334	TCTGGGCAATGACACACA	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

<u>CP051306.1</u>	1149953	GCAATGACACACAATCC	1149937
<u>CP050532.1</u>	1163761	GCAATGACACACAATCC	1163745
<u>CP026662.1</u>	3670433	GCAATGACACACAATCC	3670417
<u>CP046860.1</u>	220600	GCAATGACACACAATCC	220584
<u>CP045672.1</u>	1149573	GCAATGACACACAATCC	1149557
<u>CP031693.1</u>	1784581	GCAATGACACACAATCC	1784565
<u>CP028215.1</u>	3409502	GCAATGACACACAATCC	3409486
<u>CP028201.1</u>	958052	GCAATGACACACAATCC	958036
<u>CP047564.1</u>	55777957	GGGCAATGACACACAAT	55777941
<u>CP047564.1</u>	127201	CAATGACACACAATC	127187
<u>CP047485.1</u>	1188146	GCAATGACACACAATCC	1188130
<u>CP023409.1</u>	1217994	GCAATGACACACAATCC	1217978
<u>CP047325.1</u>	2666125	GCAATGACACACAATCC	2666109
<u>XM 023214463.2</u>	2390	GCAATGACACACAATCC	2374
<u>CP045811.1</u>	1149504	GCAATGACACACAATCC	1149488
<u>CP045812.1</u>	1170071	GCAATGACACACAATCC	1170055
<u>CP045922.1</u>	1170071	GCAATGACACACAATCC	1170055
<u>CP045816.1</u>	1179248	GCAATGACACACAATCC	1179232
<u>CP045817.1</u>	1136046	GCAATGACACACAATCC	1136030
<u>CP045818.1</u>	1138918	GCAATGACACACAATCC	1138902
<u>CP045819.1</u>	1157077	GCAATGACACACAATCC	1157061
<u>CP045820.1</u>	1138568	GCAATGACACACAATCC	1138552
<u>CP045821.1</u>	1149568	GCAATGACACACAATCC	1149552
<u>CP045823.1</u>	1138556	GCAATGACACACAATCC	1138540
<u>CP045824.1</u>	1148638	GCAATGACACACAATCC	1148622
<u>CP045825.1</u>	1183154	GCAATGACACACAATCC	1183138
<u>CP045826.1</u>	1214198	GCAATGACACACAATCC	1214182
<u>CP031784.1</u>	218244	GCAATGACACACAATCC	218228
<u>CP031783.1</u>	1658707	GCAATGACACACAATCC	1658691
<u>CP045425.1</u>	1137129	GCAATGACACACAATCC	1137113
<u>CP044498.1</u>	1123295	GCAATGACACACAATCC	1123279
<u>XM 030153818.1</u>	4465	GGGCAATGACACACAAT	4449
<u>CP041757.1</u>	1170071	GCAATGACACACAATCC	1170055
<u>CP041360.1</u>	3508421	GGCAATGACACACAATC	3508405
<u>LR606196.1</u>	9886770	TCGGGCAATGACTCACAATCC	9886750
<u>LR597471.1</u>	22884527	GGGCAATGACACACAAT	22884511
<u>LR597461.1</u>	29961218	GCAATGACACACAATCC	29961202
<u>AP019714.1</u>	1170075	GCAATGACACACAATCC	1170059
<u>CP034484.1</u>	1170078	GCAATGACACACAATCC	1170062
<u>CP040528.1</u>	1145059	GCAATGACACACAATCC	1145043
<u>CP039935.1</u>	1153914	GCAATGACACACAATCC	1153898
<u>CP033205.1</u>	1149966	GCAATGACACACAATCC	1149950
<u>LR584416.1</u>	58542907	GCAATGACACACAATCC	58542891
<u>LR584416.1</u>	50619263	CAATGACACACAATCC	50619248
<u>LR584416.1</u>	36250074	CAATGACACACAATC	36250060
<u>LR584416.1</u>	45280676	CAATGACACACAATC	45280662
<u>LR584416.1</u>	41598409	AATGACACACAATCC	41598423
<u>CP029609.1</u>	1187677	GCAATGACACACAATCC	1187661
<u>CP024767.1</u>	4158911	CTCGGGCAATGACACAC	4158895
<u>CP035414.1</u>	1160645	GCAATGACACACAATCC	1160629
<u>CP035413.1</u>	1186728	GCAATGACACACAATCC	1186712
<u>CP035411.1</u>	1144667	GCAATGACACACAATCC	1144651
<u>CP035403.1</u>	1233149	GCAATGACACACAATCC	1233133
<u>CP035406.1</u>	1181052	GCAATGACACACAATCC	1181036
<u>CP035397.1</u>	1163029	GCAATGACACACAATCC	1163013
<u>CP035400.1</u>	1160029	GCAATGACACACAATCC	1160013
<u>CP046592.1</u>	1187398	GCAATGACACACAATCC	1187382
<u>CP062497.1</u>	1163796	GCAATGACACACAATCC	1163780

<u>CP060710.1</u>	1170515	GCAATGACACACAATCC	1170499
<u>CP060417.1</u>	1135143	GCAATGACACACAATCC	1135127
<u>CP049924.1</u>	1128410	GCAATGACACACAATCC	1128394
<u>CP058242.1</u>	1170479	GCAATGACACACAATCC	1170463
<u>CP051860.1</u>	4093524	GCAATGACACACAATCC	4093540
<u>CP050582.1</u>	4860768	CTCGGGCAATGACACAC	4860784
<u>CP050640.1</u>	4858108	CTCGGGCAATGACACAC	4858124
<u>CP050319.1</u>	3061266	GCAATGACACACAATCC	3061282
<u>LR778280.1</u>	30594062	GCAATGACACACAATCC	30594078
<u>LR778256.1</u>	2397833	GGCAATGACACACAATC	2397849
<u>LR778256.1</u>	55782119	GGGCAATGACACACA	55782133
<u>XM 032439116.1</u>	354	CTCGGGCAATGACACAC	370
<u>XM 032439114.1</u>	333	CTCGGGCAATGACACAC	349
<u>CP028218.1</u>	903496	GCAATGACACACAATCC	903512
<u>CP028217.1</u>	3858288	GCAATGACACACAATCC	3858304
<u>CP028213.1</u>	2531942	GCAATGACACACAATCC	2531958
<u>CP028212.1</u>	1598342	GCAATGACACACAATCC	1598358
<u>CP028209.1</u>	3607020	GCAATGACACACAATCC	3607036
<u>CP028202.1</u>	1844442	GCAATGACACACAATCC	1844458
<u>CP046448.1</u>	3051899	GCAATGACACACAATCC	3051915
<u>CP046047.1</u>	2932105	GCAATGACACACAATCC	2932121
<u>CP045006.1</u>	2934342	GCAATGACACACAATCC	2934358
<u>XM 031106775.1</u>	1434	GGGCAATGACACACAAT	1450
<u>LR697108.1</u>	33470624	GCAATGACACACAATCC	33470640
<u>LR697108.1</u>	52963245	GGGCAATGACACACA	52963231
<u>LR606189.1</u>	2584263	GCAATGACACACAATCC	2584279
<u>CP021123.1</u>	2980154	GCAATGACACACAATCC	2980170
<u>CP025987.1</u>	2995424	GGCAATGACACACAATC	2995440
<u>LR584417.1</u>	17130184	GCAATGACACACAATCC	17130200
<u>LR584417.1</u>	25903436	AATGACACACAATCC	25903422
<u>LR535845.1</u>	6973460	GGCAATGACACACAATC	6973476
<u>LR535845.1</u>	8298486	GCAATGACACACAAT	8298472
<u>LR877217.1</u>	81321755	GGGCAATGACACACAAT	81321771
<u>LR877217.1</u>	64453847	GCAATGACACACAAT	64453833
<u>LR862365.1</u>	22466815	GCAATGACACACAATCC	22466831
<u>LR862365.1</u>	22505624	GCAATGACACACAATCC	22505640
<u>LR828303.1</u>	653605	CGGGCAATGACACACAA	653621
<u>CP028812.1</u>	2442752	GCAATGACACACAATCC	2442768

PRODOTTI DEGLI ALLINEAMENTI SIGNIFICATIVI (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=149598](https://www.ncbi.nlm.nih.gov/nucleotide/CP054177.1?report=genbank&log$=nuclalign&blast_rank=10&RID=W5C1YPBB013&from=1149582&to=149598)

[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/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/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=148553](https://www.ncbi.nlm.nih.gov/nucleotide/CP051466.1?report=genbank&log$=nuclalign&blast_rank=13&RID=W5C1YPBB013&from=1148537&to=148553)

[https://www.ncbi.nlm.nih.gov/nucleotide/CP051465.1?report=genbank&log\\$=nuclalign&blast_rank=14&RID=W5C1YPBB013&from=1148288&to=148304](https://www.ncbi.nlm.nih.gov/nucleotide/CP051465.1?report=genbank&log$=nuclalign&blast_rank=14&RID=W5C1YPBB013&from=1148288&to=148304)

[https://www.ncbi.nlm.nih.gov/nucleotide/CP051462.1?report=genbank&log\\$=nuclalign&blast_rank=15&RID=W5C1YPBB013&from=1286742&to=286758](https://www.ncbi.nlm.nih.gov/nucleotide/CP051462.1?report=genbank&log$=nuclalign&blast_rank=15&RID=W5C1YPBB013&from=1286742&to=286758)

[https://www.ncbi.nlm.nih.gov/nucleotide/CP051306.1?report=genbank&log\\$=nuclalign&blast_rank=16&RID=W5C1YPBB013&from=1149937&to=149953](https://www.ncbi.nlm.nih.gov/nucleotide/CP051306.1?report=genbank&log$=nuclalign&blast_rank=16&RID=W5C1YPBB013&from=1149937&to=149953)

[https://www.ncbi.nlm.nih.gov/nucleotide/CP050532.1?report=genbank&log\\$=nuclalign&blast_rank=17&RID=W5C1YPBB013&from=1163745&to=163761](https://www.ncbi.nlm.nih.gov/nucleotide/CP050532.1?report=genbank&log$=nuclalign&blast_rank=17&RID=W5C1YPBB013&from=1163745&to=163761)

[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/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=784581](https://www.ncbi.nlm.nih.gov/nucleotide/CP031693.1?report=genbank&log$=nuclalign&blast_rank=21&RID=W5C1YPBB013&from=1784565&to=784581)

[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/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/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=188146](https://www.ncbi.nlm.nih.gov/nucleotide/CP047485.1?report=genbank&log$=nuclalign&blast_rank=25&RID=W5C1YPBB013&from=1188130&to=188146)

[https://www.ncbi.nlm.nih.gov/nucleotide/CP023409.1?report=genbank&log\\$=nuclalign&blast_rank=26&RID=W5C1YPBB013&from=1217978&to=217994](https://www.ncbi.nlm.nih.gov/nucleotide/CP023409.1?report=genbank&log$=nuclalign&blast_rank=26&RID=W5C1YPBB013&from=1217978&to=217994)

[https://www.ncbi.nlm.nih.gov/nucleotide/CP047325.1?report=genbank&log\\$=nuclalign&blast_rank=27&RID=W5C1YPBB013&from=2666109&to=666125](https://www.ncbi.nlm.nih.gov/nucleotide/CP047325.1?report=genbank&log$=nuclalign&blast_rank=27&RID=W5C1YPBB013&from=2666109&to=666125)

[https://www.ncbi.nlm.nih.gov/nucleotide/CP045811.1?report=genbank&log\\$=nuclalign&blast_rank=29&RID=W5C1YPBB013&from=1149488&to=149504](https://www.ncbi.nlm.nih.gov/nucleotide/CP045811.1?report=genbank&log$=nuclalign&blast_rank=29&RID=W5C1YPBB013&from=1149488&to=149504)

[https://www.ncbi.nlm.nih.gov/nucleotide/CP045812.1?report=genbank&log\\$=nuclalign&blast_rank=30&RID=W5C1YPBB013&from=1170055&to=170071](https://www.ncbi.nlm.nih.gov/nucleotide/CP045812.1?report=genbank&log$=nuclalign&blast_rank=30&RID=W5C1YPBB013&from=1170055&to=170071)

[https://www.ncbi.nlm.nih.gov/nucleotide/CP045922.1?report=genbank&log\\$=nuclalign&blast_rank=31&RID=W5C1YPBB013&from=1170055&to=170071](https://www.ncbi.nlm.nih.gov/nucleotide/CP045922.1?report=genbank&log$=nuclalign&blast_rank=31&RID=W5C1YPBB013&from=1170055&to=170071)

[https://www.ncbi.nlm.nih.gov/nucleotide/CP045816.1?report=genbank&log\\$=nuclalign&blast_rank=32&RID=W5C1YPBB013&from=1179232&to=179248](https://www.ncbi.nlm.nih.gov/nucleotide/CP045816.1?report=genbank&log$=nuclalign&blast_rank=32&RID=W5C1YPBB013&from=1179232&to=179248)

[https://www.ncbi.nlm.nih.gov/nucleotide/CP045817.1?report=genbank&log\\$=nuclalign&blast_rank=33&RID=W5C1YPBB013&from=1136030&to=136046](https://www.ncbi.nlm.nih.gov/nucleotide/CP045817.1?report=genbank&log$=nuclalign&blast_rank=33&RID=W5C1YPBB013&from=1136030&to=136046)

[https://www.ncbi.nlm.nih.gov/nucleotide/CP045818.1?report=genbank&log\\$=nuclalign&blast_rank=34&RID=W5C1YPBB013&from=1138902&to=138918](https://www.ncbi.nlm.nih.gov/nucleotide/CP045818.1?report=genbank&log$=nuclalign&blast_rank=34&RID=W5C1YPBB013&from=1138902&to=138918)

[https://www.ncbi.nlm.nih.gov/nucleotide/CP045819.1?report=genbank&log\\$=nuclalign&blast_rank=35&RID=W5C1YPBB013&from=1157061&to=157077](https://www.ncbi.nlm.nih.gov/nucleotide/CP045819.1?report=genbank&log$=nuclalign&blast_rank=35&RID=W5C1YPBB013&from=1157061&to=157077)

[https://www.ncbi.nlm.nih.gov/nucleotide/CP045820.1?report=genbank&log\\$=nuclalign&blast_rank=36&RID=W5C1YPBB013&from=1138552&to=138568](https://www.ncbi.nlm.nih.gov/nucleotide/CP045820.1?report=genbank&log$=nuclalign&blast_rank=36&RID=W5C1YPBB013&from=1138552&to=138568)

[https://www.ncbi.nlm.nih.gov/nucleotide/CP045821.1?report=genbank&log\\$=nuclalign&blast_rank=37&RID=W5C1YPBB013&from=1149552&to=149568](https://www.ncbi.nlm.nih.gov/nucleotide/CP045821.1?report=genbank&log$=nuclalign&blast_rank=37&RID=W5C1YPBB013&from=1149552&to=149568)

[https://www.ncbi.nlm.nih.gov/nucleotide/CP045823.1?report=genbank&log\\$=nuclalign&blast_rank=38&RID=W5C1YPBB013&from=1138540&to=138556](https://www.ncbi.nlm.nih.gov/nucleotide/CP045823.1?report=genbank&log$=nuclalign&blast_rank=38&RID=W5C1YPBB013&from=1138540&to=138556)

[https://www.ncbi.nlm.nih.gov/nucleotide/CP045824.1?report=genbank&log\\$=nuclalign&blast_rank=39&RID=W5C1YPBB013&from=1148622&to=148638](https://www.ncbi.nlm.nih.gov/nucleotide/CP045824.1?report=genbank&log$=nuclalign&blast_rank=39&RID=W5C1YPBB013&from=1148622&to=148638)

[https://www.ncbi.nlm.nih.gov/nucleotide/CP045825.1?report=genbank&log\\$=nuclalign&blast_rank=40&RID=W5C1YPBB013&from=1183138&to=183154](https://www.ncbi.nlm.nih.gov/nucleotide/CP045825.1?report=genbank&log$=nuclalign&blast_rank=40&RID=W5C1YPBB013&from=1183138&to=183154)

[https://www.ncbi.nlm.nih.gov/nucleotide/CP045826.1?report=genbank&log\\$=nuclalign&blast_rank=41&RID=W5C1YPBB013&from=1214182&to=214198](https://www.ncbi.nlm.nih.gov/nucleotide/CP045826.1?report=genbank&log$=nuclalign&blast_rank=41&RID=W5C1YPBB013&from=1214182&to=214198)

[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/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=658707](https://www.ncbi.nlm.nih.gov/nucleotide/CP031783.1?report=genbank&log$=nuclalign&blast_rank=43&RID=W5C1YPBB013&from=1658691&to=658707)

[https://www.ncbi.nlm.nih.gov/nucleotide/CP041757.1?report=genbank&log\\$=nuclalign&blast_rank=47&RID=W5C1YPBB013&from=1170055&to=170071](https://www.ncbi.nlm.nih.gov/nucleotide/CP041757.1?report=genbank&log$=nuclalign&blast_rank=47&RID=W5C1YPBB013&from=1170055&to=170071)

[https://www.ncbi.nlm.nih.gov/nucleotide/AP019714.1?report=genbank&log\\$=nuclalign&blast_rank=52&RID=W5C1YPBB013&from=1170059&to=170075](https://www.ncbi.nlm.nih.gov/nucleotide/AP019714.1?report=genbank&log$=nuclalign&blast_rank=52&RID=W5C1YPBB013&from=1170059&to=170075)

[https://www.ncbi.nlm.nih.gov/nucleotide/CP034484.1?report=genbank&log\\$=nuclalign&blast_rank=53&RID=W5C1YPBB013&from=1170062&to=170078](https://www.ncbi.nlm.nih.gov/nucleotide/CP034484.1?report=genbank&log$=nuclalign&blast_rank=53&RID=W5C1YPBB013&from=1170062&to=170078)

[https://www.ncbi.nlm.nih.gov/nucleotide/CP040528.1?report=genbank&log\\$=nuclalign&blast_rank=54&RID=W5C1YPBB013&from=1145043&to=145059](https://www.ncbi.nlm.nih.gov/nucleotide/CP040528.1?report=genbank&log$=nuclalign&blast_rank=54&RID=W5C1YPBB013&from=1145043&to=145059)

[https://www.ncbi.nlm.nih.gov/nucleotide/CP033205.1?report=genbank&log\\$=nuclalign&blast_rank=56&RID=W5C1YPBB013&from=1149950&to=149966](https://www.ncbi.nlm.nih.gov/nucleotide/CP033205.1?report=genbank&log$=nuclalign&blast_rank=56&RID=W5C1YPBB013&from=1149950&to=149966)
[https://www.ncbi.nlm.nih.gov/nucleotide/CP029609.1?report=genbank&log\\$=nuclalign&blast_rank=58&RID=W5C1YPBB013&from=1187661&to=187677](https://www.ncbi.nlm.nih.gov/nucleotide/CP029609.1?report=genbank&log$=nuclalign&blast_rank=58&RID=W5C1YPBB013&from=1187661&to=187677)
[https://www.ncbi.nlm.nih.gov/nucleotide/CP035414.1?report=genbank&log\\$=nuclalign&blast_rank=60&RID=W5C1YPBB013&from=1160629&to=160645](https://www.ncbi.nlm.nih.gov/nucleotide/CP035414.1?report=genbank&log$=nuclalign&blast_rank=60&RID=W5C1YPBB013&from=1160629&to=160645)
[https://www.ncbi.nlm.nih.gov/nucleotide/CP035413.1?report=genbank&log\\$=nuclalign&blast_rank=61&RID=W5C1YPBB013&from=1186712&to=186728](https://www.ncbi.nlm.nih.gov/nucleotide/CP035413.1?report=genbank&log$=nuclalign&blast_rank=61&RID=W5C1YPBB013&from=1186712&to=186728)
[https://www.ncbi.nlm.nih.gov/nucleotide/CP035411.1?report=genbank&log\\$=nuclalign&blast_rank=62&RID=W5C1YPBB013&from=1144651&to=144667](https://www.ncbi.nlm.nih.gov/nucleotide/CP035411.1?report=genbank&log$=nuclalign&blast_rank=62&RID=W5C1YPBB013&from=1144651&to=144667)
[https://www.ncbi.nlm.nih.gov/nucleotide/CP035403.1?report=genbank&log\\$=nuclalign&blast_rank=63&RID=W5C1YPBB013&from=1233133&to=1233149](https://www.ncbi.nlm.nih.gov/nucleotide/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=181052](https://www.ncbi.nlm.nih.gov/nucleotide/CP035406.1?report=genbank&log$=nuclalign&blast_rank=64&RID=W5C1YPBB013&from=1181036&to=181052)
[https://www.ncbi.nlm.nih.gov/nucleotide/CP035397.1?report=genbank&log\\$=nuclalign&blast_rank=65&RID=W5C1YPBB013&from=1163013&to=163029](https://www.ncbi.nlm.nih.gov/nucleotide/CP035397.1?report=genbank&log$=nuclalign&blast_rank=65&RID=W5C1YPBB013&from=1163013&to=163029)
[https://www.ncbi.nlm.nih.gov/nucleotide/CP035400.1?report=genbank&log\\$=nuclalign&blast_rank=66&RID=W5C1YPBB013&from=1160013&to=160029](https://www.ncbi.nlm.nih.gov/nucleotide/CP035400.1?report=genbank&log$=nuclalign&blast_rank=66&RID=W5C1YPBB013&from=1160013&to=160029)
[https://www.ncbi.nlm.nih.gov/nucleotide/CP046592.1?report=genbank&log\\$=nuclalign&blast_rank=67&RID=W5C1YPBB013&from=1187382&to=187398](https://www.ncbi.nlm.nih.gov/nucleotide/CP046592.1?report=genbank&log$=nuclalign&blast_rank=67&RID=W5C1YPBB013&from=1187382&to=187398)
[https://www.ncbi.nlm.nih.gov/nucleotide/CP062497.1?report=genbank&log\\$=nuclalign&blast_rank=68&RID=W5C1YPBB013&from=1163780&to=163796](https://www.ncbi.nlm.nih.gov/nucleotide/CP062497.1?report=genbank&log$=nuclalign&blast_rank=68&RID=W5C1YPBB013&from=1163780&to=163796)
[https://www.ncbi.nlm.nih.gov/nucleotide/CP060710.1?report=genbank&log\\$=nuclalign&blast_rank=69&RID=W5C1YPBB013&from=1170499&to=170515](https://www.ncbi.nlm.nih.gov/nucleotide/CP060710.1?report=genbank&log$=nuclalign&blast_rank=69&RID=W5C1YPBB013&from=1170499&to=170515)
[https://www.ncbi.nlm.nih.gov/nucleotide/CP060417.1?report=genbank&log\\$=nuclalign&blast_rank=70&RID=W5C1YPBB013&from=1135127&to=135143](https://www.ncbi.nlm.nih.gov/nucleotide/CP060417.1?report=genbank&log$=nuclalign&blast_rank=70&RID=W5C1YPBB013&from=1135127&to=135143)
[https://www.ncbi.nlm.nih.gov/nucleotide/CP049924.1?report=genbank&log\\$=nuclalign&blast_rank=71&RID=W5C1YPBB013&from=1128394&to=128410](https://www.ncbi.nlm.nih.gov/nucleotide/CP049924.1?report=genbank&log$=nuclalign&blast_rank=71&RID=W5C1YPBB013&from=1128394&to=128410)
[https://www.ncbi.nlm.nih.gov/nucleotide/CP058242.1?report=genbank&log\\$=nuclalign&blast_rank=72&RID=W5C1YPBB013&from=1170463&to=170479](https://www.ncbi.nlm.nih.gov/nucleotide/CP058242.1?report=genbank&log$=nuclalign&blast_rank=72&RID=W5C1YPBB013&from=1170463&to=170479)
[https://www.ncbi.nlm.nih.gov/nucleotide/CP051860.1?report=genbank&log\\$=nuclalign&blast_rank=73&RID=W5C1YPBB013&from=4093524&to=4093540](https://www.ncbi.nlm.nih.gov/nucleotide/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/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/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/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/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/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/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/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/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/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/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/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](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>

SEQUENZA N° 2/2

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

SEQUENZA N° 2/2 (T=3; C=2; G=1; A=0)

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° 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
Sequenza 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 AP023477.1	Eukaryotic synthetic construct chromosome 17	eukaryotic synthetic construct	38.2	128	95%	2.2	100.00%	88299790
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	Homo sapiens chromosome 17 , clone hRPK.721_K_1, complete sequence	human	38.2	38.2	86%	2.2	100.00%	163908
1/2 3/2 3/3 5/1 6/1 10/1	Takifugu rubripes genome assembly, chromosome: 20	torafugu	36.2	36.2	81%	8.7	100.00%	18239318
LR584236.2 LR584433.1	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 LR862367.1	Cyprinus carpio genome assembly common carp genome, scaffold: LG39, chromosome: 39	common carp	36.2	66.4	81%	8.7	100.00%	16255708
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	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	Sphaerama 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 8/1	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	Canis lupus familiaris breed Labrador retriever chromosome	dog	34.2	94.6	81%	35	100.00%	73584800

1/2 3/3 3/4 3/5 3/6 5/1 10/1	11a								
CP050630.1	Canis lupus familiaris breed								
1/2 3/3 3/4 3/5 3/6 5/1 10/1	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	Coregonus sp. 'balchen' genome assembly, chromosome: 1	Coregonus sp. 'balchen'	34.2	34.2	77%	35	100.00%		93459789
1/2 3/3 8/1 9/1	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_030551832.1	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_030551831.1	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_030551829.1	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_030551828.1	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_030551827.1	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_030551826.1	PREDICTED: Sparus aurata cytohesin 2 (cyth2), transcript variant X2, mRNA	gilthead seabream	34.2	34.2	77%	35	100.00%		4687
XM_030394229.1	PREDICTED: Sparus aurata cytohesin 2 (cyth2), transcript variant X1, mRNA	gilthead seabream	34.2	34.2	77%	35	100.00%		4697
XM_030394228.1	Sparus aurata genome assembly, chromosome: 17	gilthead seabream	34.2	64.4	77%	35	100.00%		37948182
LR537137.1	Myripristis murdjan genome assembly, chromosome: 11	pinecone soldierfish	34.2	34.2	77%	35	100.00%		31823041
LR597560.1	Salarias fasciatus genome assembly, chromosome: 10	jewelled blenny	34.2	98.6	77%	35	100.00%		28124721
LR597445.1	Denticeps clupeioides genome assembly, chromosome: 4	denticle herring	34.2	34.2	77%	35	100.00%		35364899
LR535816.1	Macrobrachium nipponense isolate FS-2020 chromosome 28	Macrobrachium nipponense	34.2	34.2	77%	35	100.00%		75602129
CP062029.1	Aspergillus flavus strain CA14 chromosome 2	Aspergillus flavus	34.2	34.2	95%	35	95.24%		6306092
CP061805.1	Acomys russatus genome assembly, chromosome: 8	golden spiny mouse	34.2	96.6	86%	35	100.00%		77196262
LR877219.1									
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

CONSRM 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	

VISTA DI ALLINEAMENTO VERTICALE AGGANCIATO A QUERY CON LETTERE D'IDENTITA'

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	GGTCCTGACACACAAGA	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	GGTCCTGACACACAAGA	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

<u>LR812084.1</u>	12278486	CCTGACACACAAGAA	12278472
<u>XM 035010550.1</u>	9541	GGGTCCTGACACACAAG	9525
<u>CP051060.1</u>	1764302	TCGGGTGCTGACACACAAGAA	1764282
<u>CP051036.1</u>	1808656	TCGGGTGCTGACACACAAGAA	1808636
<u>CP051020.1</u>	2070075	TCGGGTGCTGACACACAAGAA	2070055
<u>CP051068.1</u>	1763679	TCGGGTGCTGACACACAAGAA	1763659
<u>CP051076.1</u>	1905409	TCGGGTGCTGACACACAAGAA	1905389
<u>CP051084.1</u>	1826983	TCGGGTGCTGACACACAAGAA	1826963
<u>CP051044.1</u>	1474962	TCGGGTGCTGACACACAAGAA	1474942
<u>CP051092.1</u>	1840194	TCGGGTGCTGACACACAAGAA	1840174
<u>CP051052.1</u>	1741091	TCGGGTGCTGACACACAAGAA	1741071
<u>CP051028.1</u>	1806024	TCGGGTGCTGACACACAAGAA	1806004
<u>CP051010.1</u>	8085892	TCGGGTCCTGACACACA	8085876
<u>CP050974.1</u>	7984726	TCGGGTCCTGACACACA	7984710
<u>CP041650.2</u>	8366161	TCGGGTCCTGACACACA	8366145
<u>XM 032393171.1</u>	9671	GGGTCCTGACACACAAG	9655
<u>CP041612.2</u>	8160571	TCGGGTCCTGACACACA	8160555
<u>CP041607.2</u>	7937840	TCGGGTCCTGACACACA	7937824
<u>CP041609.2</u>	8065585	TCGGGTCCTGACACACA	8065569
<u>CP041613.2</u>	8368507	TCGGGTCCTGACACACA	8368491
<u>CP041604.2</u>	8289159	TCGGGTCCTGACACACA	8289143
<u>CP041602.2</u>	8020607	TCGGGTCCTGACACACA	8020591
<u>CP041610.2</u>	7993768	TCGGGTCCTGACACACA	7993752
<u>XM 032318161.1</u>	9488	GGGTCCTGACACACAAG	9472
<u>XM 032318160.1</u>	9515	GGGTCCTGACACACAAG	9499
<u>XM 032318159.1</u>	9667	GGGTCCTGACACACAAG	9651
<u>XM 032318158.1</u>	9530	GGGTCCTGACACACAAG	9514
<u>CP047558.1</u>	13102812	TCGGGTCCTGACACACA	13102796
<u>CP047558.1</u>	23959045	TCGGGTCCTGACACACA	23959059
<u>LR738417.1</u>	15156877	GGGTCCTGACACACAAG	15156861
<u>LR738417.1</u>	58724401	GTCCTGACACACAAGAA	58724417
<u>LR738417.1</u>	61670400	GGGTCCTGACACACA	61670414
<u>CP047250.1</u>	1740771	TCGGGTGCTGACACACAAGAA	1740751
<u>CP044622.1</u>	1812873	TCGGGTGCTGACACACAAGAA	1812853
<u>XM 031040735.1</u>	9464	GGGTCCTGACACACAAG	9448
<u>CP035529.1</u>	1961507	TCGGGTGCTGACACACAAGAA	1961487
<u>LR699164.1</u>	34911047	GGGTCCTGACACACAAG	34911031
<u>LR697115.1</u>	32421543	TCGGGTCCTGACACACA	32421527
<u>LR697115.1</u>	27036679	GGTCCTGACACACAAG	27036664
<u>LR697115.1</u>	4290476	GGTCCTGACACACAA	4290462
<u>LR537134.1</u>	2921913	GGTCCTGACACACAAGA	2921897
<u>CP041651.1</u>	7902720	TCGGGTCCTGACACACA	7902704
<u>CP041611.1</u>	8165294	TCGGGTCCTGACACACA	8165278
<u>CP041654.1</u>	8164184	TCGGGTCCTGACACACA	8164168
<u>XM 029457131.1</u>	140	TCGGGTCCTGACACACA	124
<u>LR594556.1</u>	53940900	GGGTCCTGACAAACAAGAA	53940882
<u>LR594556.1</u>	1399214	TCGGGTCCTGACACACA	1399198
<u>LR132013.2</u>	14345330	CGGGTCCTGACACACAA	14345314
<u>XM 028096253.1</u>	8289	GGGTCCTGACACACAAG	8273
<u>XM 036099982.1</u>	9366	GGGTCCTGACACACAAG	9350
<u>CP059859.1</u>	1812250	TCGGGTGCTGACACACAAGAA	1812230
<u>CP059867.1</u>	1812830	TCGGGTGCTGACACACAAGAA	1812810
<u>XM 027612786.2</u>	9360	GGGTCCTGACACACAAG	9344
<u>LR812059.1</u>	22791268	GTCCTGACACACAAGAA	22791252
<u>LR812590.1</u>	17194139	GTCCTGACACACAAGAA	17194123
<u>LR812590.1</u>	13780417	CCTGACACACAAGAA	13780431
<u>LR812565.1</u>	18830476	GTCCTGACACACAAGAA	18830460
<u>AP023164.1</u>	71550743	GGTCCTGACACACAAGA	71550759

<u>AP023164.1</u>	43772683	GGTCCTGACACACAA	43772669
<u>AP023164.1</u>	76006964	CCTGACACACAAGAA	76006950
<u>AP023164.1</u>	89775056	GGGTCCTGACACACA	89775070
<u>XM 034761652.1</u>	1347	GGGTCCTGACACACAAG	1363
<u>XM 034761651.1</u>	1347	GGGTCCTGACACACAAG	1363
<u>XM 034761650.1</u>	1347	GGGTCCTGACACACAAG	1363
<u>XR 004644615.1</u>	1347	GGGTCCTGACACACAAG	1363
<u>XM 034761648.1</u>	1347	GGGTCCTGACACACAAG	1363
<u>XM 034761647.1</u>	1347	GGGTCCTGACACACAAG	1363
<u>LR792560.1</u>	5337604	CGGGTCCTGACACACAA	5337620
<u>LR792560.1</u>	25042884	TCCTGACACACAAGA	25042870
<u>CP050600.1</u>	63813944	GGTCCTGACACACAAGA	63813960
<u>CP050600.1</u>	6594778	GTCCTGACACACAAG	6594764
<u>CP050600.1</u>	7332222	GGGTCCTGACACACA	7332236
<u>CP050630.1</u>	63791832	GGTCCTGACACACAAGA	63791848
<u>CP050630.1</u>	6585343	GTCCTGACACACAAG	6585329
<u>CP050630.1</u>	7323174	GGGTCCTGACACACA	7323188
<u>XM 032790304.1</u>	1858	GGGTCCTGACACACAAG	1874
<u>XM 032790303.1</u>	1858	GGGTCCTGACACACAAG	1874
<u>XM 032790302.1</u>	1858	GGGTCCTGACACACAAG	1874
<u>XM 032790301.1</u>	1858	GGGTCCTGACACACAAG	1874
<u>XM 032790300.1</u>	1858	GGGTCCTGACACACAAG	1874
<u>LR778253.1</u>	74667960	GTCCTGACACACAAGAA	74667976
<u>XM 030551832.1</u>	1874	GGGTCCTGACACACAAG	1890
<u>XM 030551831.1</u>	1874	GGGTCCTGACACACAAG	1890
<u>XM 030551829.1</u>	1874	GGGTCCTGACACACAAG	1890
<u>XM 030551828.1</u>	1874	GGGTCCTGACACACAAG	1890
<u>XM 030551827.1</u>	1897	GGGTCCTGACACACAAG	1913
<u>XM 030551826.1</u>	1874	GGGTCCTGACACACAAG	1890
<u>XM 030394229.1</u>	1874	GGGTCCTGACACACAAG	1890
<u>XM 030394228.1</u>	1884	GGGTCCTGACACACAAG	1900
<u>LR537137.1</u>	23280262	GGGTCCTGACACACAAG	23280278
<u>LR537137.1</u>	15586339	GTCCTGACACACAAG	15586325
<u>LR597560.1</u>	7549408	CGGGTCCTGACACACAA	7549424
<u>LR597445.1</u>	4090641	GTCCTGACACACAAGAA	4090657
<u>LR597445.1</u>	21828624	TCCTGACACACAAGAA	21828609
<u>LR597445.1</u>	24913684	TCCTGACACACAAGAA	24913699
<u>LR535816.1</u>	2764452	GGGTCCTGACACACAAG	2764468
<u>CP062029.1</u>	57872644	CGGGTCCTGACACACAA	57872660
<u>CP061805.1</u>	4502060	TCGGGTGCTGACACACAAGAA	4502080
<u>LR877219.1</u>	73395407	GGGTGCTGACACACAAGAA	73395389
<u>LR877219.1</u>	67213109	GGGTCCTGACACACAAG	67213125
<u>LR877219.1</u>	71189442	GTCCTGACACACAAGA	71189427
<u>CP046697.1</u>	13106661	TCGGGTCCTGAYACACA	13106645

PRODOTTI DEGLI ALLINEAMENTI SIGNIFICATIVI (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](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_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_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_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_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_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_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_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_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_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_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_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_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_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_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_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](https://www.ncbi.nlm.nih.gov/nucleotide/XM_030551826.1?report=genbank&log$=nuclalign&blast_rank=89&RID=W6XMJ10P013&from=1874&to=1890)

SEQUENZA N° 3/2

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

SEQUENZA N° 3/2 (T=2; C=0; G=3; A=1)

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° 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
Sequenza 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: **TATCCATTACATGCTCAACAA**

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	Zea mays	38.2	38.2	86%	2.2	100.00%	1784

	variant X2, mRNA								
	PREDICTED: Zea mays B3 domain-containing protein IDEF1 (LOC103628923), transcript								
XM_020539751.2	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 RP11-47311, complete sequence	human	36.2	36.2	81%	8.9	100.00%		159712
LR812129.1	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),	cotton	36.2	36.2	81%	8.9	100.00%		1309

	mRNA								
XM_012581734.1	PREDICTED: Gossypium raimondii transcription factor bHLH19-like (LOC105763496), mRNA	Gossypium raimondii	36.2	36.2	81%	8.9	100.00%		1622
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 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	Erithacus rubecula genome assembly, chromosome: 6	European robin	34.2	94.6	86%	35	100.00%		60684490
LR812106.1	Erithacus rubecula genome assembly, chromosome: 4	European robin	34.2	64.4	95%	35	95.24%		68599267
XM_034203981.1	PREDICTED: Gymnodraco acuticeps saccin-like (LOC117538367), mRNA	Gymnodraco acuticeps	34.2	34.2	77%	35	100.00%		9925
XM_034203980.1	PREDICTED: Gymnodraco acuticeps saccin-like (LOC117538366), mRNA	Gymnodraco acuticeps	34.2	34.2	77%	35	100.00%		11751
XM_034203979.1	PREDICTED: Gymnodraco acuticeps saccin (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 saccin (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 saccin-like (LOC117485916), mRNA	emerald rockcod	34.2	34.2	77%	35	100.00%		5584
XM_034134621.1	PREDICTED: Trematomus bernacchii saccin-like (LOC117485915), mRNA	emerald rockcod	34.2	34.2	77%	35	100.00%		12460
XM_034134161.1	PREDICTED: Trematomus bernacchii saccin-like (LOC117485571), mRNA	emerald rockcod	34.2	34.2	77%	35	100.00%		14753
XM_034110707.1	Pseudochaenichthys georgianus saccin (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

	mRNA								
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 saccin-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	Rosevirga pacifica strain 9dgc1 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	Aquila chrysaetos chrysaetos	Aquila chrysaetos	34.2	34.2	95%	35	95.24%		29699254

1/2 3/3 3/5 9/1	genome assembly, chromosome: 17	chrysaetos							
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	

VISTA DI ALLINEAMENTO VERTICALE AGGANCIATO A QUERY CON LETTERE D'IDENTITA'

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

<u>LR880663.1</u>	20588898	CCCATTACATGCTCAACA	20588881
<u>AC087190.5</u>	95049	TATCCCATTACATGCTCA	95032
<u>LR812129.1</u>	6141125	TATCCCATTACATGCTCA	6141142
<u>LR812129.1</u>	4095966	TCCCATTACATGCTC	4095980
<u>AP022681.1</u>	42443338	CCATTACATGCTCAACAA	42443355
<u>CP032564.1</u>	37529827	ATCCCATTACATGCTCAA	37529844
<u>CP032564.1</u>	44871891	TTACATGCTCAACAA	44871877
<u>LR584085.1</u>	17507700	TATCCCATTACATGCTCA	17507717
<u>LR584085.1</u>	8204435	CCATTACATGCTCAA	8204421
<u>LR584085.1</u>	14388452	TCCCATTACATGCTC	14388466
<u>LR584085.1</u>	14575511	TCCCATTACATGCTC	14575525
<u>LR584069.1</u>	15589628	ATCCCATTACATGCTAAACA	15589609
<u>LR584069.1</u>	36872323	CCATTACATGCTCAACAA	36872340
<u>LR584069.1</u>	24491114	TCCCATTACATGCTC	24491100
<u>CP034494.1</u>	52008196	TCCCATGACATGCTCAACA	52008178
<u>CP034494.1</u>	9113208	TATCCCATTACATGCTCA	9113225
<u>CP034494.1</u>	66183523	CCATTACATGCTCAA	66183537
<u>XM 020714013.2</u>	1121	CCATTACATGCTCAACAA	1138
<u>XM 023953470.1</u>	1133	CCATTACATGCTCAACAA	1150
<u>XM 011489176.3</u>	1413	CCATTACATGCTCAACAA	1430
<u>XM 023953469.1</u>	1425	CCATTACATGCTCAACAA	1442
<u>XM 004066996.4</u>	1425	CCATTACATGCTCAACAA	1442
<u>XM 023953468.1</u>	1477	CCATTACATGCTCAACAA	1494
<u>XM 023953467.1</u>	1488	CCATTACATGCTCAACAA	1505
<u>XM 023953466.1</u>	1467	CCATTACATGCTCAACAA	1484
<u>XM 016879281.1</u>	349	ATCCCATTACATGCTCAA	366
<u>XM 012581734.1</u>	662	ATCCCATTACATGCTCAA	679
<u>CP009355.1</u>	1269741	CCATTACATGCTCAACAA	1269758
<u>CP061205.1</u>	3194946	CCATTACATGCTCAACAA	3194963
<u>CP027557.1</u>	4946095	CCCATTACATGCTCAACA	4946112
<u>XM 036210237.1</u>	1692	CCATTACATGCTCAACAA	1709
<u>XM 024295636.2</u>	1704	CCATTACATGCTCAACAA	1721
<u>XM 024295635.2</u>	1757	CCATTACATGCTCAACAA	1774
<u>XM 024295634.2</u>	1766	CCATTACATGCTCAACAA	1783
<u>AP023476.1</u>	56430623	TCCCATGACATGCTCAACA	56430605
<u>AP023476.1</u>	9156872	TATCCCATTACATGCTCA	9156889
<u>AP023476.1</u>	70607256	CCATTACATGCTCAA	70607270
<u>LR812108.1</u>	14024173	TATCCCATTACATGCTC	14024157
<u>LR812108.1</u>	55816198	CCATTACATGCTCAA	55816184
<u>LR812108.1</u>	25243419	ATCCCATTACATGCT	25243433
<u>LR812106.1</u>	17540369	ATCCCATTACTTGCTCAACAA	17540349
<u>LR812106.1</u>	1724166	TTACATGCTCAACAA	1724152
<u>XM 034203981.1</u>	1725	CATTACATGCTCAACAA	1709
<u>XM 034203980.1</u>	3355	CATTACATGCTCAACAA	3339
<u>XM 034203979.1</u>	3340	CATTACATGCTCAACAA	3324
<u>XM 034203978.1</u>	3340	CATTACATGCTCAACAA	3324
<u>XM 034134622.1</u>	3322	CATTACATGCTCAACAA	3306
<u>XM 034134621.1</u>	257	CATTACATGCTCAACAA	241
<u>XM 034134161.1</u>	3353	CATTACATGCTCAACAA	3337
<u>XM 034110707.1</u>	3324	CATTACATGCTCAACAA	3308
<u>LR792566.1</u>	22736379	CATTACATGCTCAACAA	22736363
<u>CP048847.1</u>	30598354	TCCCATTACATGCTCAA	30598338
<u>CP048847.1</u>	34594943	CCATTACATGCTCAA	34594929
<u>AP022682.1</u>	44113831	CATTACATGCTCAACAA	44113815
<u>XM 031810749.1</u>	2203	CCATTACATGCTCAACA	2187
<u>XM 031810748.1</u>	2399	CCATTACATGCTCAACA	2383
<u>XM 020467159.2</u>	2328	CCATTACATGCTCAACA	2312
<u>XM 020467158.2</u>	2317	CCATTACATGCTCAACA	2301

<u>XM 020467155.2</u>	2213	CCATTACATGCTCAACA	2197
<u>XM 020467160.2</u>	2428	CCATTACATGCTCAACA	2412
<u>CP032578.1</u>	31033463	TCCCATTACATGCTCAA	31033447
<u>CP032578.1</u>	53587352	TTACATGCTCAACAA	53587338
<u>LR633960.1</u>	22633765	TCCCATTACATGCTCAA	22633749
<u>LR633960.1</u>	22634484	TCCCATTACATGCTCAA	22634468
<u>LR633953.1</u>	19265064	CCATTACATGCTCAACA	19265048
<u>LR597560.1</u>	14470542	ATCCCATTACAAGCTCAACAA	14470522
<u>XM 029459375.1</u>	135	CATTACATGCTCAACAA	119
<u>CP032560.1</u>	23750225	TCCCATTACATGCTCAA	23750209
<u>LR584231.2</u>	1752792	CATTACATGCTCAACAA	1752776
<u>CP028923.1</u>	2912772	CATTACATGCTCAACAA	2912756
<u>CP041165.1</u>	2188370	TCCCATTACATGCTCAA	2188354
<u>CP060294.1</u>	4655806	CCATTACATGCTCAACA	4655790
<u>LR877197.1</u>	75781442	CATTACATGCTCAACAA	75781426
<u>LR877197.1</u>	55851315	ATCCCATTACATGCT	55851301
<u>LR877197.1</u>	78953222	ATCCCATTACATGCT	78953208
<u>LR877197.1</u>	45008148	CATTACATGCTCAAC	45008162
<u>LR877197.1</u>	50164947	ATCCCATTACATGCT	50164961
<u>LR877197.1</u>	76807168	ATCCCATTACATGCT	76807182
<u>LR862394.1</u>	2205206	TCCCATTACATGCTCAA	2205190
<u>XM 034535788.1</u>	2631	TATCCCATTACATGGTCAACA	2651
<u>XM 034535769.1</u>	2828	TATCCCATTACATGGTCAACA	2848
<u>LR792831.1</u>	16950994	CATTACATGCTCAACAA	16951010
<u>LR761616.1</u>	16950994	CATTACATGCTCAACAA	16951010
<u>CP047344.1</u>	2305404	CCATTACATGCTCAACA	2305420
<u>CP047344.1</u>	2022003	TTACATGCTCAACAA	2022017
<u>AP019652.1</u>	979168	CATTACATGCTCAACAA	979184
<u>AP019650.1</u>	977013	CATTACATGCTCAACAA	977029
<u>LR738604.1</u>	43729402	TATCCCATTACATGTTCAACA	43729422
<u>AP018694.1</u>	1891606	CCCATTACATGCTCAAC	1891622
<u>CP045339.1</u>	404180	CATTACATGCTCAACAA	404196
<u>LR698965.1</u>	1848004	TCCCATTACATGCTCAA	1848020
<u>LR633950.1</u>	27856735	TCCCATTACATGCTCAA	27856751
<u>LR633950.1</u>	27857454	TCCCATTACATGCTCAA	27857470
<u>LR633950.1</u>	10352973	TATCCCATTACATGCT	10352988
<u>LR606197.1</u>	6047717	ATCCCATTACTTGCTCAACAA	6047737
<u>LR597448.1</u>	19915008	CATTACATGCTCAACAA	19915024
<u>AP019297.1</u>	4751659	CATTACATGCTCAACAA	4751675
<u>AP019297.1</u>	30436690	CCCATTACATGCTCA	30436676
<u>AP019297.1</u>	17094636	TTACATGCTCAACAA	17094650
<u>CP064939.1</u>	3184051	CCATTACATGCTCAACA	3184067
<u>LR813497.2</u>	978326	CCCATTACATGCTCAAC	978342
<u>LR884475.1</u>	22370334	CATTACATGCTCAACAA	22370350
<u>CP051488.1</u>	4062423	CCATTACATGCTCAACA	4062439
<u>LR862150.1</u>	14830253	CATTACATGCTCAACAA	14830269
<u>AP022274.1</u>	3949516	CCATTACATGCTCAACA	3949532
<u>LR812551.1</u>	30495017	CATTACATGCTCAACAA	30495033
<u>LR812551.1</u>	20403021	CCATTACATGCTCAAC	20403036
<u>CP046702.1</u>	2199171	CATYACATGCTCAACAA	2199187
<u>CP046702.1</u>	35161657	TCCCATTACATGCTC	35161643

PRODOTTI DEGLI ALLINEAMENTI SIGNIFICATIVI (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: sacin-like** [GenBank](#); [GenBank](#); [GenBank](#); [GenBank](#); [GenBank](#); [GenBank](#); [GenBank](#)
[GenBank](#)
- **Sacin** [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](#)

SEQUENZA N° 3/2 (T=2; C=0; G=3; A=1)

Query: **TATCCATTACATGATATTATT**

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	Pipistrellus pipistrellus genome assembly, chromosome: 12	common pipistrelle	32.2	124	90%	138	100.00%	58291996
1/2 2/2 3/3 3/5 6/1 7/2 8/1 9/1 10/1								
LR862363.1	Pipistrellus pipistrellus genome assembly, chromosome: 7	common pipistrelle	32.2	96.6	95%	138	100.00%	84548104
1/2 2/2 3/3 3/5 6/1 7/2 8/1 9/1 10/1								
.....
LR862367.1	Pipistrellus pipistrellus genome assembly, chromosome: 11	common pipistrelle	32.2	62.4	90%	138	100.00%	69386144
1/2 2/2 3/3 3/5 6/1 7/2 8/1 9/1 10/1								
LR862366.1	Pipistrellus pipistrellus genome assembly, chromosome: 10 Pipistrellus pipistrellus genome assembly, chromosome: 6	common pipistrelle	32.2	62.4	72%	138	100.00%	71504835
1/2 2/2 3/3 3/5 6/1 7/2 8/1 9/1 10/1								
LR862362.1	Pipistrellus pipistrellus genome assembly, chromosome: 6	common pipistrelle	32.2	62.4	95%	138	100.00%	85793147
1/2 2/2 3/3 3/5 6/1 7/2 8/1 9/1 10/1								

SEQUENZA N° 3/3

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

SEQUENZA N° 3/3 (T=2; C=0; G=3; A=1)

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° 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
Sequenza 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: **ACTCCATTACATGCTATTCAA**

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 Schistosoma mansoni strain Puerto Rico genome assembly, chromosome: 4	Timema shepardii	38.2	38.2	86%	2.2	100.00%	371
HE601627.2	Plectropomus leopardus DNA, chromosome 22, nearly complete sequence	Schistosoma mansoni leopard coralgroup	38.2	38.2	86%	2.2	100.00%	47279781
AP022721.1			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 Eukaryotic synthetic construct	Ipomoea trifida eukaryotic	36.2	66.4	81%	8.9	100.00%	28619729
CP034499.1	chromosome 20	synthetic construct	36.2	96.6	86%	8.9	100.00%	68480253
CP023762.1	Solanum lycopersicum cultivar I-3 chromosome 6 PREDICTED: Monodelphis domestica uncharacterized LOC103105194 (LOC103105194), ncRNA	tomato	36.2	36.2	81%	8.9	100.00%	49756038
XR_470386.2	Apteryx australis mantelli genome assembly AptMant0, scaffold scaffold806	gray short-tailed opossum	36.2	36.2	81%	8.9	100.00%	530
LK065388.1	Solanum lycopersicum chromosome ch06, complete genome	Apteryx australis mantelli	36.2	36.2	81%	8.9	100.00%	1316800
HG975518.1	tomato	tomato	36.2	36.2	81%	8.9	100.00%	46045610
HG975445.1	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	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	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	Canis lupus familiaris breed Labrador retriever chromosome 15a	dog	34.2	94.6	100%	35	100.00%	64284575
CP050620.1	Canis lupus familiaris breed Labrador retriever chromosome 15b	dog	34.2	94.6	100%	35	100.00%	64292917
LR778278.1	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	Solanum pinnatisectum cultivar CGN17745 chromosome 1	tansyleaf nightshade	34.2	34.2	77%	35	100.00%	88611892
CP047558.1	Solanum pinnatisectum cultivar CGN17745 chromosome 12	tansyleaf nightshade	34.2	66.4	77%	35	100.00%	61096591
	Solanum tuberosum cultivar P8	potato	34.2	66.4	77%	35	100.00%	45472623

CP046684.1	chromosome 11								
1/2 2/2 3/2 3/5 6/1 8/1 9/1 10/1									
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	Solanum tuberosum cultivar MSH/14-112 chromosome 1	potato	34.2	34.2	77%	35	100.00%	88648484	
1/2 2/2 3/2 3/5 6/1 8/1 9/1 10/1									
CP046699.1	Solanum tuberosum cultivar MSH/14-112 chromosome 4	potato	34.2	34.2	77%	35	100.00%	72170609	
1/2 2/2 3/2 3/5 6/1 8/1 9/1 10/1									
CP046697.1	Solanum tuberosum cultivar MSH/14-112 chromosome 12	potato	34.2	66.4	77%	35	100.00%	61134531	
1/2 2/2 3/2 3/5 6/1 8/1 9/1 10/1									
CP046691.1	Solanum tuberosum cultivar MSH/14-112 chromosome 11	potato	34.2	66.4	77%	35	100.00%	45472774	
1/2 2/2 3/2 3/5 6/1 8/1 9/1 10/1									
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	Chanos chanos genome assembly, chromosome: 2	milkfish	34.2	94.6	100%	35	100.00%	59736555	
1/2 2/2 3/1 3/4 8/1									
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	Gadus morhua genome assembly, chromosome: 15	Atlantic cod	34.2	34.2	77%	35	100.00%	28657694	
1/2 2/2 3/2									
LR633944.1	Gadus morhua genome assembly, chromosome: 2	Atlantic cod	34.2	98.6	90%	35	100.00%	28732775	
1/2 2/2 3/2									
LR606182.1	Aquila chrysaetos chrysaetos genome assembly, chromosome: 2	Aquila chrysaetos chrysaetos	34.2	96.6	90%	35	100.00%	83001952	
1/2 3/2 3/5 9/1									
LR597457.1	Salarias fasciatus genome assembly, chromosome: 23	jewelled blenny	34.2	94.6	77%	35	100.00%	41705463	
CP040924.1	Clostridium therrmarum 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	
Salmo trutta genome assembly,	river trout	river trout	34.2	34.2	77%	35	100.00%	41918599	

LR584444.1	chromosome: 35								
LR584420.1	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 SCY1 like pseudokinase 2 (<i>scyl2</i>), transcript variant X2, mRNA	blunt-snouted clingfish	34.2	34.2	77%	35	100.00%	6817	
XM_028449732.1	PREDICTED: Gouania willdenowi SCY1 like pseudokinase 2 (<i>scyl2</i>), 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	Gouania willdenowi genome assembly, chromosome: 6	blunt-snouted clingfish	34.2	66.4	90%	35	100.00%	72451976	
LR131993.1	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 HSOX 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	uncultured	34.2	34.2	77%	35	100.00%	592	

CP062013.1	DGGE gel band AOA-a ammonia monooxygenase gene, partial cds Macrobrachium nipponense isolate FS-2020 chromosome 22	archaeon 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 Acomys russatus genome assembly, chromosome: 7	potato	34.2	34.2	77%	35	100.00%	72008707
LR877218.1		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
source      Location/Qualifiers
            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="IGSDYEETLKPSSIVGTSSTSTCTPISQRKETSTRRSSDHRATSS
TLVQDNTVPSGGTEEFPPPPSSLQTPIDVMAFSQSPEFPSPRMPVSKELCSKQRN
LYELNHLYKHIHPELRKNLEKDYISQVSEIVSSQVNPSSSVSADVQQARYVFENTNIS
CPKGLNSEREYLECNEILKGEVQSMRWIFENQPLDAINNGSPEEGNTXKGIADQEIIA
VGDVKYTTWMFETQPIDMLGDHSSGTEENAEEKIPELARGDVCTARWMFETKPLDSMNK
MHPSQEEPQVSSIKDIIIGGDVKTVRYMFETLHLDQLGQFHTVDQVHLLQLRSELQEIK
GNVKQSIKCFETQPFVIRDLGLGQTLKIKTVHREDVEKGDVKTAQGLTQPLDTINKD
ITEIKVIQEI SMEEKVKGVSRAKWL FETQPLEKIKESEEDMIEKETIMGTDVSRKWCW
MFETQPSDILKEVPDADSVRSEEEIIGGDVQTTKPLFQTLPIEALKDSPDVGKQLKITT
SEEEKGDVRHQKWI FETQPLEEIREDKKEYIRTVKLEEVDRGDIRNYTNTFESNNLIK
FDASHKIEVKVTRGAVELNKS L FETTPHAIQDHLGKCHQVKTVQQEELRGEVRSRSC
RWFETRPIDQFDESIHNFQIIRGISAQDIQTGNVKS AKWLFETQPLDSIKYFNSMEV
ESKIEQATDIIKGNVKTCSQPMESLYEKVSLMTDSEEIPKGDIACTWLFETRPLDTI
KDDSEETVKSQAVKQEEINGVDVCSACFLFETENLDSIKGEEGKEIKAVEVDIQAGDV
SSMRHKFENQSLDSIRSSEVLRKITSPATLCRQIKIETRRRDSPPITIPVSVNHVD
SSSFRESMEAQEEVRKVGRATYIHKDRISSSGCI GPDTECFDAVEI IHKVEGPLRSE
HTQRXEAANQTVQMAENFMNDCENEINRWXREFEDGPVLEAKSKRRVYVNGEANHNIT
LESHTFCKEEFGLTSSSENSSFTGFSGNHPRELQEKIPVKQPSICSETRYLSEHFSGVD
AFESQVVGKMTVSSPHNSEAGKSGFDFKHAPPTYEDVIAGHILDVSDSPKELRRNFQ
EMWQESQRVFKKLG YATS DSSATEMRTAFQEESAFISDKEICILCQKTVYP EECLIAD
MQIFHESCFRCHNCD SKLSLGYASLHGQIYCKPHFEQLFKSKGNYDEGFGHKQHKDR
WNCKIQSSSVDCIPNEAPNMCKNSIENTPTLRDLNKHLD AENSKGQRDELKRVERRK
LKI IWPPSRKLP EKTFFLEELKMSKPKWPPEMTAPASAEFKDELREHNNTLENKGQE
QDTSLLETYLQHIHMCQKDV TGIQEMEXYKAKXDEKEGSKNVPDNLNEAEEIKSKRKS
EMDVSKNNIGVVQIAEEKNEPGNELDDAEVIQVTDTDDEVV PENHKENLNKNNNYVA
VSDLNNGRQKTSILEFPNLLPLSSEANCVASDYVIENLESVPRI SEETYSRNVPAMAC
KKQAGRATAGSPVQSAPKSVLSSSLRGRNSSMASPPNTNTLN IKENHNSNNKLCFFF
SNTVKIAAFSKKNEDPD LIESVDPIKNMPCLYLRELGN I KHHWGEAGAAGNTSFDAL
SSECATKPSFPRVEDQFEQLTVEEQIKRNR CYSETE"

```

ORIGIN

```

1 attggaagtg actatgaaga gactcttaag ccatcatcca ttgtgggtac ctcttcgact
61 tcttgactc caatcagcca gaggaaggaa acatcgacca gacgatctag tgaccacagg

```

121 gccacctcct caaactggt gcaagataat accgttcctt caggagggac agaagaattt
181 cctcctcccc caccagcag cttcaaaact ccaatagatg tgatggcatt ttcccagtcc
241 cctgaattcc ccagccctcc tagaatgcca ccagtctcca aagaattatg ttctaaacaa
301 agaaatthgt atgagctaaa ccatttatat aaacatattc atccggagtt aaggaagaac
361 ttagaaaaag attatatcag tcaggtttct gagattgttt ctagtcaagt gaatccaagc
421 agctcagttt ctgcagatgt tcagcaagcc cggatggtt ttgaaaacac aaatatcagt
481 tgtccaaaag gcctgaactc agaaagagaa ttttggagt gtaatgagat tctgaagggg
541 gaggtgcagt ccatgaggtg gatthttgaa aatcagccat tagatgccat caacaatggc
601 tctccagagg aaggtaacac tnccaagggc attgctgatc aagaaatcat tgctgthgt
661 gatgtgaaat ataccacatg gatgthttgaa acacaacca tagatatgct gggggatcat
721 tcttctggta ctgaagaaaa tgctgagaaa attcctgagc tagccagagg agatgthttg
781 acagccagat ggatgthttga acaaaagcca ttagattcaa tgaataaaat gcatccaagt
841 caagaagaac caggggtatc ttccataaag gatataattg gcggggatgt caagactgtg
901 agatacatgt ttgaaactct acatctggat caacttggac agthttcatalc agthggatcag
961 gthcacttac tgcagctcag gtctgagctc caagaaatta agggaaatgt taagcaaagc
1021 ataaaatgth ttgaaactca accattctgt gthattagag atgthtttag tcaaacgcta
1081 gaaatthaaa ctgthtcacag agaagatgth gaaaagggg atgtgaaaac agcacaaggg
1141 cthgaaacac agcccttggc cacaatcaat aaggatatca cagaaatthaa agthcatcaa
1201 gaaatatcca tggagaaaaa ggtcaaaggt ggggtgagca gggcaaagt gthgthttgaa
1261 actcagcctt tggagaaaaat caaagagtca gaagaggaca tgattgagaa ggaacaatt
1321 atgggtacag atgtctccag aaagtgttg atgthttgaaa cacagccatc agacattcta
1381 aaggaagthc ccgacgcaga thctgthaga tctgaagaga thaatagggg cgatgtacaa
1441 actaccaagc ctctattthca aacactthca atagaagctt thaaaagatag cctgthgta
1501 gthaaactthc aaaaaattac cacctctgaa gaagaaaagg gggatgthtag gcacaaaaag
1561 tggatthttcg aaaccaacc thctggaagag attagagaag ataaaaaga gthacatacga
1621 acagtgaaac thgaagaggt tgacagagga gatatcagaa atthacacaaa thcctthtag
1681 thaaacaatt thaatthaaat tgatgcatca cataagatag aggtgaaagg agthcacaaga
1741 gthgctgtag aatthaaata atcactctth gaacaacac cthacatgc thattcaagat
1801 cacctthgaa aatgthcatca gthaaagaca gthccagcaag aagaaatcct aagaggtgaa
1861 gthagaagct gcaggtggct thttgthaaaca aggcccattg accagthttga thgaaagcatt
1921 cataactthc aataatcag aggaatatct gthcaggaca thacagactgg aatgthgaa
1981 thctgthagt gthgthttga aaccaacct cthgattcaa thaaatthtt tagthaatag
2041 gaagthgaaa gthaaattga acaagctaca gatatcatta aagggatgt caaacctgt
2101 agccagccaa tggagthctct thtatgthaaa gthctcatta thgacagacag thgagaaatt
2161 cthaaagggag acatcaaagc ctgtactthg ctctthgaaa ctcgcccct thgataccata
2221 aaagatgact ctgaagaaac agthcaaatca caagctgthaa aacaggagga gthcaatggt
2281 gthagatgtht gthcagcgtg thttctthtt gagacagaaa atctggacag cthaaaagga
2341 gaagaaggaa aggaaatcaa ggctgthgaa gthagatthcc aagctgggga thgtctccagc
2401 atgctggcata aatthgthaaa thcagthctta gactctatca gthctcaga ggaagthttg
2461 agaaagatca catctctgth aacactthgt cgcaaatta agatagaaac thgtctgtagg
2521 gactctccac ctataattac aataccgtha agthgthaaatc atgthgthtag thgtctctc
2581 agagaatcca tggagcgcga agaggaagth aggaaagthag gthaaaagggc gactthacatc
2641 cthaaagaca gaataagthc ctctgggtg atagggccag atactgthag thttgacgca
2701 gthgaaatca thcacaagth thgagggccc ctgagthcag agcacacaca gthagthgaa

2761 gcagccaacc aaactgttca aatggctgaa aatttcatga atgactgtga aatgaaata
 2821 aacagatggn tcaggaatt tgaggatggc ccagtttgg aagcaaagtc aaagagaaga
 2881 gtttatgtaa atggagaagc aaaccataat ataactag aaagtcatac attttgtaag
 2941 gaggaatttg gattaacatc ttcagagaac tctagtttta caggcttttc aggcaatcat
 3001 cctagggaac tgcaagaaaa aattcctggt aagcagccta gcatatgctc tgaacaagc
 3061 tacctaagtg agcatttctc aggtgtggat gcatttgaga gtcaagttgt ggggttgaag
 3121 atgacagttt cttcaccaca taactcagaa gctggcaaat ctggctttga cttcaagcat
 3181 gccccaccaa cctatgaaga cgtcatcgtc ggccacattt tagatgtttc tgattcacct
 3241 aaagaactca ggaggaattt tcaagagatg tggcaggaga gtcaaagagt ctttaaaaa
 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 tcaactcatg gacaaatata ttgtaaacca
 3541 cactttgagc aacttttcaa atctaaagga aattatgatg aaggtttgg acacaaacag
 3601 cataaagacc gatggaattg caaaattcag agcagctcag tggactgtat tcctaagaa
 3661 gcacaaata tgtgtaaaaa cagtatagaa aacaccccca cacttaggga tcttaacaaa
 3721 catttagatg ctgagaacag caaagggcaa agggatgagc tgagaaaagt caggagagg
 3781 agaaaactaa aatcatttg gccccatct aggaagctgc ctgagaaaac ctttcccctt
 3841 gaggaagaac tcaaatgag caaacggaaa tggccacctg aatgacagc ccctgcatct
 3901 gctgaattta aagatgaact aagagaacac aataaactc tggaaaataa aggacaagag
 3961 caagatactt ctctccttga gacatatctg cagcacatcc atatgtgtca gaaagatgtt
 4021 acaggaatcc aagaaatgga angttataaa gcaaagtaag atgagaagga agggagtaag
 4081 aatgtgccag ataactgaa tgaggctgaa gaaataaaga gtaagaggaa aagtgaaatg
 4141 gatgttagca agaataatat tggggttgtt cagatagctg aggagaaaaa tgaaccaggt
 4201 aatgaacttg atgatgcaga agttatacag gttactgaca ctgatgatga ggtggtacca
 4261 gaaaatcata aagagaattt gaataagaag aataataatt atgtagcagt ctcgatctg
 4321 aataatggca ggcagaagac atctatttta gaatttctta atctattacc actgtcaagt
 4381 gaagcaaact gtgttgcaag tgactatgtg attgaaaact tagaaagtgt acccaggatc
 4441 tctgaagaga cgtactcgag gaatgtacca gctatggctt gcaagaagca ggctggcaga
 4501 gctactgctg gcagtctgt gcagtctgcc ccaagtcag tcctcagtag cagcttaaga
 4561 ggaaagagga acagctcaat ggctctcct ctaacacaa ataccttaa catcaaagaa
 4621 aaccactcaa ataacaagaa tctatgcttc ttcttttcta aactgtgaa aattgctgct
 4681 ttttccaaga agaatgagga tctgattta atagagtctg tagatccaat taaaacatg
 4741 ccatgtttgt atttaagaga acttgggaag aacatcaaac attggcatgg tgaagcagga
 4801 gcagcaggaa acacaagctt tgatgctctg agcagtgaat gtgcaactaa gccttcattt
 4861 ccagagtgg aggaccagtt tgaacaactc actgtggaag agcaaattaa aaggaatagg
 4921 tgctacagtg aaacagagta aaacatctct ggccacttgc agtggacttc aggttggaa
 4981 acacgcaggc atggacttta gggatttga tgacgtttag tagaacttgt aagcttaatg
 5041 ttcagaaatc tcatgtctat cacaatggga tattccttgt gttacaccgt atgttttcc
 5101 caattggttt gaatctactt ttaatccgaa tggaatgaaa aaaaatgaaa caatattgct
 5161 atgttttctg tcaagtggca cttacatgta tcgctctctt tttcaacatg ccagagaaca
 5221 tcaccaggga gattctattc tgctaagtgt tatcataatt cttatacctt ct**ctataa**att
 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](#)

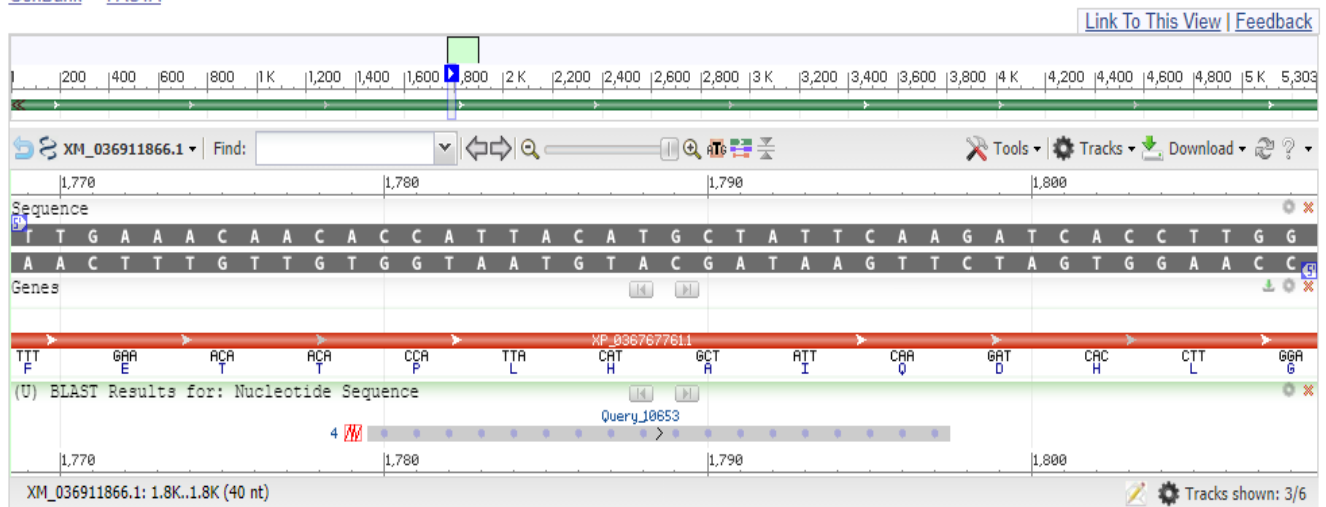
Score Expect Identities Gaps Strand
36.2 bits(18) 9.1 18/18(100%) 0/18(0%) Plus/Plus

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	

PREDICTED: Manis pentadactyla xin actin-binding repeat-containing protein 2-like (LOC118925732), partial mRNA

NCBI Reference Sequence: XM_036911866.1

[GenBank](#) [FASTA](#)



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	

VISTA DI ALLINEAMENTO VERTICALE AGGANCIATO A QUERY CON LETTERE D'IDENTITA'

Query range 1: 1 to 22

Query	1	ACTCCATTACATGCTATTCAA	22
CP017762.1	862396	TCCATTACATGCTATTCAA	862377
CP027633.1	19889483	CCATTACATGCTATTCAA	19889465
LT972194.1	21776087	CTCCATTACATGCTATTC	21776105
OC164279.1	271	CCATTACATGCTATTCAA	289
HE601627.2	7312442	CTCCATTACATGCTATTC	7312460
AP022721.1	18314661	CTCCATTACATGCTATT	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	ACTCCATTACATGCTAT	39954936
XR_003327131.1	1995	ACTCCATTACATGCTAT	1978
XR_003327129.1	2368	ACTCCATTACATGCTAT	2351
XR_003327128.1	2648	ACTCCATTACATGCTAT	2631
CP025675.1	2093443	CTCCATTACATGCTATT	2093460
CP025662.1	8191839	CTCCATTACATGCTATT	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

<u>CP045894.1</u>	21543532	ACTCTCATTACATGCTATTCA	21543512
<u>CP045894.1</u>	21568810	ACTCTCATTACATGCTATTCA	21568830
<u>CP045894.1</u>	21573646	ACTCTCATTACATGCTATTCA	21573666
<u>XM 031403826.1</u>	461	CCCATTACATGCTATTC	445
<u>XM 031412472.1</u>	211	CCCATTACATGCTATTC	195
<u>CP032575.1</u>	40054738	ACTCCCATTACAAGCTATT	40054756
<u>CP032575.1</u>	59502103	CCCATTACATGCTAGTCAA	59502121
<u>CP032575.1</u>	59985713	CATTACATGCTATTCAA	59985729
<u>LR697107.1</u>	14547457	ACTCCCATTACATGCTA	14547473
<u>LR697107.1</u>	22114195	TTACATGCTATTCAA	22114181
<u>LR697107.1</u>	24648831	CCATTACATGCTATT	24648817
<u>CP035805.1</u>	46204447	CCCATTACATGCTATTC	46204431
<u>CP035805.1</u>	55256433	CTCCCATTACATGCT	55256447
<u>CP035805.1</u>	55258099	CTCCCATTACATGCT	55258113
<u>CP035805.1</u>	55259851	CTCCCATTACATGCT	55259865
<u>LR633958.1</u>	28541402	TCCCATTACATGCTATT	28541386
<u>LR633958.1</u>	5320960	TCCCATTACATGCTA	5320974
<u>LR633958.1</u>	19337100	ATTACATGCTATTCA	19337114
<u>LR633957.1</u>	9359176	TCCCATTACATGCTATT	9359160
<u>LR633944.1</u>	328293	TCCCATTACATGCTATT	328277
<u>LR633944.1</u>	329012	TCCCATTACATGCTATT	328996
<u>LR633944.1</u>	27927790	TTACATGCTATTCAA	27927776
<u>LR606182.1</u>	36877406	TCCCATTACATGCTATT	36877422
<u>LR606182.1</u>	18920084	CCATTACATGCTATTC	18920099
<u>LR606182.1</u>	63221325	TTACATGCTATTCAA	63221339
<u>LR597457.1</u>	5998417	CATTACATGCTATTCAA	5998401
<u>LR597457.1</u>	22387956	TTACATGCTATTCAA	22387942
<u>LR597457.1</u>	29432160	TTACATGCTATTCAA	29432146
<u>CP040924.1</u>	2843564	TCCCATTACATGCTATT	2843580
<u>CP032557.1</u>	20366047	ACTCCCATTACAAGCTATT	20366029
<u>CP032557.1</u>	490197	CATTACATGCTATTCAA	490181
<u>CP032557.1</u>	19549316	CATTACATGCTATTC	19549302
<u>LR594556.1</u>	25948892	TCCCATTGCATGCTATTCA	25948874
<u>LR594556.1</u>	34441629	TCCCATTACATGCTATT	34441613
<u>LR594589.1</u>	503296	CATTACATGCTATTCAA	503312
<u>LR594589.1</u>	4281444	TTACATGCTATTCAA	4281430
<u>LR594589.1</u>	12950896	CATTACATGCTATTC	12950910
<u>CP033012.1</u>	266895	CATTACATGCTATTCAA	266879
<u>XM 028731005.1</u>	2961	CTCCCATTACATGCTAT	2945
<u>LR584444.1</u>	331044	CATTACATGCTATTCAA	331028
<u>LR584420.1</u>	38651378	CATTACATGCTATTCAA	38651394
<u>LR584403.1</u>	49326681	TCCCATTACATTCTATTCAA	49326700
<u>LR584403.1</u>	39254652	ACTCCCATTACATGCTA	39254636
<u>LR584403.1</u>	26460935	CCATTACATGCTATT	26460949
<u>LR584403.1</u>	45730826	CCATTACATGCTATT	45730840
<u>LR584403.1</u>	48250383	CCATTACATGCTATT	48250397
<u>LR584403.1</u>	60477111	CCATTACATGCTATT	60477097
<u>XM 028449733.1</u>	16	TCCCATTACATGCTATT	32
<u>XM 028449732.1</u>	16	TCCCATTACATGCTATT	32
<u>LR535852.1</u>	7366302	TCCCATTACATGCTATT	7366286
<u>LR132035.1</u>	5782616	CATTACATGCTATTCAA	5782600
<u>LR132035.1</u>	424592	CATTACATGCTATTC	424606
<u>LR132000.1</u>	17549045	ACTCCCATTACACGCTATTC	17549026
<u>LR132000.1</u>	8021281	TCCCATTACATGCTATT	8021297
<u>LR131993.1</u>	10909371	CTCCCATTACATGCTTTTCAA	10909351
<u>LR131993.1</u>	33793925	CATTACATGCTATTC	33793911
<u>AP019308.1</u>	5689581	CCCATTACATGCTATTC	5689565
<u>XM 026092999.1</u>	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

PRODOTTI DEGLI ALLINEAMENTI SIGNIFICATIVI (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](#)

SEQUENZA N° 3/3 (T=2; C=0; G=3; A=1)

Query: **ACTCCATTACATGCTATTCAA**

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

SEQUENZA N° 3/4

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

SEQUENZA N° 3/4 (T=2; C=0; G=3; A=1)

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° 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
Sequenza 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 willdenowii 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	goldfish	34.2	34.2	77%	35	100.00%	4417

	(LOC113116706), ncRNA								
CP034970.1	Vibrio chagasii strain ECSMB14107 chromosome 1, complete sequence PREDICTED: Morus notabilis DNA-directed RNA polymerases II, IV and V subunit 11 (LOC21387719), mRNA	Vibrio chagasii	34.2	34.2	95%	35	95.24%	3563604	
XM_010099710.2	PREDICTED: Seriola lalandi dorsalis uncharacterized LOC111654400	Morus notabilis	34.2	34.2	77%	35	100.00%	777	
XR_002760741.1	(LOC111654400), ncRNA PREDICTED: Xenopus laevis sterile alpha motif domain-containing protein 9-like (LOC108708729), mRNA	Seriola lalandi dorsalis	34.2	34.2	77%	35	100.00%	1769	
XM_018247752.1	Comamonas aquatica strain CJG genome	African clawed frog	34.2	34.2	77%	35	100.00%	5335	
CP016603.1	Olsenella sp. oral taxon 807	Comamonas aquatica	34.2	34.2	77%	35	100.00%	3764434	
CP012069.2	strain F0089, complete genome PREDICTED: Sebastes umbrosus saccin-like (LOC119500421), mRNA	Olsenella sp. oral taxon 807	34.2	34.2	77%	35	100.00%	3213585	
XM_037790127.1	Macrobrachium nipponense isolate FS-2020 chromosome 14	honeycomb rockfish	34.2	34.2	77%	35	100.00%	9521	
CP062014.1	Spirochaeta africana DSM 8902, complete genome	Macrobrachium nipponense	34.2	34.2	77%	35	100.00%	99588160	
CP003282.1	Populus trichocarpa clone ACSB2229-H14, complete sequence	Spirochaeta africana DSM 8902	34.2	34.2	77%	35	100.00%	3285855	
AC211727.1	Drosophila C virus strain EB, complete genome	black cottonwood	34.2	34.2	77%	35	100.00%	38255	
AF014388.1	Phallusia mammillata mRNA for superkiller viralicidic activity 2-like 2 (Skiv2l2 gene)	Drosophila C virus	34.2	34.2	95%	35	95.24%	9264	
LR790227.1	Sphingobacterium lactis strain DSM 22361 chromosome, complete genome	Phallusia mammillata	34.2	34.2	77%	35	100.00%	3384	
CP049246.1	Candidatus Syntrophocurvum alkaliphilum isolate B(2M) chromosome, complete genome	Sphingobacterium lactis Candidatus Syntrophocurvum alkaliphilum	34.2	34.2	95%	35	95.24%	4030188	
CP046457.1	Cydia pomonella cytochrome P450 4V2 mRNA, complete cds	Candidatus Syntrophocurvum alkaliphilum	34.2	34.2	77%	35	100.00%	2360781	
MH138257.1	Spirosoma rigui strain KCTC 12531 chromosome, complete genome	codling moth	34.2	34.2	77%	35	100.00%	999	
CP020105.1 1/2 2/2 3/3 8/1	Corynebacterium mustelae strain DSM 45274, complete genome	Spirosoma rigui	34.2	34.2	77%	35	100.00%	5828404	
CP011542.1	PREDICTED: Ceratosolen solmsi marchali uncharacterized LOC105360997	Corynebacterium mustelae	34.2	34.2	77%	35	100.00%	3391554	
XR_912820.1	(LOC105360997), transcript variant X2, misc_RNA PREDICTED: Ceratosolen solmsi marchali uncharacterized LOC105360997	Ceratosolen solmsi marchali	34.2	34.2	77%	35	100.00%	1433	
XM_011498069.1	(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	Canis lupus familiaris breed Labrador retriever chromosome	dog	32.2	32.2	72%	138	100.00%	62857718	

1/2 5/1	2/2 10/1	3/3	3/5 3/6	14b									
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 clupeioides 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 coral grouper	32.2	94.6	90%	138	100.00%		36063594	
CP035042.1				Halomonas sp. BC-M4-5	Halomonas sp. BC-M4-5	32.2	32.2	90%	138	95.00%		5015118	

	chromosome, complete genome	5							
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	
LR697112.1 1/2 2/2 3/1 3/3 8/1	Chanos chanos genome assembly, chromosome: 7 PREDICTED: Aedes albopictus progesterin and adiponQ receptor family member 3-like (LOC109432372), transcript variant X3, mRNA	milkfish	32.2	92.7	81%	138	100.00%	49187548	
XM_029861887.1	PREDICTED: Aedes albopictus progesterin and adiponQ receptor family member 3-like (LOC109432372), transcript variant X2, mRNA	Asian tiger mosquito	32.2	32.2	72%	138	100.00%	1991	
XM_029861886.1	PREDICTED: Aedes albopictus progesterin and adiponQ receptor family member 3-like (LOC109432372), transcript variant X1, mRNA	Asian tiger mosquito	32.2	32.2	72%	138	100.00%	1836	
XM_029861885.1	PREDICTED: Labrus bergylta protein THEM6-like (LOC109985015), mRNA	Asian tiger mosquito	32.2	32.2	72%	138	100.00%	1416	
XM_020635225.2	Psychroserpens sp. NJDZ02 chromosome, complete genome	ballan wrasse Psychroserpens sp. NJDZ02	32.2	32.2	72%	138	100.00%	1880	
CP039451.1			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 Dehalococcoides mccartyi strain FL2 chromosome, complete genome	river trout	32.2	32.2	72%	138	100.00%	91488822	
CP038470.1	PREDICTED: Dendronephthya gigantea NACHT, LRR and PYD domains-containing protein 12-like (LOC114522204), mRNA	Dehalococcoides mccartyi	32.2	32.2	72%	138	100.00%	1422358	
XM_028542838.1	Corylus avellana genome assembly, chromosome: ca4	Dendronephthya gigantea	32.2	32.2	72%	138	100.00%	3953	
LR899426.1	Shewanella sp. YLB-08 chromosome, complete genome	Corylus avellana	32.2	32.2	72%	138	100.00%	36845065	
CP045503.1	PREDICTED: Rhipicephalus microplus uncharacterized LOC119173993 (LOC119173993), mRNA	Shewanella sp. YLB-08	32.2	32.2	72%	138	100.00%	5785459	
XM_037424772.1	Colletotrichum aenigma Cytochrome P450 4F4 (Cyp4f4), partial mRNA	southern cattle tick	32.2	32.2	72%	138	100.00%	9238	
XM_037323409.1	Colletotrichum aenigma (S)-mandelate dehydrogenase (CGCA056_v004203), partial mRNA	Colletotrichum aenigma	32.2	32.2	72%	138	100.00%	1659	
XM_037319458.1	Colletotrichum siamense Cytochrome P450 4F8 (CYP4F8), partial mRNA	Colletotrichum aenigma	32.2	32.2	72%	138	100.00%	2187	
XM_036640383.1	Fusarium subglutinans fad binding domain protein (FSUBG_7736), partial mRNA	Colletotrichum siamense	32.2	32.2	72%	138	100.00%	1659	
XM_036686071.1	Halomonas sp. MCCC 1A13316 chromosome, complete genome	Fusarium subglutinans	32.2	32.2	90%	138	95.00%	1515	
CP053382.1	Ktedonobacteria bacterium brp13 DNA, complete genome	Halomonas sp. MCCC 1A13316	32.2	32.2	90%	138	95.00%	4490931	
AP024096.1	Fusarium oxysporum f. sp. koae isolate 44 chromosome 2	Ktedonobacteria bacterium brp13	32.2	32.2	72%	138	100.00%	8879929	
CP052898.1	Rhodococcus rhodochrous strain BX2 chromosome, complete genome	Fusarium oxysporum f. sp. koae	32.2	32.2	72%	138	100.00%	4660957	
CP027557.1	PREDICTED: Anopheles stephensi aminopeptidase N (LOC118507521), mRNA	Rhodococcus rhodochrous	32.2	32.2	90%	138	95.00%	5459556	
XM_036046082.1	PREDICTED: Zea mays B3 domain-containing protein IDEF1 (LOC103628923), transcript variant X4, 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 X3, mRNA	Zea mays	32.2	32.2	90%	138	95.00%	1726	
XM_020539753.3		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

VISTA DI ALLINEAMENTO VERTICALE AGGANCIATO A QUERY CON LETTERE D'IDENTITA'

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

<u>CP050623.1</u>	10487775	CCATTACATGCGCAAC	10487760
<u>AP022695.1</u>	23257140	CCCATTACATGCGCAA	23257125
<u>AP022690.1</u>	15029198	ACCCATTACATGCGCA	15029183
<u>CP032675.1</u>	1712573	CACCCATTACATGCTCAACA	1712554
<u>XM 031178426.1</u>	2061	CCCATTACATGCGCAA	2046
<u>CP032375.1</u>	5326937	ATTACATGCGCAACAA	5326922
<u>CP032375.1</u>	17112705	TTACATGCGCAACAA	17112691
<u>CP032372.1</u>	5318831	ATTACATGCGCAACAA	5318816
<u>CP032372.1</u>	5328027	ATTACATGCGCAACAA	5328012
<u>CP032372.1</u>	17083568	TTACATGCGCAACAA	17083554
<u>CP032369.1</u>	5317970	ATTACATGCGCAACAA	5317955
<u>CP032369.1</u>	5327176	ATTACATGCGCAACAA	5327161
<u>CP032369.1</u>	17083083	TTACATGCGCAACAA	17083069
<u>LR699094.1</u>	11771640	CACCCATTACATGCGC	11771625
<u>CP042389.1</u>	772295	ATTACATGCGCAACAA	772280
<u>LR537136.1</u>	25217121	CATTACATGCGCAACA	25217106
<u>CP035801.1</u>	56629303	CACCCATTACATGCGC	56629288
<u>CP035801.1</u>	62158347	ACACCCATTACATGC	62158333
<u>LR607350.1</u>	737908	ATTACATGCGCAACAA	737893
<u>CP027858.1</u>	6473139	CCATTACATGCGCAAC	6473124
<u>XM 021039339.2</u>	621	ATTACATGCGCAACAA	606
<u>LR584401.1</u>	61207559	ATTACATGCGCAACAA	61207544
<u>LR535831.1</u>	12835665	ATTACATGCGCAACAA	12835650
<u>LR890312.1</u>	2041547	ATTACATGCGCAACAA	2041532
<u>CP052886.1</u>	833445	CCCATTACATGCGCAA	833430
<u>LR880646.1</u>	34989247	CATTACATGCGCAACA	34989232
<u>LR880646.1</u>	29182986	ACACCCATTACATGC	29183000
<u>LR877148.1</u>	262317	ATTACATGCGCAACAA	262302
<u>CP057330.1</u>	2169365	ATTACATGCGCAACAA	2169350
<u>CP056567.1</u>	2180977	ATTACATGCGCAACAA	2180962
<u>CP045504.1</u>	978787	CACCCATTACATGCGC	978772
<u>LR812606.1</u>	13576925	ATTACATGCGCAACAA	13576910
<u>LR812581.1</u>	11898295	ATTACATGCGCAACAA	11898280
<u>LR812556.1</u>	14423538	ATTACATGCGCAACAA	14423523
<u>CP053263.1</u>	1029999	CCCATTACATGCGCAA	1030014
<u>LR792560.1</u>	29256488	CATTACATGCGCAACA	29256503
<u>LR778311.1</u>	48051809	ATTACATGCGCAACAA	48051824
<u>AP022706.1</u>	33783454	ACCCATTACATGCGCA	33783469
<u>AP022706.1</u>	33804113	ACCCATTACATGCGCA	33804128
<u>AP022706.1</u>	5021490	TTACATGCGCAACAA	5021476
<u>CP035042.1</u>	3998223	CACCCAGTACATGCGCAACA	3998242
<u>CP045427.1</u>	3331462	ATTACATGCGCAACAA	3331477
<u>CP032580.1</u>	40222466	CATTACATGCGCAACA	40222481
<u>LR697112.1</u>	7022426	CCATTACATGCGCAAC	7022441
<u>LR697112.1</u>	38998851	TTACATGCGCAACAA	38998837
<u>LR697112.1</u>	39042207	TTACATGCGCAACAA	39042193
<u>XM 029861887.1</u>	935	CCCATTACATGCGCAA	950
<u>XM 029861886.1</u>	360	CCCATTACATGCGCAA	375
<u>XM 029861885.1</u>	360	CCCATTACATGCGCAA	375
<u>XM 020635225.2</u>	338	CCATTACATGCGCAAC	353
<u>CP039451.1</u>	1153313	ACCCATTTTCATGCGCAACAA	1153332
<u>LR584428.1</u>	69879931	ACCCATTACATGCGCA	69879946
<u>CP038470.1</u>	1202561	ACCCATTACATGCGCA	1202576
<u>XM 028542838.1</u>	617	ACACCCATTACATGCG	632
<u>LR899426.1</u>	3011913	ACCCATTACATGCGCA	3011928
<u>CP045503.1</u>	2956236	ATTACATGCGCAACAA	2956251
<u>XM 037424772.1</u>	7105	ATTACATGCGCAACAA	7120
<u>XM 037323409.1</u>	461	ATTACATGCGCAACAA	476

XM_037319458.1	748	ATTACATGCGCAACAA	763
XM_036640383.1	461	ATTACATGCGCAACAA	476
XM_036686071.1	1106	ACACCCATTTTCATGCGCAAC	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

PRODOTTI DEGLI ALLINEAMENTI SIGNIFICATIVI (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](#)
- Progesterin 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](#)

SEQUENZA N° 3/5

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

SEQUENZA N° 3/5 (T=2; C=0; G=3; A=1)

Sequenza Originale	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
Sequenza 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: **TCTCCATTACATGAGATTCAA**

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	Arabidopsis thaliana genome assembly, chromosome: 1	gray rockcross	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 coralgroup	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 PREDICTED: Nylanderia fulva cytochrome c1-2, heme protein, mitochondrial-like (LOC114941085), transcript variant X2, mRNA	pinecone soldierfish Nylanderia fulva	34.2	64.4	100%	35	95.24%	47119472 1393
XM_029315943.1	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
XM_029315942.1	Gossypioides kirkii chromosome KI_10	Nylanderia fulva	34.2	34.2	77%	35	100.00%	1979
CP032252.1	Gossypioides kirkii chromosome KI_09	Gossypioides kirkii	34.2	223	81%	35	100.00%	47257441
CP032251.1	Gossypioides kirkii chromosome KI_01	Gossypioides kirkii	34.2	64.4	77%	35	100.00%	38743931
CP032244.1	Mastacembelus armatus genome assembly, chromosome: 9	Gossypioides kirkii	34.2	94.6	86%	35	100.00%	38174113
LR535841.1	Anabas testudineus genome assembly, chromosome: 4	zig-zag eel	34.2	64.4	100%	35	100.00%	25822229
LR132043.1	Hyperionvirus sp. clone Hyperionvirus_4 genomic sequence	climbing perch Hyperionvirus sp.	34.2	66.4	86%	35	100.00%	26090302 145918
MK072386.1	PREDICTED: Papaver somniferum 1-aminocyclopropane-1-carboxylate oxidase homolog 1-like (LOC113324819), mRNA	opium poppy	34.2	34.2	77%	35	100.00%	1439
XM_026573109.1	Paenisporosarcina sp. K2R23-3 chromosome, complete genome	Paenisporosarcina sp. K2R23-3	34.2	34.2	77%	35	100.00%	2537131
CP032418.1	Ipomoea triloba cultivar NCNSP0323 chromosome 13	trilobed morning glory	34.2	64.4	77%	35	100.00%	32066333
CP025672.1	Ipomoea trifida cultivar NCNSP0306 chromosome 13	Ipomoea trifida	34.2	34.2	77%	35	100.00%	22914374
CP025656.1	Haemonchus contortus, ISE/inbred ISE, WGS project CAVP01000000 data, chromosome: X	barber pole worm	34.2	34.2	77%	35	100.00%	46012677
LS997567.1	Lateolabrax maculatus linkage group 7 sequence	spotted sea bass	34.2	64.4	77%	35	100.00%	27713553
CP032604.1	Lateolabrax maculatus chromosome Lm20	spotted sea bass	34.2	64.4	77%	35	100.00%	21152183
CP027281.1	Cryptococcus neoformans var. neoformans JEC21 hypothetical protein (CND04685), mRNA	Cryptococcus neoformans var. neoformans JEC21	34.2	34.2	77%	35	100.00%	2087
XM_024658498.1	PREDICTED: Morus notabilis serine/threonine-protein kinase BLUS1 (LOC21410377), mRNA	Morus notabilis	34.2	34.2	77%	35	100.00%	2094
XM_024168143.1	PREDICTED: Cyanistes caeruleus uncharacterized LOC111928373	blue tit	34.2	34.2	77%	35	100.00%	1664
XR_002864067.1	(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 PREDICTED: Athalia rosae uncharacterized LOC105689598	long-grained rice	34.2	34.2	77%	35	100.00%	35849732
XM_012406738.2	(LOC105689598), mRNA	coleseed sawfly	34.2	34.2	77%	35	100.00%	1824

	PREDICTED: Lates calcarifer microtubule-associated serine/threonine-protein kinase 2-like (LOC108883545), mRNA	barramundi perch	34.2	34.2	77%	35	100.00%	10871
XM_018676841.1	Buchnera aphidicola (Schlechtendalia chinensis) strain SC, complete genome	Buchnera aphidicola (Schlechtendalia chinensis)	34.2	34.2	77%	35	100.00%	607835
CP011299.1	Colchicum autumnale chloroplast, complete genome	Colchicum autumnale	34.2	34.2	77%	35	100.00%	156462
KP125337.1	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
AP015034.1	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
AP015039.1	Sardinella longiceps isolate CMFRIMBTDS 419 control region, partial sequence; mitochondrial	Indian oil sardine	34.2	34.2	77%	35	100.00%	1110
KJ888251.1	Sardinella longiceps isolate CMFRIMBTDS 325 control region, partial sequence; mitochondrial	Indian oil sardine	34.2	34.2	77%	35	100.00%	1076
KJ888227.1	Oryza sativa Japonica Group DNA, chromosome 12, cultivar: Nipponbare, complete sequence	Japanese rice	34.2	64.4	100%	35	100.00%	27531856
AP014968.1	Oryza sativa Japonica Group DNA, chromosome 4, cultivar: Nipponbare, complete sequence	Japanese rice	34.2	34.2	77%	35	100.00%	35502694
AP014960.1	Oryza sativa Japonica Group DNA, chromosome 2, cultivar: Nipponbare, complete sequence	Japanese rice	34.2	34.2	77%	35	100.00%	35937250
AP014958.1	Oryza sativa Indica Group cultivar RP Bio-226 chromosome 2 sequence	long-grained rice	34.2	34.2	77%	35	100.00%	36385228
CP012610.1	Apteryx australis mantelli genome assembly AptMant0, scaffold scaffold11160	Apteryx australis mantelli	34.2	34.2	77%	35	100.00%	203429
LK075972.1	Cucumis melo genomic chromosome, chr_4	muskmelon	34.2	64.4	90%	35	100.00%	33123230
LN713258.1	Cucumis melo genomic chromosome, chr_1	muskmelon	34.2	34.2	77%	35	100.00%	35383099
LN713255.1	Cucumis melo genomic scaffold, anchoredscaffold00011	muskmelon	34.2	34.2	77%	35	100.00%	6909323
LN681841.1	Cucumis melo genomic scaffold, anchoredscaffold00046	muskmelon	34.2	34.2	77%	35	100.00%	3261579
LN681802.1	Cyprinus carpio genome assembly common carp genome, scaffold: LG47, chromosome: 47	common carp	34.2	64.4	77%	35	100.00%	16759857
LN590681.1								
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	Cryptococcus neoformans var.	34.2	34.2	77%	35	100.00%	1783081

	4 sequence	neoformans JEC21						
AC125233.4	Mus musculus BAC clone RP23-9D24 from chromosome 5, complete sequence	house mouse	34.2	34.2	77%	35	100.00%	212792
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

<u>AP022632.1</u>	81609444	CTCCCATTACATGAGAT	81609460
<u>AP022632.1</u>	1185563	CTCCCATTACATGAG	1185549
<u>AP022704.1</u>	31638031	CCATTACATGAGATTCA	31638047
<u>AP022704.1</u>	13523044	ATTACATGAGATTCA	13523058
<u>XM 002936994.4</u>	1647	TCTCCCATTACATGAGA	1631
<u>LR722981.1</u>	16834644	CATTACATGAGATTCAA	16834660
<u>LR722981.1</u>	89069024	CATTACATGAGATTCA	89069009
<u>LR722981.1</u>	89112696	CATTACATGAGATTCA	89112681
<u>CP042455.1</u>	1241841	CCATTACATGAGATTCA	1241825
<u>XM 030198566.1</u>	996	TCTCCCATTACATGAGA	980
<u>XM 030198557.1</u>	996	TCTCCCATTACATGAGA	980
<u>XM 030198551.1</u>	1122	TCTCCCATTACATGAGA	1106
<u>XM 030198543.1</u>	1122	TCTCCCATTACATGAGA	1106
<u>XM 030198536.1</u>	1122	TCTCCCATTACATGAGA	1106
<u>XM 030198529.1</u>	1122	TCTCCCATTACATGAGA	1106
<u>XM 030198521.1</u>	1122	TCTCCCATTACATGAGA	1106
<u>LR537130.1</u>	21695109	CATTACATGAGATTCAA	21695093
<u>LR537130.1</u>	29966081	CATTACATGAGATTCA	29966066
<u>CP035801.1</u>	33249112	TCTCCCATTATATGAGATTCA	33249132
<u>CP035801.1</u>	6199412	TCCCATTACATGAGA	6199398
<u>CP035800.1</u>	186911	CATTACATGAGATTCAA	186927
<u>LR606198.1</u>	25126242	TCTCCCATTACATGAGA	25126258
<u>LR606198.1</u>	27821319	ATTACATGAGATTCAA	27821304
<u>LR597552.1</u>	23845435	TCTCCCAGTACATGAGATTCA	23845455
<u>LR597552.1</u>	16379503	TTACATGAGATTCAA	16379489
<u>XM 029315943.1</u>	1108	CATTACATGAGATTCAA	1124
<u>XM 029315942.1</u>	1107	CATTACATGAGATTCAA	1123
<u>CP032252.1</u>	6434823	CCATTACATGAGATTCA	6434807
<u>CP032252.1</u>	44215530	CCATTACATGAGATTCA	44215514
<u>CP032252.1</u>	26950127	CATTACATGAGATTCA	26950142
<u>CP032252.1</u>	33839702	CCATTACATGAGATTTC	33839717
<u>CP032252.1</u>	34237892	CATTACATGAGATTTC	34237906
<u>CP032252.1</u>	37887277	CATTACATGAGATTTC	37887263
<u>CP032252.1</u>	44549642	TTACATGAGATTCAA	44549656
<u>CP032251.1</u>	35471183	CCATTACATGAGATTCA	35471199
<u>CP032251.1</u>	37264375	CATTACATGAGATTTC	37264361
<u>CP032244.1</u>	18935716	TCCCATTACATAAGATTCA	18935734
<u>CP032244.1</u>	19967470	CCATTACATGAGATTCA	19967486
<u>CP032244.1</u>	4230653	CATTACATGAGATTTC	4230639
<u>LR535841.1</u>	1112990	CATTACATGAGATTCAA	1112974
<u>LR535841.1</u>	25095307	TCTCCCATTACATGA	25095293
<u>LR132043.1</u>	24477558	CCATTACATGAGATTCA	24477542
<u>LR132043.1</u>	18950349	TCCCATTACATGAGAT	18950364
<u>MK072386.1</u>	18967	TCTCCAATACATGAGATTCA	18947
<u>XM 026573109.1</u>	1343	TCTCCCATTACATGAGA	1327
<u>CP032418.1</u>	1289823	TCTCCCATTACATGAGA	1289839
<u>CP025672.1</u>	2204754	CATTACATGAGATTCAA	2204770
<u>CP025672.1</u>	18960837	TTACATGAGATTCAA	18960851
<u>CP025656.1</u>	1330796	CATTACATGAGATTCAA	1330812
<u>LS997567.1</u>	259351	CATTACATGAGATTCAA	259367
<u>CP032604.1</u>	21005603	CATTACATGAGATTCAA	21005619
<u>CP032604.1</u>	21347900	TTACATGAGATTCAA	21347886
<u>CP027281.1</u>	8637131	CATTACATGAGATTCAA	8637115
<u>CP027281.1</u>	8191845	ATTACATGAGATTCA	8191831
<u>XM 024658498.1</u>	633	CTCCCATTACATGAGAT	617
<u>XM 024168143.1</u>	1717	CATTACATGAGATTCAA	1701
<u>XR 002864067.1</u>	1518	CATTACATGAGATTCAA	1534
<u>LT972181.1</u>	7109283	TCCCATTACATGAGATT	7109299

<u>CP023113.2</u>	20058368	CTCCCATTACATGAAATTCAA	20058348
<u>CP023113.2</u>	19208163	TTACATGAGATTCAA	19208177
<u>XM 022430592.1</u>	1510	CATTACATGAGATTCAA	1494
<u>CP021239.1</u>	700004	CCATTACATGAGATTCA	699988
<u>CP021239.1</u>	965864	CTCCCATTACATGAGAT	965880
<u>CP018160.1</u>	2265361	TCTCCCATTACATGAGA	2265377
<u>XM 012406738.2</u>	1057	TCTCCCATTACATGAGA	1041
<u>XM 018676841.1</u>	10516	CCATTACATGAGATTCA	10532
<u>CP011299.1</u>	555657	CATTACATGAGATTCAA	555641
<u>KP125337.1</u>	81891	TCCCATTACATGAGATT	81875
<u>AP015034.1</u>	54524717	CATTACATGAGATTCAA	54524733
<u>AP015034.1</u>	66723276	CATTACATGAGATTCA	66723291
<u>AP015039.1</u>	37843921	TCCCATTACATGAGATT	37843905
<u>KJ888251.1</u>	748	CCATTACATGAGATTCA	732
<u>KJ888227.1</u>	742	CCATTACATGAGATTCA	726
<u>AP014968.1</u>	21053515	TCTCCCATTACATGAGA	21053531
<u>AP014968.1</u>	8709153	TTACATGAGATTCAA	8709139
<u>AP014960.1</u>	2251499	TCTCCCATTACATGAGA	2251483
<u>AP014958.1</u>	9627530	TCTCCCATTACATGAGA	9627546
<u>CP012610.1</u>	9941563	TCTCCCATTACATGAGA	9941579
<u>LK075972.1</u>	10878	TCTCCCATTACATGAGA	10862
<u>LN713258.1</u>	29807408	TCCCATTACATGAGATT	29807424
<u>LN713258.1</u>	1409995	TTACATGAGATTCAA	1410009
<u>LN713255.1</u>	26026003	CATTACATGAGATTCAA	26025987
<u>LN681841.1</u>	3315823	TCCCATTACATGAGATT	3315807
<u>LN681802.1</u>	2219811	CATTACATGAGATTCAA	2219795
<u>LN590681.1</u>	3767544	CATTACATGAGATTCAA	3767528
<u>LN590681.1</u>	16263666	TTACATGAGATTCAA	16263652
<u>HG975451.1</u>	27599434	CATTACATGAGATTCAA	27599450
<u>HG975451.1</u>	63830996	TTACATGAGATTCAA	63830982
<u>HG975451.1</u>	66010204	CATTACATGAGATTC	66010218
<u>HG975451.1</u>	68655790	CATTACATGAGATTC	68655804
<u>OA566334.1</u>	71577	TCCCATTACATGAGATT	71593
<u>CP062021.1</u>	15212325	CTCCCATTACATGAGAT	15212309
<u>CP062051.1</u>	9417621	CATTACATGAGATTCAA	9417605
<u>CP062051.1</u>	40451573	TTACATGAGATTCAA	40451559
<u>NC 049026.1</u>	82718	TCCCATTACATGAGATT	82702
<u>NC 049025.1</u>	81902	TCCCATTACATGAGATT	81886
<u>NC 049014.1</u>	80875	TCCCATTACATGAGATT	80859
<u>LS423614.1</u>	65154382	TCCCATTACATGAGATT	65154398
<u>LS423614.1</u>	10082855	TCCCATTACATGAGA	10082841
<u>LS423614.1</u>	12504604	CATTACATGAGATTC	12504590
<u>LS423614.1</u>	41585139	TTACATGAGATTCAA	41585153
<u>LS423614.1</u>	45137039	TTACATGAGATTCAA	45137053
<u>LS423614.1</u>	58730326	TTACATGAGATTCAA	58730312
<u>LR812533.1</u>	34706224	CATTACATGAGATTCAA	34706208
<u>LR812533.1</u>	45428608	ATTACATGAGATTCAA	45428623
<u>LR812533.1</u>	17485149	ATTACATGAGATTCA	17485135
<u>XM 771023.1</u>	460	CTCCCATTACATGAGAT	444
<u>CU075459.1</u>	1524	TCTCCCATTACATGAGA	1508
<u>AC102166.13</u>	98747	CATTACATGAGATTCAA	98763
<u>AE017344.1</u>	1290155	CTCCCATTACATGAGAT	1290171
<u>AC125233.4</u>	153179	CCATTACATGAGATTCA	153195
<u>CR382121.1</u>	700120	CCATTACATGAGATTCA	700104
<u>CR382121.1</u>	965985	CTCCCATTACATGAGAT	966001
<u>AC158950.3</u>	54888	CATTACATGAGATTCAA	54872

PRODOTTI DEGLI ALLINEAMENTI SIGNIFICATIVI (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](#)

SEQUENZA N° 3/5 (T=2; C=0; G=3; A=1)

Query: **TCTCCATTACATGAGATTCAA**

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
.....

SEQUENZA N° 3/6

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

SEQUENZA N° 3/6 (T=2; C=0; G=3; A=1)

Sequenza Originaria

Trend n° 3 (T=2; C=0; G=3; A=1)

Trend n° 3 (I i D d)

Trend n° 3/6

Sequenza n° 3/6 (54,54%)

G	T	G	A	A	A	T	G	G	T	C	A	T	G	T	G	T	G	G	C	G	G
3	2	3	1	1	1	2	3	3	2	0	1	2	3	2	3	2	3	3	0	3	3
-	D	I	D	d	d	I	I	i	D	D	I	I	I	D	I	D	I	i	D	I	i
3	2	3	2	2	2	1	2	2	1	0	1	2	3	0	2	0	3	3	0	1	1
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	Salmo trutta genome assembly, chromosome: 13	river trout	40.1	40.1	90%	0.57	100.00%	91488822
1/2 2/2 3/3 3/4 8/1	Bacillus velezensis strain SYBC H47, complete genome	Bacillus velezensis	38.2	38.2	86%	2.2	100.00%	3884433
CP017747.1	Bacillus amyloliquefaciens strain S499, complete genome	Bacillus amyloliquefaciens	38.2	38.2	86%	2.2	100.00%	3927922
CP014700.1	Bacillus amyloliquefaciens IT-45, complete genome	Bacillus amyloliquefaciens IT-45	38.2	38.2	86%	2.2	100.00%	3928857
CP004065.1	Bacillus velezensis strain DTU001 chromosome, complete genome	Bacillus velezensis	38.2	38.2	86%	2.2	100.00%	3927210
CP035533.1	Bacillus velezensis strain KMU01 chromosome, complete genome	Bacillus velezensis	38.2	38.2	86%	2.2	100.00%	3932437
CP063768.1	Bacillus velezensis strain B268 chromosome, complete genome	Bacillus velezensis	38.2	38.2	86%	2.2	100.00%	3908940
CP053764.1	Bacillus velezensis strain LPL061 chromosome, complete genome	Bacillus velezensis	38.2	38.2	86%	2.2	100.00%	3907268
CP042271.1	Bacillus velezensis strain KD1 chromosome, complete genome	Bacillus velezensis	38.2	38.2	86%	2.2	100.00%	3921966
CP014990.2	Bacillus velezensis strain GFP-2 chromosome, complete genome	Bacillus velezensis	38.2	38.2	86%	2.2	100.00%	3975220
CP021011.1	Bacillus velezensis strain GH1-13, complete genome	Bacillus velezensis	38.2	38.2	86%	2.2	100.00%	4071980
CP019040.1	Bacillus amyloliquefaciens strain Y14, complete genome	Bacillus amyloliquefaciens	38.2	38.2	86%	2.2	100.00%	3957164
CP017953.1	Apteryx australis mantelli genome assembly AptMant0, scaffold scaffold152	Apteryx australis mantelli	38.2	38.2	86%	2.2	100.00%	6182939
LK064797.1	Bacillus velezensis strain L-S60 chromosome, complete genome	Bacillus velezensis	38.2	38.2	86%	2.2	100.00%	3903017
CP011278.1	Bacillus velezensis strain L-H15 chromosome, complete genome	Bacillus velezensis	38.2	38.2	86%	2.2	100.00%	3905973
CP010556.1	Bacillus velezensis strain CLA178 chromosome, complete genome	Bacillus velezensis	38.2	38.2	86%	2.2	100.00%	4068479
CP061087.1	Danio rerio genome assembly, chromosome: 12	zebrafish	36.2	66.4	81%	8.9	100.00%	36192843
LR812074.1	Cytophagaceae bacterium isolate ECT2AJA-110-C chromosome	Cytophagaceae bacterium	36.2	36.2	81%	8.9	100.00%	4167162
CP030886.1	Danio rerio genome assembly, chromosome: 12	zebrafish	36.2	36.2	81%	8.9	100.00%	49740644
LR812049.1	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	Sphaeramia orbicularis genome assembly, chromosome: 15	orbiculate cardinalfish	34.2	34.2	77%	35	100.00%	54816899
LR597464.1	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 ATCC4355T 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	Ovis canadensis	34.2	126	95%	35	95.24%	86551329

	43U chromosome 10 sequence	canadensis								
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		
XM_008876883.1	Rhabditophanes_sp_KR3021, scaffold RSKR_scaffold0000001	Aphanomyces invadans hypothetical protein, variant mRNA	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	Canis lupus familiaris breed									
1/2 2/2 3/3 3/4 3/5 5/1 10/1	Labrador retriever chromosome 05a	dog	34.2	64.4	77%	35	100.00%	86926571		
CP050642.1	Canis lupus familiaris breed									
1/2 2/2 3/3 3/4 3/5 5/1 10/1	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 alloenum filamin-C (LOC107042647), mRNA	Diachasma alloenum	34.2	34.2	77%	35	100.00%	9274		
LR132000.1	Gouania willdenowii genome assembly, chromosome: 6	blunt-snouted clingfish	34.2	128	86%	35	100.00%	72451976		
1/2 3/3 3/4	PREDICTED: Acinonyx jubatus intersectin 1 (ITSN1), transcript variant X6, mRNA	cheetah	34.2	34.2	77%	35	100.00%	5435		
XM_027041670.1	Pseudoalteromonas agarivorans strain Hao 2018 chromosome I, complete sequence	Pseudoalteromonas agarivorans	34.2	34.2	95%	35	95.24%	3611742		
CP033065.1	PREDICTED: Solenopsis invicta WD repeat-containing protein on Y chromosome (LOC105207815), mRNA	red fire ant	34.2	34.2	77%	35	100.00%	2976		
XM_026134041.1	Arachis hypogaea cultivar Shitouqi chromosome A08	peanut	34.2	64.4	81%	35	100.00%	51955169		
CP030990.1										
CP027087.1	Bos mutus isolate yakQH1 chromosome 19	wild yak	34.2	34.2	77%	35	100.00%	60756558		
1/2 7/2 10/1	PREDICTED: Felis catus intersectin 1 (ITSN1), transcript variant X6, mRNA	domestic cat	34.2	34.2	77%	35	100.00%	5421		
XM_023238782.1	Pseudoalteromonas atlantica strain ECSMB14104 chromosome, complete genome	Pseudoalteromonas atlantica	34.2	34.2	95%	35	95.24%	3669595		
CP023464.1	Pseudoalteromonas agarivorans DSM 14585 chromosome I, complete sequence	Pseudoalteromonas agarivorans DSM 14585	34.2	34.2	95%	35	95.24%	3701940		
CP011011.1										

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

VISTA DI ALLINEAMENTO VERTICALE AGGANCIATO A QUERY CON LETTERE D'IDENTITA'

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

<u>LR597464.1</u>	3410873	TGTTTATTACATGCT	3410887
<u>LR597464.1</u>	56438488	TGTTTATTACATGCT	56438502
<u>LR597464.1</u>	3473140	GTTTATTACATGCTC	3473126
<u>LR597464.1</u>	9261435	GTGTTTATTACATGC	9261421
<u>CP027523.1</u>	773271	GTGTTTAATACATGCTCGGCA	773291
<u>LR131964.1</u>	2784092	GTGTTTATTACATGCTC	2784108
<u>XM 026419049.1</u>	749	TGTTTATTACATGCTCG	765
<u>CP022985.1</u>	263866	TGTTTATTACATTCTCGGCAA	263886
<u>CP022985.1</u>	1767496	TGTTTATTACATTCTCGGCAA	1767476
<u>CP035704.1</u>	2627575	TTTATTACATGCTCGGC	2627591
<u>LT965928.1</u>	767567	GTGTTTAATACATGCTCGGCA	767587
<u>CP022522.1</u>	1439540	GTTTATTACATGCTCGG	1439556
<u>XM 022836198.1</u>	1412	TGTTGATTACATGCTCGGCAA	1432
<u>XM 022836197.1</u>	1572	TGTTGATTACATGCTCGGCAA	1592
<u>XM 022792557.1</u>	1612	TGTTGATTACATGCTCGGCAA	1632
<u>XM 022792556.1</u>	1811	TGTTGATTACATGCTCGGCAA	1831
<u>XM 015846147.1</u>	4378	TATTACATGCTCGGCAA	4394
<u>CP011895.1</u>	47859170	GTGTTTATTACATGCTAGGCA	47859190
<u>CP011895.1</u>	32793092	GTGTTTATTACATGCT	32793107
<u>CP011895.1</u>	8531593	TGTTTATTACATGCT	8531607
<u>CP011895.1</u>	70986903	TGTTTATTACATGCT	70986917
<u>XM 010757741.1</u>	4378	TATTACATGCTCGGCAA	4394
<u>LK995546.1</u>	4646482	GTGTTTATTACATGCTC	4646498
<u>XM 008876883.1</u>	1892	TTATTACATGCTCGGCA	1908
<u>XM 008876882.1</u>	1957	TTATTACATGCTCGGCA	1973
<u>XM 007408047.1</u>	923	GTGTTTATTACATGCTC	939
<u>FO081946.3</u>	2338	GTGTTTATTACATGCTC	2354
<u>LR812053.1</u>	23405961	TGTTTATTACATGCTCG	23405977
<u>LR812053.1</u>	13835351	TGTTTATTACATGCT	13835365
<u>LR812053.1</u>	13475574	TGTTTATTACATGCT	13475560
<u>LR812053.1</u>	14202541	GTGTTTATTACATGC	14202527
<u>LR812584.1</u>	23158624	TGTTTATTACATGCTCG	23158640
<u>LR812584.1</u>	13560226	TGTTTATTACATGCT	13560240
<u>LR812584.1</u>	13220015	TGTTTATTACATGCT	13220001
<u>LR812520.1</u>	33859889	TTATTACATGCTCGGCA	33859905
<u>LR812520.1</u>	48573693	TGTTTATTACATGCT	48573707
<u>LR812520.1</u>	11621984	TGTTTATTACATGCT	11621970
<u>LR812520.1</u>	35856217	GTGTTTATTACATGC	35856203
<u>CP000612.1</u>	864290	GTGTTTATTACATGCTC	864306
<u>CR854942.16</u>	198069	TGTTTATTACATGCTCG	198085
<u>CP050595.1</u>	51054656	GTGTTTATTACATGCTC	51054640
<u>CP050595.1</u>	43698638	GTGTTTATTACATGC	43698652
<u>CP050642.1</u>	51047486	GTGTTTATTACATGCTC	51047470
<u>CP050642.1</u>	43691978	GTGTTTATTACATGC	43691992
<u>XM 030328380.2</u>	4435	TTATTACATGCTCGGCA	4419
<u>XM 030328379.2</u>	4450	TTATTACATGCTCGGCA	4434
<u>XM 030328378.2</u>	4648	TTATTACATGCTCGGCA	4632
<u>XM 030328376.2</u>	4663	TTATTACATGCTCGGCA	4647
<u>CP040558.1</u>	2730007	GTGTTTAATACATGCTCGGCA	2729987
<u>XM 015263782.2</u>	4941	TGTTTATTACATGCTCG	4925
<u>LR132000.1</u>	13471894	GTGTTTATTACATGCTC	13471878
<u>LR132000.1</u>	15682035	GTGTTTATTACATGCT	15682050
<u>LR132000.1</u>	17723161	TTTATTACATGCTCGG	17723146
<u>LR132000.1</u>	39203337	TGTTTATTACATGCT	39203323
<u>XM 027041670.1</u>	4594	TTATTACATGCTCGGCA	4578
<u>CP033065.1</u>	2858632	GTGTTTAATACATGCTCGGCA	2858612
<u>XM 026134041.1</u>	1180	TGTTTATTACATGCTCG	1164
<u>CP030990.1</u>	47809991	GTTTATTACATGCTCGG	47809975

<u>CP030990.1</u>	50122539	TGTTTATTACATGCT	50122553
<u>CP027087.1</u>	16137289	TTTATTACATGCTCGGC	16137273
<u>XM 023238782.1</u>	4580	TTATTACATGCTCGGCA	4564
<u>CP023464.1</u>	2993631	GTGTTTAATACATGCTCGGCA	2993611
<u>CP011011.1</u>	3015947	GTGTTTAATACATGCTCGGCA	3015927
<u>CP011028.1</u>	3054898	GTGTTTAATACATGCTCGGCA	3054878
<u>CP015399.2</u>	4853392	GTGTTTATTACATGCTC	4853376
<u>LT634572.1</u>	1541167	GTGTTTATTACATGCTC	1541151
<u>XM 015541209.1</u>	4358	TTATTACATGCTCGGCA	4342
<u>XM 012986555.1</u>	1456	TGTTTATTACATGCTCG	1440
<u>XM 012986554.1</u>	1574	TGTTTATTACATGCTCG	1558
<u>XM 012699782.1</u>	69	TTTATTACATGCTCGGC	53
<u>LR812609.1</u>	22130831	TGTTTATTACATGCTCG	22130815
<u>LR812609.1</u>	14729804	TGTTTATTACATGCT	14729818
<u>LR812609.1</u>	14396777	TGTTTATTACATGCT	14396763
<u>LR812609.1</u>	15096455	GTGTTTATTACATGC	15096441
<u>FP074886.7</u>	124177	GTGTTTATTACATGCTC	124161
<u>FP929039.1</u>	1812939	TTTATTACATGCTCGGC	1812923
<u>AC149883.2</u>	21346	GTGTTTATTACATGCTC	21330
<u>AY016023.1</u>	163893	GTGTTTATTACATGCTC	163877
<u>AB006531.1</u>	5925	GTGTTTATTACATGCTC	5909
<u>OA566189.1</u>	69235	TGTTTATTACTTGCTCGGCA	69254
<u>OA564290.1</u>	68076	TTATTACATGCTCGGC	68091
<u>CP064845.1</u>	1321270	TATTACATGCTCGGCA	1321285
<u>CP064846.1</u>	182505	TATTACATGCTCGGCA	182520
<u>OA565377.1</u>	104438	GTGTTTATTACATGCT	104423
<u>CP064904.1</u>	1014034	TTTATTACATGCTCGG	1014019
<u>CP064851.1</u>	27375	TATTACATGCTCGGCA	27360
<u>CP043357.1</u>	27582	TATTACATGCTCGGCA	27567
<u>CP043353.1</u>	28507	TATTACATGCTCGGCA	28492
<u>CP064820.1</u>	2571746	TATTACATGCTCGGCA	2571731

PRODOTTI DEGLI ALLINEAMENTI SIGNIFICATIVI (CDS FEATURE)

- **Citrate synthase 3** [GenBank](#) [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](#)
- **Tricyclic synthase 0e23, chloroplastic-like** [GenBank](#) [GenBank](#)
- **Alpha-catulin-like** [GenBank](#)
- **Nonstructural polyprotein** [GenBank](#)
- **D-serine transporter DsdX** [GenBank](#) [GenBank](#) [GenBank](#) [GenBank](#)

SEQUENZA N° 4/1

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

SEQUENZA N° 4/1 (T=1; C=3; G=0; A=2)

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° 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
Sequenza 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: **TATCCATTACATGATATTATT**

Le basi della **Sequenza n° 4/1** (T=1; C=3; G=0; A=2) sono identiche alle basi della **Sequenza n° 3/1** (T=2; C=0; G=3; A=1). Per i dettagli si rimanda a quanto scritto riguardo la **Sequenza n° 3/1**.

SEQUENZA N° 5/1

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

SEQUENZA N° 5/1 (T=3; C=1; G=2; A=0)

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° 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
Sequenza 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	Canis lupus familiaris breed							
1/2 2/2 3/3 3/4 3/5 3/6 10/1	Labrador retriever chromosome 10a	dog	38.2	68.4	100%	2.2	100.00%	69938001
CP050611.1	Canis lupus familiaris breed							
1/2 2/2 3/3 3/4 3/5 3/6 10/1	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 APyT genome assembly, chromosome: 1	Akkermansia glycaniphila	38.2	70.4	95%	2.2	100.00%	3074078	
LR812118.1	Erithacus rubecula genome assembly, chromosome: 15	European robin	36.2	36.2	81%	8.7	100.00%	19110993	
1/2 3/2	Felis catus Senzu DNA, chromosome: E1, American Shorthair breed	domestic cat	36.2	66.4	86%	8.7	100.00%	65441144	
AP023165.1	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_002189001.3	PREDICTED: Taeniopygia guttata protein Wnt-11b-2 (LOC100226560), transcript variant X1, mRNA	zebra finch	36.2	36.2	81%	8.7	100.00%	2886	
XM_030272468.2	Lachnospiraceae bacterium isolate MGYG-HGUT-01373 genome assembly, chromosome: 1	Lachnospiraceae bacterium	36.2	36.2	81%	8.7	100.00%	2219029	
LR698958.1	Myripristis murdjan genome assembly, chromosome: 18	pinecone soldierfish	36.2	36.2	81%	8.7	100.00%	29228082	
LR597567.1	Salarias fasciatus genome assembly, chromosome: 3	jewelled blenny	36.2	66.4	81%	8.7	100.00%	30411982	
LR597438.1	Echeneis naucrates genome assembly, chromosome: 19	live sharksucker	36.2	36.2	100%	8.7	95.45%	17971980	
LR584060.1	Acidovorax sp. 1608163 chromosome, complete genome	Acidovorax sp. 1608163	36.2	36.2	100%	8.7	95.45%	5045878	
CP033069.1	Streptomyces lydicus strain WYEC 108 chromosome, complete genome	Streptomyces lydicus	36.2	96.6	100%	8.7	100.00%	9125666	
CP029042.1	PREDICTED: Dromaius novaehollandiae pygopus family PHD finger 2 (PYGO2), mRNA	emu	36.2	36.2	100%	8.7	95.45%	1454	
XM_026121011.1	Sphingopyxis sp. FD7 plasmid pSFD01 DNA, complete sequence	Sphingopyxis sp. FD7	36.2	36.2	81%	8.7	100.00%	242612	
AP017899.1	Methylovulum psychrotolerans strain HV10_M2, complete genome	Methylovulum psychrotolerans	36.2	66.4	81%	8.7	100.00%	4923400	
CP022129.1	Sodalis glossinidius str. 'morsitans' isolate B4 genome assembly, chromosome:	Sodalis glossinidius str. 'morsitans'	36.2	66.4	81%	8.7	100.00%	4171874	
LN854557.1	SgGMMB4_chromosome PREDICTED: Calidris pugnax dachsous cadherin-related 2 (DCHS2), mRNA	ruff	36.2	36.2	100%	8.7	95.45%	10705	
XM_014943042.1	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_012646526.1	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	
XM_017700999.2	Komagataeibacter hansenii strain C110 chromosome, complete genome	Komagataeibacter hansenii	36.2	36.2	81%	8.7	100.00%	3530618	
CP062147.1	Escherichia coli strain EC96 chromosome, complete genome	Escherichia coli	36.2	36.2	81%	8.7	100.00%	4461910	
CP060748.1	Mycobacterium gordonae strain 24T chromosome, complete genome	Mycobacterium gordonae	36.2	36.2	81%	8.7	100.00%	6266765	
CP059165.1	Chlamydomonas reinhardtii strain CC425 Mn superoxide dismutase 4 (MSD4) mRNA, complete cds	Chlamydomonas reinhardtii	36.2	36.2	81%	8.7	100.00%	1342	
GU134344.1	Oryza minuta clone OM_Ba0185B12, complete sequence	Oryza minuta	36.2	36.2	81%	8.7	100.00%	188666	
AC236542.1	Oryza minuta clone OM_Ba0059E04, complete sequence	Oryza minuta	36.2	36.2	81%	8.7	100.00%	131828	
AC236527.1	Chlamydomonas reinhardtii superoxide dismutase [Mn] (MSD4), partial mRNA	Chlamydomonas reinhardtii	36.2	36.2	81%	8.7	100.00%	657	
XM_001690884.1	Sodalis glossinidius str. 'morsitans' DNA, complete genome	Sodalis glossinidius str. 'morsitans'	36.2	66.4	81%	8.7	100.00%	4171146	
AP008232.1	Danio rerio genome assembly, chromosome: 16	zebrafish	34.2	96.6	95%	35	100.00%	55350661	
LR812078.1	PREDICTED: Zootoca vivipara glycyl-tRNA synthetase 1 (GARS1), mRNA	common lizard	34.2	34.2	77%	35	100.00%	2327	
XM_035129734.1	Felis catus Senzu DNA, chromosome: D2, American Shorthair breed	domestic cat	34.2	34.2	77%	35	100.00%	90643714	
AP023162.1									

XM_034315246.1	PREDICTED: <i>Osmia lignaria</i> uncharacterized LOC117600181 (LOC117600181), transcript variant X2, mRNA	orchard mason bee	34.2	34.2	77%	35	100.00%	1225
XM_034315245.1	PREDICTED: <i>Osmia lignaria</i> 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	<i>Pseudomonas putida</i> strain JQ581 chromosome, complete genome	<i>Pseudomonas putida</i>	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: <i>Pecten maximus</i> excitatory amino acid transporter-like (LOC117336294), mRNA	<i>Pecten maximus</i>	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	<i>Digitaria exilis</i> annotation	<i>Digitaria exilis</i>	34.2	94.6	77%	35	100.00%	38778756
LR792828.1	<i>Digitaria exilis</i> annotation	<i>Digitaria exilis</i>	34.2	64.4	77%	35	100.00%	40393928
LR792821.1	<i>Digitaria exilis</i> annotation	<i>Digitaria exilis</i>	34.2	64.4	77%	35	100.00%	35142973
LR792820.1	<i>Digitaria exilis</i> annotation	<i>Digitaria exilis</i>	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	<i>Digitaria exilis</i> genome assembly, chromosome: 5B	<i>Digitaria exilis</i>	34.2	94.6	77%	35	100.00%	38778756
LR761613.1	<i>Digitaria exilis</i> genome assembly, chromosome: 5A	<i>Digitaria exilis</i>	34.2	64.4	77%	35	100.00%	40393928
LR761606.1	<i>Digitaria exilis</i> genome assembly, chromosome: 1B	<i>Digitaria exilis</i>	34.2	64.4	77%	35	100.00%	35142973
LR761605.1	<i>Digitaria exilis</i> genome assembly, chromosome: 1A	<i>Digitaria exilis</i>	34.2	94.6	77%	35	100.00%	37200191
XM_017289033.2	PREDICTED: <i>Drosophila miranda</i> E3 ubiquitin-protein ligase MARCH5 (LOC108157133), transcript variant X2, mRNA	<i>Drosophila miranda</i>	34.2	34.2	77%	35	100.00%	1747
XM_017289032.2	PREDICTED: <i>Drosophila miranda</i> E3 ubiquitin-protein ligase MARCH5 (LOC108157133), transcript variant X1, mRNA	<i>Drosophila miranda</i>	34.2	34.2	77%	35	100.00%	1726
CP050976.1	<i>Streptomyces</i> sp. RPA4-5 chromosome	<i>Streptomyces</i> 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	<i>Canis lupus familiaris</i> 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	<i>Canis lupus familiaris</i> breed Labrador retriever chromosome 20b	dog	34.2	64.4	77%	35	100.00%	58321546
XM_033060848.1	PREDICTED: <i>Catharus ustulatus</i> dachsous cadherin-related 2 (DCHS2), mRNA	Swainson's thrush	34.2	34.2	95%	35	95.24%	1839
AP022720.1	<i>Plectropomus leopardus</i> DNA, chromosome 21, nearly complete sequence	leopard coralgroup	34.2	64.4	81%	35	100.00%	25625633
XM_015886902.2	PREDICTED: <i>Coturnix japonica</i> 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: <i>Coturnix japonica</i> 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: <i>Coturnix japonica</i> SWI/SNF related, matrix associated, actin dependent regulator of chromatin subfamily c member 2	Japanese quail	34.2	34.2	77%	35	100.00%	3542

	(SMARCC2), transcript variant X4, mRNA 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_015886899.2								
	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_015886898.2								
	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
XM_015886897.2								
	Streptomyces aureovercillatus strain HN6 chromosome, complete genome	Streptomyces aureovercillatus	34.2	64.4	100%	35	100.00%	9617533
CP048641.1								
	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_032198087.1								
	PREDICTED: Rhipicephalus microplus probable cation-transporting ATPase 13A3 (LOC119161628), mRNA	southern cattle tick	34.2	34.2	77%	35	100.00%	12141
XM_037414197.1								
	PREDICTED: Peromyscus leucopus inorganic pyrophosphatase 1 (Ppa1), transcript variant X2, mRNA	white-footed mouse	34.2	34.2	77%	35	100.00%	1301
XM_028886172.2								
	Trueperella sp. 19M2397 strain 19OD0592 chromosome, complete genome	Trueperella sp. 19M2397	34.2	64.4	86%	35	100.00%	2373092
CP063212.1								
	Trueperella sp. 19M2397 strain 15A0121 chromosome, complete genome	Trueperella sp. 19M2397	34.2	34.2	77%	35	100.00%	2327149
CP063213.1								
	Trueperella sp. 19M2397 chromosome, complete genome	Trueperella sp. 19M2397	34.2	34.2	77%	35	100.00%	2282684
CP053291.1								
	Janibacter indicus strain TT2 chromosome, complete genome	Janibacter indicus	34.2	34.2	77%	35	100.00%	3663756
CP062789.1								
	Jeongeupia sp. HS-3 DNA, complete genome	Jeongeupia sp. HS-3	34.2	34.2	77%	35	100.00%	3396565
AP024094.1								
	Pseudomonas putida strain ZXP-20 chromosome, complete genome	Pseudomonas putida	34.2	64.4	86%	35	100.00%	6511858
CP061723.1								
1/2 7/1 7/2 9/1								
AP023477.1								
	Homo sapiens DNA chromosome 17 nearly complete genome	human	34.2	64.4	100%	35	100.00%	80688777
1/2 2/2 3/2 3/3 6/1 10/1								
	PREDICTED: Onychomys torridus coordinator of PRMT5 and differentiation stimulator (Copr), transcript variant X2, mRNA	southern grasshopper mouse	34.2	34.2	77%	35	100.00%	809
XM_036166397.1								
	PREDICTED: Onychomys torridus coordinator of PRMT5 and differentiation stimulator (Copr), transcript variant X1, mRNA	southern grasshopper mouse	34.2	34.2	77%	35	100.00%	866
XM_036166396.1								
	Beauveria bassiana strain HN6 chromosome 1	Beauveria bassiana	34.2	94.6	100%	35	100.00%	6424025
CP045886.1								
	Streptomyces libani subsp. rufus NBRC 15424 DNA, complete genome	Streptomyces libani subsp. rufus	34.2	94.6	95%	35	95.24%	8522827
AP023408.1								
	Onychomys torridus genome assembly, chromosome: 17	southern grasshopper mouse	34.2	64.4	86%	35	100.00%	64256681
LR877204.1								
	Streptomyces sp. SCUT-3 chromosome, complete genome	Streptomyces sp. SCUT-3	34.2	64.4	86%	35	100.00%	6075149
CP046907.1								
	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_035424987.1								
	PREDICTED: Anguilla anguilla N-acyl-aromatic-L-amino acid amidohydrolase (carboxylate-forming) B-like (LOC118231297),	European eel	34.2	34.2	77%	35	100.00%	1452
XM_035424986.1								

	transcript variant X3, mRNA								
XM_035424985.1	PREDICTED: Anguilla anguilla N-acyl-aromatic-L-amino acid amidohydrolase (carboxylate-forming) B-like (LOC118231297), transcript variant X2, mRNA	European eel	34.2	34.2	77%	35	100.00%	1581	
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	

VISTA DI ALLINEAMENTO VERTICALE AGGANCIATO A QUERY CON LETTERE D'IDENTITA'

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	CTGCCCGCCGCCACAGCCACC	92
XM 012646526.1	2611	CTGCCCGCCGCCACAGCCACC	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

<u>LR761614.1</u>	6876829	CCGCCGCACACAGCCAC	6876845
<u>LR761614.1</u>	5539866	CGCCGCACACAGCCA	5539852
<u>LR761614.1</u>	37962811	CGCCGCACACAGCCA	37962825
<u>LR761613.1</u>	7121733	CCGCCGCACACAGCCAC	7121749
<u>LR761613.1</u>	39921398	CGCCGCACACAGCCA	39921412
<u>LR761606.1</u>	5378840	CCGCCGCACACAGCCAC	5378824
<u>LR761606.1</u>	14804528	GCCGCACACAGCCAC	14804542
<u>LR761605.1</u>	5334243	CCGCCGCACACAGCCAC	5334227
<u>LR761605.1</u>	19482845	CGCCGCACACAGCCA	19482831
<u>LR761605.1</u>	29045300	CGCCGCACACAGCCA	29045286
<u>XM 017289033.2</u>	248	GCCCGCCGCACACAGCC	264
<u>XM 017289032.2</u>	227	GCCCGCCGCACACAGCC	243
<u>CP050976.1</u>	8937379	TGCCCGCCGCCACAGCCACC	8937399
<u>CP050976.1</u>	5011301	CGCCGCACACAGCC	5011287
<u>CP050571.1</u>	55652137	CTGCCCGCCGCACACAG	55652121
<u>CP050571.1</u>	39983513	CTGCCCGCCGCACAC	39983527
<u>CP050631.1</u>	55647723	CTGCCCGCCGCACACAG	55647707
<u>CP050631.1</u>	39973297	CTGCCCGCCGCACAC	39973311
<u>XM 033060848.1</u>	110	CTGCCCGCCCCACACAGCCAC	90
<u>AP022720.1</u>	23859507	GCCCGCCGCACACAGCC	23859491
<u>AP022720.1</u>	9168476	CGCCGCACACAGCCA	9168490
<u>XM 015886902.2</u>	3236	GCCCGCCGCACACAGCC	3252
<u>XM 015886901.2</u>	3428	GCCCGCCGCACACAGCC	3444
<u>XM 015886900.2</u>	3431	GCCCGCCGCACACAGCC	3447
<u>XM 015886899.2</u>	3518	GCCCGCCGCACACAGCC	3534
<u>XM 015886898.2</u>	3521	GCCCGCCGCACACAGCC	3537
<u>XM 015886897.2</u>	3524	GCCCGCCGCACACAGCC	3540
<u>CP048641.1</u>	2537708	CGCCGCACACAGCCACC	2537692
<u>CP048641.1</u>	9124121	CTGCCCGCCGCACAC	9124135
<u>XM 032198087.1</u>	51	CCCGCCGCACACAGCCA	35
<u>XM 032198086.1</u>	51	CCCGCCGCACACAGCCA	35
<u>XM 032198085.1</u>	51	CCCGCCGCACACAGCCA	35
<u>XM 032198084.1</u>	51	CCCGCCGCACACAGCCA	35
<u>XM 037414197.1</u>	3337	CGCCGCACACAGCCACC	3321
<u>XM 028886172.2</u>	94	CCGCCGCACACAGCCAC	78
<u>CP062789.1</u>	509128	GCCCGCCGCACACAGCC	509144
<u>AP024094.1</u>	384518	CGCCGCACACAGCCACC	384502
<u>CP061723.1</u>	1554612	TGCCCGCCGCACACAGCCA	1554594
<u>CP061723.1</u>	3017879	TGCCCGCCGCACACAGC	3017863
<u>AP023477.1</u>	76694453	CTGCCCGCCGCACACAG	76694437
<u>AP023477.1</u>	74580794	CCGCACACAGCCACC	74580808
<u>XM 036166397.1</u>	56	CCCGCCGCACACAGCCA	40
<u>XM 036166396.1</u>	56	CCCGCCGCACACAGCCA	40
<u>CP045886.1</u>	2323905	CTGCCCGCCGCACACAG	2323889
<u>CP045886.1</u>	5608051	CCGCACACAGCCACC	5608037
<u>CP045886.1</u>	6346327	TGCCCGCCGCACACA	6346313
<u>AP023408.1</u>	27661	TGCCCGCCGCCACAGCCACC	27641
<u>AP023408.1</u>	613141	CCGCCGCACACAGCC	613155
<u>AP023408.1</u>	7134375	CCCGCCGCACACAGC	7134361
<u>LR877204.1</u>	29872798	CCCGCCGCACACAGTCACC	29872816
<u>LR877204.1</u>	10928480	CCCGCCGCACACAGCCA	10928496
<u>CP046907.1</u>	2339206	CTGCCCGCCGCACACAGCC	2339188
<u>CP046907.1</u>	2869276	GCCCGCCGCACACAGCC	2869260
<u>XM 035424987.1</u>	217	CCGCCGCACACAGCCAC	201
<u>XM 035424986.1</u>	314	CCGCCGCACACAGCCAC	298
<u>XM 035424985.1</u>	443	CCGCCGCACACAGCCAC	427
<u>XM 035424984.1</u>	483	CCGCCGCACACAGCCAC	467
<u>CP054635.1</u>	839409	GCCCGCCGCACACAGCC	839425

<u>CP056779.1</u>	3830421	GCCCGCCGCACACAGCC	3830437
<u>CP056117.1</u>	2102096	GCCCGCCGCACACAGCC	2102112
<u>LR812053.1</u>	11096994	CCGCCGCACACAGCCAC	11096978
<u>LR812053.1</u>	32205844	GCCGCACACAGCCACC	32205859
<u>LR812053.1</u>	16778702	TGCCCGCCGCACACA	16778716
<u>LS423635.1</u>	2666403	CTGCCCGCCGCACACAG	2666387
<u>LR812609.1</u>	10255845	CCGCCGCACACAGCCAC	10255829
<u>LR812609.1</u>	30652221	GCCGCACACAGCCACC	30652236
<u>LR812609.1</u>	15868371	TGCCCGCCGCACACA	15868357
<u>LR812584.1</u>	10880970	CCGCCGCACACAGCCAC	10880954
<u>LR812584.1</u>	32007184	GCCGCACACAGCCACC	32007199
<u>LR812584.1</u>	16376697	TGCCCGCCGCACACA	16376711
<u>LR812534.1</u>	29479406	CCGCCGCACACAGCCAC	29479422
<u>LR812534.1</u>	40709593	GCCGCACACAGCCACC	40709608
<u>LR812534.1</u>	23502108	TGCCCGCCGCACACA	23502094

PRODOTTI DEGLI ALLINEAMENTI SIGNIFICATIVI (CDS FEATURE)

- RNA 3'-phosphate cyclase [GenBank](#); [GenBank](#)
- **Zinc**/iron permease [GenBank](#)
- Hypothetical protein [GenBank](#); [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](#); [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](#)

SEQUENZA N° 5/2

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

SEQUENZA N° 5/2 (T=3; C=1; G=2; A=0)

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° 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
Sequenza 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: **CGCAAAGCCTGCGCGGCCACC**

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	Glycocalis alkaliphilus strain 6B-8, complete genome	Glycocalis 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	Xanthomonas campestris pv. campestris str. CN16	36.2	161	100%	9.1	100.00%	5023147

	genome								
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 756C, complete genome	Xanthomonas campestris pv. raphani 756C	36.2	128	81%	9.1	100.00%	4941214
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

VISTA DI ALLINEAMENTO VERTICALE AGGANCIATO A QUERY CON LETTERE D'IDENTITA'

Query range 1: 1 to 22

Query	1	CGCAAAGCCTGCGCGGCCACC	22
CP007444.1	3529116	CGCAAAGCCTGCGCGGCCCA	3529135
CP007444.1	2514930	AAAGCCTGCGCGCGC	2514916
CP007444.1	2948117	CGCAAAGCCTGCGCG	2948131
CP060388.1	51295	CGCAAAGCCTGCGCGGCCA	51314
CP060388.1	45720	GCCTGCGCGGCCAC	45734
CP049736.1	450092	CGCAAAGCCTGCGCGGCC	450074
CP049736.1	429383	GCCTGCGCGGCCAC	429397
CP049736.1	2423660	AAAGCCTGCGCGCGC	2423646
CP049736.1	4680990	AGCCTGCGCGGCCA	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	GCAAAGCCTGCGCGCGC	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

<u>CP032509.1</u>	1460723	CGCAAAGCCTGCGCG	1460737
<u>CP024676.1</u>	1844362	CGCAAAGCCTGCGCGCGC	1844379
<u>CP031606.1</u>	2902444	CAAAGCCTGCGCGCGCCA	2902427
<u>CP031606.1</u>	4679086	AAGCCTGCGCGCGCC	4679072
<u>CP034811.1</u>	170428	CGCAAAGCCTGCGCGCGC	170445
<u>CP025522.1</u>	2728803	CGCAAAGCCTGCGCGCGC	2728820
<u>CP025521.1</u>	2746473	CGCAAAGCCTGCGCGCGC	2746490
<u>LR134485.1</u>	2222998	CGCAAAGCCTGCGCGCGC	2222981
<u>MH449685.1</u>	21601	CAAAGCCTGCGCGCGCCA	21618
<u>CP025997.1</u>	284550	AAGCCTGCGCGCGCCACC	284567
<u>CP025997.1</u>	1037954	CAAAGCCTGCGCGCG	1037940
<u>CP028843.1</u>	5110156	CGCAAAGCCTGCGCGCGC	5110139
<u>CP023467.1</u>	2972650	CGCAAAGCCTGCGCGCGC	2972633
<u>CP023467.1</u>	3990021	GCCTGCGCGCGCCAC	3990007
<u>CP023467.1</u>	4291883	CGCAAAGCCTGCGCG	4291897
<u>CP014030.2</u>	2843726	CGCAAAGCCTGCGCGCGC	2843743
<u>CP026124.1</u>	2935136	AAGCCTGCGCGCGCCACC	2935119
<u>CP026124.1</u>	1189807	GCCTGCGCGCGCCAC	1189821
<u>CP025750.1</u>	3786086	CGCAAAGCCTGCGCGCGC	3786069
<u>CP025750.1</u>	137235	GCCTGCGCGCGCCACC	137250
<u>CP025750.1</u>	4208327	CGCAAAGCCTGCGCGC	4208342
<u>CP025750.1</u>	4420324	GCAAAGCCTGCGCGC	4420310
<u>CP025750.1</u>	5050122	CGCAAAGCCTGCGCG	5050136
<u>CP021913.1</u>	1359938	CGCAAAGCCTGCGGGCGCCACC	1359917
<u>CP021913.1</u>	460680	GCCTGCGCGCGCCAC	460666
<u>CP017389.1</u>	3706375	CGCAAAGCCTGCGCGCGC	3706358
<u>CP017389.1</u>	133681	GCCTGCGCGCGCCACC	133696
<u>CP017389.1</u>	4144639	CGCAAAGCCTGCGCGC	4144654
<u>CP017389.1</u>	4355298	GCAAAGCCTGCGCGC	4355284
<u>CP017389.1</u>	4990731	CGCAAAGCCTGCGCG	4990745
<u>CP017323.1</u>	3723299	CGCAAAGCCTGCGCGCGC	3723282
<u>CP017323.1</u>	133677	GCCTGCGCGCGCCACC	133692
<u>CP017323.1</u>	4161583	CGCAAAGCCTGCGCGC	4161598
<u>CP017323.1</u>	4372242	GCAAAGCCTGCGCGC	4372228
<u>CP017323.1</u>	5007672	CGCAAAGCCTGCGCG	5007686
<u>CP017319.1</u>	3751445	CGCAAAGCCTGCGCGCGC	3751428
<u>CP017319.1</u>	167322	GCCTGCGCGCGCCACC	167337
<u>CP017319.1</u>	4196367	CGCAAAGCCTGCGCGC	4196382
<u>CP017319.1</u>	3039234	GCCTGCGCGCGCCAC	3039248
<u>CP017319.1</u>	4407303	GCAAAGCCTGCGCGC	4407289
<u>CP017319.1</u>	5110592	CGCAAAGCCTGCGCG	5110606
<u>CP017317.1</u>	3706634	CGCAAAGCCTGCGCGCGC	3706617
<u>CP017317.1</u>	133679	GCCTGCGCGCGCCACC	133694
<u>CP017317.1</u>	4144914	CGCAAAGCCTGCGCGC	4144929
<u>CP017317.1</u>	4375615	GCAAAGCCTGCGCGC	4375601
<u>CP017317.1</u>	5011045	CGCAAAGCCTGCGCG	5011059
<u>CP017310.1</u>	1296408	CGCAAAGCCTGCGCGCGC	1296425
<u>CP017310.1</u>	135689	GCCTGCGCGCGCCACC	135704
<u>CP017310.1</u>	849581	CGCAAAGCCTGCGCGC	849566
<u>CP017310.1</u>	638652	GCAAAGCCTGCGCGC	638666
<u>CP017310.1</u>	1989563	GCCTGCGCGCGCCAC	1989549
<u>CP017310.1</u>	5092620	CGCAAAGCCTGCGCG	5092634
<u>CP017308.1</u>	3753299	CGCAAAGCCTGCGCGCGC	3753282
<u>CP017308.1</u>	136990	GCCTGCGCGCGCCACC	137005
<u>CP017308.1</u>	4169082	CGCAAAGCCTGCGCGC	4169097
<u>CP017308.1</u>	3047944	GCCTGCGCGCGCCAC	3047958
<u>CP017308.1</u>	4378878	GCAAAGCCTGCGCGC	4378864
<u>CP017308.1</u>	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	GCAAGCCTGCGCGCGCCAC	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	GCAAGCCTGCGCGCGCCAC	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	CGCAAAGCCTGCGCCC GCCACC	390
CP004886.1	3609442	CGCAAAGCCC GCGCGCGCCACC	3609463
CP003740.1	3869610	GCAAAGCCTGCGCGCGCC	3869593
AP024109.1	1908047	CGCAAAGCCTGCGCGCGC	1908064
AP024109.1	118685	AAAGCCTGCGCGCGC	118699
AP024109.1	717368	GCAAAGCCTGCGCGC	717382

<u>XM 037898474.1</u>	57	AAGCCTGCGCGCGCCACC	40
<u>XM 037898464.1</u>	57	AAGCCTGCGCGCGCCACC	40
<u>XM 007068977.3</u>	56	AAGCCTGCGCGCGCCACC	39
<u>LR884483.1</u>	6407061	CGCAAAGCCTGCGCGCGC	6407044
<u>AP024096.1</u>	6396520	CGCAAAGCCTGCGCGCGC	6396503
<u>CP062066.1</u>	3767676	CGCAAAGCCTGCGCGCGC	3767659
<u>CP062066.1</u>	144567	GCCTGCGCGCGCCACC	144582
<u>CP062066.1</u>	4207108	CGCAAAGCCTGCGCGC	4207123
<u>CP062066.1</u>	2614591	GCAAAGCCTGCGCGC	2614577
<u>CP062066.1</u>	3072312	GCCTGCGCGCGCCAC	3072326
<u>CP062066.1</u>	4420098	GCAAAGCCTGCGCGC	4420084
<u>CP062066.1</u>	5080952	CGCAAAGCCTGCGCG	5080966
<u>CP060008.1</u>	2102525	CGCAAAGCCTGCGCGCGC	2102508
<u>CP058243.1</u>	3652732	CGCAAAGCCTGCGCGCGC	3652715
<u>CP058243.1</u>	136197	GCCTGCGCGCGCCACC	136212
<u>CP058243.1</u>	4081788	CGCAAAGCCTGCGCGC	4081803
<u>CP058243.1</u>	4292203	GCAAAGCCTGCGCGC	4292189
<u>CP058243.1</u>	4923527	CGCAAAGCCTGCGCG	4923541
<u>CP002789.1</u>	1248491	CGCAAAGCCTGCGCGCGC	1248508
<u>CP002789.1</u>	818572	CGCAAAGCCTGCGCGC	818557
<u>CP002789.1</u>	609183	GCAAAGCCTGCGCGC	609197
<u>CP002789.1</u>	4918339	CGCAAAGCCTGCGCG	4918353
<u>CP002038.1</u>	2882472	CGCAAAGCCTGCGCGCGC	2882455
<u>CP002038.1</u>	697082	CGCAAAGCCTGCGCGC	697067
<u>FN563999.1</u>	441	AAGCCTGCGCGCGCCACC	424
<u>FN563998.1</u>	441	AAGCCTGCGCGCGCCACC	424
<u>AE008922.1</u>	1184098	CGCAAAGCCTGCGCGCGC	1184115
<u>AE008922.1</u>	138222	GCCTGCGCGCGCCACC	138237
<u>AE008922.1</u>	767969	CGCAAAGCCTGCGCGC	767954
<u>AE008922.1</u>	1919634	GCCTGCGCGCGCCAC	1919620
<u>AE008922.1</u>	2363298	GCAAAGCCTGCGCGC	2363312
<u>AE008922.1</u>	4430498	GCAAAGCCTGCGCGC	4430484
<u>AE008922.1</u>	5054392	CGCAAAGCCTGCGCG	5054406
<u>AM920689.1</u>	3790946	CGCAAAGCCTGCGCGCGC	3790929
<u>AM920689.1</u>	137129	GCCTGCGCGCGCCACC	137144
<u>AM920689.1</u>	4213655	CGCAAAGCCTGCGCGC	4213670
<u>AM920689.1</u>	4425418	GCAAAGCCTGCGCGC	4425404
<u>AM920689.1</u>	5056686	CGCAAAGCCTGCGCG	5056700
<u>AM167904.1</u>	2733000	CGCAAAGCCTGCGCGCGC	2733017
<u>AM167904.1</u>	1453432	AAGCCTGCGCGCGCC	1453446
<u>CP000050.1</u>	3856451	CGCAAAGCCTGCGCGCGC	3856434
<u>CP000050.1</u>	137818	GCCTGCGCGCGCCACC	137833
<u>CP000050.1</u>	4272579	CGCAAAGCCTGCGCGC	4272594
<u>CP000050.1</u>	2616238	GCAAAGCCTGCGCGC	2616224
<u>CP000050.1</u>	3120500	GCCTGCGCGCGCCAC	3120514
<u>CP000050.1</u>	4488081	GCAAAGCCTGCGCGC	4488067
<u>CP000050.1</u>	5126916	CGCAAAGCCTGCGCG	5126930
<u>BA000012.4</u>	3988442	AAGCCTGCGCGCGCCACC	3988459

SEQUENZA N° 5/2 (T=3; C=1; G=2; A=0)

Query: **CGCAAAGCCTGCGCGGCCACC**

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
.....

SEQUENZA N° 6/1

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

SEQUENZA N° 6/1 (T=2; C=3; G=0; A=1)

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° 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
Sequenza 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	Parabassiss 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 PREDICTED: Cicer arietinum pentatricopeptide repeat-containing protein At1g18485 (LOC101509188), transcript variant X1, mRNA	gilthead seabream	34.2	34.2	77%	35	100.00%	25861820
XM_004516808.3	PREDICTED: Papaver somniferum uncharacterized LOC113300348 (LOC113300348), transcript variant X7, ncRNA	chickpea	34.2	34.2	77%	35	100.00%	3685
XR_003335731.1	PREDICTED: Papaver somniferum uncharacterized LOC113300348 (LOC113300348), transcript variant X6, ncRNA	opium poppy	34.2	34.2	77%	35	100.00%	4266
XR_003335730.1	PREDICTED: Papaver somniferum uncharacterized LOC113300348 (LOC113300348), transcript variant X5, ncRNA	opium poppy	34.2	34.2	77%	35	100.00%	4281
XR_003335729.1	PREDICTED: Papaver somniferum uncharacterized LOC113300348 (LOC113300348), transcript variant X4, ncRNA	opium poppy	34.2	34.2	77%	35	100.00%	4264
XR_003335728.1	PREDICTED: Papaver somniferum uncharacterized LOC113300348 (LOC113300348), transcript variant X3, ncRNA	opium poppy	34.2	34.2	77%	35	100.00%	4165
XR_003335727.1	PREDICTED: Papaver somniferum uncharacterized LOC113300348 (LOC113300348), transcript variant X2, ncRNA	opium poppy	34.2	34.2	77%	35	100.00%	4509
XR_003335726.1	PREDICTED: Papaver somniferum uncharacterized LOC113300348 (LOC113300348), transcript variant X1, ncRNA	opium poppy	34.2	34.2	77%	35	100.00%	4270
XR_003335725.1	PREDICTED: Eptesicus fuscus homeobox C6 (HOXC6), transcript variant X1, mRNA	big brown bat	34.2	34.2	77%	35	100.00%	10240
XM_008140785.2	Anopheles stephensi strain Indian chromosome 2R	Asian malaria mosquito	34.2	126	90%	35	100.00%	58051481
CP032299.1	Anopheles stephensi strain SDA-500 chromosome 2R	Asian malaria mosquito	34.2	64.4	90%	35	100.00%	62568927
CP032232.1	PREDICTED: Alligator sinensis cystathionine beta-synthase (LOC102379548), transcript variant X9, mRNA	Chinese alligator	34.2	34.2	77%	35	100.00%	4432
XM_025214551.1	PREDICTED: Alligator sinensis cystathionine beta-synthase (LOC102379548), transcript variant X8, mRNA	Chinese alligator	34.2	34.2	77%	35	100.00%	4437
XM_025214550.1	PREDICTED: Alligator sinensis cystathionine beta-synthase (LOC102379548), transcript variant X7, mRNA	Chinese alligator	34.2	34.2	77%	35	100.00%	4660
XM_025214549.1	PREDICTED: Alligator sinensis cystathionine beta-synthase (LOC102379548), transcript variant X6, mRNA	Chinese alligator	34.2	34.2	77%	35	100.00%	4505
XM_025214548.1	PREDICTED: Alligator sinensis cystathionine beta-synthase (LOC102379548), transcript variant X5, mRNA	Chinese alligator	34.2	34.2	77%	35	100.00%	4544
XM_025214547.1	PREDICTED: Alligator sinensis cystathionine beta-synthase (LOC102379548), transcript variant X4, mRNA	Chinese alligator	34.2	34.2	77%	35	100.00%	4507
XM_025214546.1	PREDICTED: Alligator sinensis cystathionine beta-synthase (LOC102379548), transcript variant X3, mRNA	Chinese alligator	34.2	34.2	77%	35	100.00%	4538
XM_025214545.1	PREDICTED: Alligator sinensis cystathionine beta-synthase (LOC102379548), transcript variant X2, mRNA	Chinese alligator	34.2	34.2	77%	35	100.00%	4528
XM_025214544.1	PREDICTED: Alligator sinensis cystathionine beta-synthase (LOC102379548), transcript variant X1, mRNA	Chinese alligator	34.2	34.2	77%	35	100.00%	4767

AC277723.1	Homo sapiens chromosome 7 clone VMRC62-424L4, complete sequence	human	34.2	34.2	77%	35	100.00%	162561
1/2 2/2 3/2 3/3 5/1 10/1								
CP020671.1	Oryzias latipes strain Hd-rR chromosome 7 sequence	Japanese medaka	34.2	34.2	77%	35	100.00%	34573382
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 lm 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 (pad2), 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 (pad2), 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 (pad2), 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	Japanese medaka	34.2	34.2	77%	35	100.00%	7016

XM_004070826.4	synaptotagmin-2 (LOC101161806), transcript variant X2, mRNA PREDICTED: Oryzias latipes	Japanese medaka	34.2	34.2	77%	35	100.00%	7090
XM_023062592.1	synaptotagmin-2 (LOC101161806), transcript variant X1, mRNA PREDICTED: Onthophagus taurus uncharacterized LOC111427443 (LOC111427443), transcript variant X2, mRNA PREDICTED: Onthophagus taurus uncharacterized LOC111427443 (LOC111427443), transcript variant X1, mRNA	Onthophagus taurus	34.2	34.2	77%	35	100.00%	2585
XM_023062591.1	Lupinus angustifolius cultivar Tanjil chromosome LG-08	Onthophagus taurus narrow-leaved blue lupine	34.2	34.2	77%	35	100.00%	2735
CP023120.1	PREDICTED: Hippocampus comes uncharacterized LOC109516675 (LOC109516675), transcript variant X1, ncRNA	tiger tail seahorse Saccoglossus kowalevskii	34.2	34.2	77%	35	100.00%	30636785
XR_002168388.1	Scaffolds_166 genomic sequence	tiger tail seahorse Saccoglossus kowalevskii	34.2	34.2	77%	35	100.00%	966
JX899680.1		Scaffolds_166 genomic sequence	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 Acanthopagrus latus isolate v.2019 genome assembly, chromosome: 11	potato	34.2	34.2	77%	35	100.00%	57022023
LR884470.1	Danio rerio strain Nadia (NA) genome assembly, chromosome: 1	yellowfin seabream	34.2	34.2	77%	35	100.00%	31321781
LR812569.1	Danio rerio strain Cooch Behar (CB) genome assembly, chromosome: 1	zebrafish	34.2	34.2	95%	35	95.24%	58221445
LR812544.1		zebrafish	34.2	34.2	95%	35	95.24%	54094920
AC073347.3	Homo sapiens BAC clone RP11-775L16 from 7, complete sequence PREDICTED: Peromyscus leucopus DExD-box helicase 21 (Ddx21), transcript variant X2, mRNA	human	34.2	34.2	77%	35	100.00%	197144
XM_028886192.2	PREDICTED: Peromyscus leucopus DExD-box helicase 21 (Ddx21), transcript variant X1, mRNA	white-footed mouse	32.2	32.2	72%	137	100.00%	3521
XM_037200273.1	PREDICTED: Pollicipes pollicipes glycerol-3-phosphate acyltransferase 3-like (LOC119109572), transcript variant X3, 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 X2, 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 X1, mRNA	Pollicipes pollicipes	32.2	32.2	72%	137	100.00%	4195
XM_037233251.1	PREDICTED: Pollicipes pollicipes glycerol-3-phosphate acyltransferase 3-like (LOC119108870), transcript variant X2, mRNA	Pollicipes pollicipes	32.2	32.2	72%	137	100.00%	4198
XM_037232415.1	PREDICTED: Pollicipes pollicipes glycerol-3-phosphate acyltransferase 3-like (LOC119108870), transcript variant X1, mRNA	Pollicipes pollicipes	32.2	32.2	72%	137	100.00%	4190
XM_037232414.1	Macrobrachium nipponense isolate FS-2020 chromosome 34	Pollicipes pollicipes Macrobrachium nipponense	32.2	32.2	72%	137	100.00%	4193
CP062047.1			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 Acanthopagrus latus isolate v.2019 genome assembly, chromosome:	potato	32.2	62.4	90%	137	100.00%	68020514
LR884475.1		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 (dnah1), mRNA	red-bellied piranha	32.2	32.2	72%	137	100.00%	13212
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	Solanum tuberosum cultivar Solyntus chromosome 3	potato	32.2	32.2	72%	137	100.00%	63701590
CP055235.1	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

VISTA DI ALLINEAMENTO VERTICALE AGGANCIATO A QUERY CON LETTERE D'IDENTITA'

Query range 1: 1 to 22

Query	1	ACATTTCAATCGAGCACAATAA	22
CP017414.1	3009550	TTTCAATCGAGCACAATAA	3009532
XM_012937820.2	953	ACATTTCAATCGAGCACA	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	CATTTCAATCGAGCACA	3204766
LK052913.1	4606	CATTTCAATCGAGCACA	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	ATTTCAATCGAGCACA	15016171
XM_004516808.3	1822	ATTTCAATCGAGCACA	1806
XR_003335731.1	1679	TCAATCGAGCACAATAA	1663
XR_003335730.1	1679	TCAATCGAGCACAATAA	1663
XR_003335729.1	1679	TCAATCGAGCACAATAA	1663

<u>XR 003335728.1</u>	1679	TCAATCGAGCACAATAA	1663
<u>XR 003335727.1</u>	1679	TCAATCGAGCACAATAA	1663
<u>XR 003335726.1</u>	1679	TCAATCGAGCACAATAA	1663
<u>XR 003335725.1</u>	1679	TCAATCGAGCACAATAA	1663
<u>XM 008140785.2</u>	1155	ACATTTCAATCGAGCAC	1139
<u>CP032299.1</u>	10003448	TTTCAATCGAGCACCATAA	10003430
<u>CP032299.1</u>	22162784	ATTTCAATCGAGCACAA	22162768
<u>CP032299.1</u>	43636053	CAATCGAGCACAATAA	43636068
<u>CP032299.1</u>	57624346	AATCGAGCACAATAA	57624360
<u>CP032232.1</u>	10453175	TTTCAATCGAGCACCATAA	10453157
<u>CP032232.1</u>	22474271	ATTTCAATCGAGCACAA	22474255
<u>XM 025214551.1</u>	2673	CATTTCAATCGAGCACA	2657
<u>XM 025214550.1</u>	2678	CATTTCAATCGAGCACA	2662
<u>XM 025214549.1</u>	2901	CATTTCAATCGAGCACA	2885
<u>XM 025214548.1</u>	2746	CATTTCAATCGAGCACA	2730
<u>XM 025214547.1</u>	2785	CATTTCAATCGAGCACA	2769
<u>XM 025214546.1</u>	2748	CATTTCAATCGAGCACA	2732
<u>XM 025214545.1</u>	2779	CATTTCAATCGAGCACA	2763
<u>XM 025214544.1</u>	2769	CATTTCAATCGAGCACA	2753
<u>XM 025214543.1</u>	3008	CATTTCAATCGAGCACA	2992
<u>AC277723.1</u>	162338	ATTTCAATCGAGCACAA	162322
<u>CP020671.1</u>	20151957	ACATTTCAATCGAGCAC	20151941
<u>CP019094.1</u>	6869168	TTCAATCGAGCACAATA	6869152
<u>XM 019596250.1</u>	130	ATTTCAATCGAGCACAA	114
<u>XM 019487198.1</u>	2871	CATTTCAATCGAGCACA	2855
<u>XM 019487197.1</u>	3103	CATTTCAATCGAGCACA	3087
<u>XM 019487196.1</u>	3028	CATTTCAATCGAGCACA	3012
<u>XM 006260474.3</u>	3092	CATTTCAATCGAGCACA	3076
<u>XM 014378881.1</u>	395	TTTCAATCGAGCACAAT	379
<u>XM 013452927.1</u>	225	TTTCAATCGAGCACAAT	209
<u>XM 009826460.1</u>	1377	ACATTTCAATCGAGCAC	1361
<u>XM 007940675.1</u>	1997	ACATTTCAATCGAGCAC	1981
<u>XM 006835838.1</u>	1997	ACATTTCAATCGAGCAC	1981
<u>AC162031.12</u>	28360	ATTTCAATCGAGCACAA	28344
<u>XM 034868321.1</u>	2593	CATTTCAATCGAGCACA	2609
<u>XM 034868320.1</u>	2608	CATTTCAATCGAGCACA	2624
<u>LR792568.1</u>	10557225	ACATTTTAATCGAGCACAA	10557207
<u>LR792568.1</u>	8265588	TTCAATCGAGCACAATA	8265604
<u>XM 032513447.1</u>	2612	CATTTCAATCGAGCACA	2628
<u>LR738563.1</u>	8015903	TTTCAATCGAGCACAAT	8015919
<u>LR736850.1</u>	25837080	TTTCAATCGAGCACAAT	25837096
<u>CP032376.1</u>	11155525	TTTCAATCGAGCACAAT	11155541
<u>CP032373.1</u>	11159719	TTTCAATCGAGCACAAT	11159735
<u>CP032367.1</u>	11143339	TTTCAATCGAGCACAAT	11143355
<u>LR597437.1</u>	20047958	CATTTCAATCGAGCACA	20047974
<u>CP032252.1</u>	33076510	ATTTCAATCGAGCACAA	33076526
<u>AP019376.1</u>	4612470	CATTTCAATCGAGCACA	4612486
<u>CP025674.1</u>	2139338	CATTTCAATCGAGCACA	2139354
<u>XM 011477429.3</u>	3178	ACATTTCAATCGAGCAC	3194
<u>XM 011477428.3</u>	3113	ACATTTCAATCGAGCAC	3129
<u>XM 004070826.4</u>	3187	ACATTTCAATCGAGCAC	3203
<u>XM 023062592.1</u>	1380	TTTCAATCGAGCACAAT	1396
<u>XM 023062591.1</u>	1530	TTTCAATCGAGCACAAT	1546
<u>CP023120.1</u>	9676200	ATTTCAATCGAGCACAA	9676216
<u>XR 002168388.1</u>	71	ACATTTCAATCGAGCAC	87
<u>JX899680.1</u>	727976	TTTCAATCGAGCACAAT	727992
<u>CP055244.1</u>	20094247	TCAATCGAGCACAATAA	20094263
<u>LR884470.1</u>	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

PRODOTTI DEGLI ALLINEAMENTI SIGNIFICATIVI (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](#)

SEQUENZA N° 6/1 (T=2; C=3; G=0; A=1)

Query: **ACATTTC AATCGAGCACAATAA**

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	Pipistrellus pipistrellus							
1/2 2/2 3/2 3/3	genome assembly,	common						
3/5	chromosome: 5	pipistrelle	32.2	32.2	72%	138	100.00%	94929986
.....

SEQUENZA N° 6/2

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

SEQUENZA N° 6/2 (T=2; C=3; G=0; A=1)

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° 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
Sequenza 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: **GCATTCGGATGCACACAATAA**

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	Ipomoea triloba cultivar NCNSP0323 chromosome 1	milkfish	36.2	36.2	81%	8.9	100.00%	40347274
CP025660.1	Lateolabrax maculatus linkage group 2 sequence	trilobed morning glory	36.2	36.2	81%	8.9	100.00%	38093535
CP032595.1	Lateolabrax maculatus chromosome Lm17	spotted sea bass	36.2	36.2	81%	8.9	100.00%	29991467
CP027278.1	PREDICTED: Ipomoea nil FAD synthetase 2, chloroplastic-like (LOC109171689), transcript variant X3, mRNA	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 X2, 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 X1, mRNA	Japanese morning glory	36.2	36.2	81%	8.9	100.00%	1401
XM_019320757.1	PREDICTED: Stomoxys calcitrans homocysteine S-methyltransferase-like (LOC106090333), mRNA	Japanese morning glory	36.2	36.2	81%	8.9	100.00%	1405
XM_013256488.1	Ipomoea trifida cultivar NCNSP0306 chromosome 1	stable fly	36.2	36.2	81%	8.9	100.00%	1230
CP025644.1		Ipomoea trifida	36.2	36.2	81%	8.9	100.00%	32247286

LT94792.1	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	house mouse	36.2	36.2	81%	8.9	100.00%	39885
JN949026.1	Ggt6:tm1a(KOMP)Wtsi; transgenic	house mouse	36.2	36.2	81%	8.9	100.00%	39885
JN949025.1	Mus musculus targeted non-conditional, lacZ-tagged mutant allele	house mouse	36.2	36.2	81%	8.9	100.00%	39885
AF345929.1	Mybbp1a:tm1e(EUCOMM)Hmgu; transgenic	house mouse	36.2	36.2	81%	8.9	100.00%	38333
AL662812.12	Mus musculus targeted KO-first, conditional ready, lacZ-tagged mutant allele	house mouse	36.2	36.2	81%	8.9	100.00%	38404
LR792833.1	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
LR761618.1	Mouse DNA sequence from clone RP23-48A2 on chromosome 11, complete sequence	house mouse	36.2	36.2	81%	8.9	100.00%	258521
XM_017289635.2	Digitaria exilis annotation	Digitaria exilis	34.2	34.2	77%	35	100.00%	29094342
XM_001355143.4	Digitaria exilis genome assembly, chromosome: 7B	Digitaria exilis	34.2	34.2	77%	35	100.00%	29094342
XR_004425352.1	PREDICTED: Drosophila miranda zinc finger protein 622 (LOC108157528), mRNA	Drosophila miranda	34.2	34.2	77%	35	100.00%	1577
XM_033131157.1	PREDICTED: Drosophila pseudoobscura zinc finger protein 622 (LOC4815046), mRNA	Drosophila pseudoobscura	34.2	34.2	77%	35	100.00%	1573
XM_033131156.1	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_033131154.1	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
LR778290.1	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
CP032234.1	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
CP026254.1	Coregonus sp. 'balchen' genome assembly, chromosome: 38	Coregonus sp. 'balchen'	34.2	132	77%	35	100.00%	36774138
CP026254.1	Coregonus sp. 'balchen' genome assembly, chromosome: 37	Coregonus sp. 'balchen'	34.2	98.6	100%	35	100.00%	43663377
CP026254.1	Salmo trutta genome assembly, chromosome: 34	river trout	34.2	34.2	77%	35	100.00%	42900288
CP026254.1	Salmo trutta genome assembly, chromosome: 17	river trout	34.2	68.4	95%	35	95.24%	59764774
CP026254.1	Salmo trutta genome assembly, chromosome: 23	river trout	34.2	34.2	95%	35	95.24%	51488243
CP026254.1	Salmo trutta genome assembly, chromosome: 33	river trout	34.2	34.2	77%	35	100.00%	44893998
CP026254.1	Salmo trutta genome assembly, chromosome: 24	river trout	34.2	64.4	95%	35	100.00%	50333583
CP026254.1	PREDICTED: Drosophila persimilis zinc finger protein 622 (LOC6597574), mRNA	Drosophila persimilis	34.2	34.2	77%	35	100.00%	1532
CP026254.1	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
CP026254.1	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 mth2-24019, 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
	Bacillus paranthracis strain Gxun-30 chromosome, complete genome							
CP065149.1	Rhodobacterales bacterium strain HKCCA1288 chromosome, complete genome	Bacillus paranthracis	32.2	32.2	72%	139	100.00%	5149464
CP065161.1	1_Tdi_b3v08	Rhodobacterales bacterium	32.2	32.2	72%	139	100.00%	2663031
OA565374.1	PREDICTED: Nematolebias whitei nuclear factor, erythroid 2-like 3 (nfe2l3), mRNA	Timema douglasi	32.2	32.2	90%	139	95.00%	181011
XM_037692747.1	PREDICTED: Rhipicephalus sanguineus transmembrane protein 114-like (LOC119400263), mRNA	Rio pearlfish	32.2	32.2	72%	139	100.00%	2051
XM_037667276.1	PREDICTED: Triticum dicoccoides uncharacterized LOC119339539 (LOC119339539), mRNA	brown dog tick	32.2	32.2	72%	139	100.00%	5227
XM_037611556.1	PREDICTED: Triticum dicoccoides uncharacterized LOC119338597 (LOC119338597), mRNA	Triticum dicoccoides	32.2	32.2	72%	139	100.00%	373
XM_037610901.1	PREDICTED: Hermetia illucens uncharacterized LOC119647451 (LOC119647451), transcript variant X2, mRNA	Triticum dicoccoides	32.2	32.2	72%	139	100.00%	1260
XM_038048380.1	PREDICTED: Hermetia illucens uncharacterized LOC119647451 (LOC119647451), transcript variant X1, mRNA	Hermetia illucens	32.2	32.2	72%	139	100.00%	4697
XM_038048379.1		Hermetia illucens	32.2	32.2	72%	139	100.00%	4137

PREDICTED: Bombyx mori
 guanine nucleotide-binding protein
 subunit beta-5 (LOC119629989),
 mRNA

XM_038017770.1	domestic silkworm	32.2	32.2	72%	139	100.00%	3489
OA884554.1	Notodromas monacha	32.2	32.2	72%	139	100.00%	38594
LR899427.1	Corylus avellana genome assembly, chromosome: ca5	32.2	32.2	72%	139	100.00%	36653616
CP046680.1	Solanum tuberosum cultivar P8 chromosome 2	30.2	60.5	68%	549	100.00%	48611948

VISTA DI ALLINEAMENTO VERTICALE AGGANCIATO A QUERY CON LETTERE D'IDENTITA'

Query range 1: 1 to 22

Query	1	GCATTTTCGGATGCACACAATAA	22
LR884471.1	5872730	CATTTTCGGATGCACACAAT	5872748
LR812078.1	6416688	CATTTTCGGATGCACACAA	6416671
LR792548.1	27700847	GCATTTTCGGGAAGCACACAATAA	27700826
XM_031276121.1	386	ATTTTCGGATGCACACAAT	369
XM_030786843.1	1748	CATTTTCGGATGCACACAA	1731
LR697115.1	19324110	CATTTTCGGATGCACACAA	19324093
CP025660.1	13048241	ATTTTCGGATGCACACAAT	13048224
CP032595.1	11460181	CATTTTCGGATGCACACAA	11460164
CP027278.1	9140189	CATTTTCGGATGCACACAA	9140172
XM_019320759.1	469	ATTTTCGGATGCACACAAT	452
XM_019320758.1	449	ATTTTCGGATGCACACAAT	432
XM_019320757.1	453	ATTTTCGGATGCACACAAT	436
XM_013256488.1	851	ATTTTCGGATGCACACAAT	834
CP025644.1	8402783	ATTTTCGGATGCACACAAT	8402800
LT594792.1	27997305	ATTTTCGGATGCACACAAT	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	CATTTTCGGATGCACACA	148
XM_001355143.4	170	CATTTTCGGATGCACACA	154
XR_004425352.1	869	CATTTTCGGATGCACACA	853
XM_033131157.1	886	CATTTTCGGATGCACACA	870
XM_033131156.1	955	CATTTTCGGATGCACACA	939
XM_033131154.1	869	CATTTTCGGATGCACACA	853
LR778290.1	30972461	CATTTTCGGATGCACACA	30972445
LR778290.1	31640047	CATTTTCGGATGCACACA	31640031
LR778290.1	31724491	CATTTTCGGATGCACACA	31724507
LR778290.1	25939133	CATTTTCGGATGCACA	25939147
LR778289.1	37891224	GCATTTGGGATGCACACAATA	37891244
LR778289.1	38138938	CATTTTCGGATGCACACA	38138922
LR778289.1	1178049	GGATGCACACAATAA	1178063
LR584434.1	4737613	CATTTTCGGATGCACACA	4737597
LR584426.1	40716693	CATTTTCGGATGCACACTATAA	40716673
LR584426.1	44568833	CATTTTCGGATGCACACA	44568849
LR584421.1	6926956	CATTTGGGATGCACACAATAA	6926936
LR584414.1	8526326	CATTTTCGGATGCACACA	8526310

<u>LR584412.1</u>	27486011	GCATTTCCGGATGCACAC	27485995
<u>LR584412.1</u>	30270180	CGGATGCACACAATA	30270194
<u>XM 002022755.2</u>	129	CATTTCCGGATGCACACA	113
<u>CP032234.1</u>	29284202	CATTTCCGGATGCACACA	29284186
<u>CP032234.1</u>	30390915	TTTCCGGATGCACACA	30390901
<u>CP026254.1</u>	7772701	CATTGCCGGATGCACACAATAA	7772681
<u>XM 002071777.3</u>	408	TTTCCGGATGCACACAAT	392
<u>CP023332.1</u>	8950321	TTTCCGGATGCACACAAT	8950305
<u>AE013599.5</u>	13062817	TTTCCGGATGCACACAAT	13062801
<u>XM 007136828.1</u>	1139	TTTCCGGATGCACACAAT	1123
<u>OA888277.1</u>	4987	ATTTCCGGATGCACACAA	4971
<u>CP002563.1</u>	1553457	TTCGGATGCACACAATA	1553441
<u>AC146941.8</u>	83883	CATTTCCGGATGCACACA	83867
<u>LR812071.1</u>	20700303	TTTCCGGATGCACACAAT	20700319
<u>LR812071.1</u>	7146821	TTCGGATGCACACAA	7146835
<u>XM 034397081.1</u>	827	CATTTCCGGATGCACACA	843
<u>XM 034397072.1</u>	827	CATTTCCGGATGCACACA	843
<u>XM 034397054.1</u>	848	CATTTCCGGATGCACACA	864
<u>LR778269.1</u>	2268914	CATTTCCGGATGCAGACAAT	2268896
<u>LR778269.1</u>	12469841	CATTTGGGATGCACACAAT	12469823
<u>LR778269.1</u>	40386131	CATTTCCGGATGCACACA	40386147
<u>LR778269.1</u>	21275431	GGATGCACACAATAA	21275417
<u>LR778267.1</u>	2833714	CATTTGGGATGCACACAATAA	2833734
<u>LR778267.1</u>	42529115	GGATGCACACAATAA	42529101
<u>LR778267.1</u>	12460471	GGATGCACACAATAA	12460485
<u>LR778260.1</u>	16977017	CATTTGGGATGCACACAAT	16976999
<u>LR778260.1</u>	43845202	CATTTGGGATGCACACAAT	43845220
<u>LR778260.1</u>	18485245	CATTTCCGGATGCACACA	18485261
<u>LR778260.1</u>	13433975	ATTTCCGGATGCACAC	13433961
<u>LR738627.1</u>	9776345	CATTTCCGGATGCACACA	9776361
<u>LR738627.1</u>	43382214	GGATGCACACAATAA	43382200
<u>LR738607.1</u>	12873655	CATTTCCGGATGCACACA	12873671
<u>LR738607.1</u>	65672302	GGATGCACACAATAA	65672316
<u>CP035804.1</u>	68968159	GCATTTCCGGATGCACAC	68968175
<u>CP035804.1</u>	74077465	GGATGCACACAATAA	74077451
<u>XM 029729683.1</u>	3338	CATTTCCGGATGCACACA	3354
<u>XM 029729682.1</u>	3442	CATTTCCGGATGCACACA	3458
<u>XM 029729681.1</u>	3445	CATTTCCGGATGCACACA	3461
<u>LR584435.1</u>	4663289	CATTTAGGATGCACACAATAA	4663309
<u>LR584435.1</u>	56328727	CATTTGGGATGCACACAAT	56328709
<u>LR584435.1</u>	50726731	CATTTCCGGATGCACA	50726717
<u>LR584433.1</u>	2698528	CATTTAGGATGCACACAATAA	2698548
<u>LR584423.1</u>	23636461	CATTTCCGGATGCACACA	23636477
<u>LR131966.1</u>	1611448	TCGGATGCACACAATAA	1611464
<u>CP025673.1</u>	18984285	CATTTCCGGATGCACAGAATAA	18984305
<u>CP025673.1</u>	18456469	GCATTTCCGGATGCAC	18456483
<u>CP025673.1</u>	21343206	GCATTTCCGGATGCAC	21343220
<u>LS997565.1</u>	39271780	GCATTTCCGGATGCACAC	39271796
<u>LS997565.1</u>	16214664	CATTTCCGGATGCACA	16214650
<u>LT594791.1</u>	27812647	TTTCCGGATGCACACAAT	27812663
<u>LT594790.1</u>	5296331	TTCGGATGCACACAATA	5296347
<u>XM 014194909.1</u>	3809	CATTTCCGGATGCACACA	3825
<u>XM 014194899.1</u>	3914	CATTTCCGGATGCACACA	3930
<u>XM 014194890.1</u>	3915	CATTTCCGGATGCACACA	3931
<u>CP012024.1</u>	588716	TTCGGATGCACACAATA	588732
<u>LK066269.1</u>	301368	TCGGATGCACACAATAA	301384
<u>LK066269.1</u>	313490	TCGGATGCACACAATAA	313506
<u>LK066269.1</u>	325800	TCGGATGCACACAATAA	325816

<u>LN013793.1</u>	6993	CATTTCCGGATGCACACA	7009
<u>LR862152.1</u>	6253400	TTTCGGATGCACACAAT	6253416
<u>AP023178.1</u>	48147	GCATTTCCGGATGCACAC	48163
<u>LR812046.1</u>	20731743	TTTCGGATGCACACAAT	20731759
<u>HE601631.2</u>	58795411	CATTTTCGTATGCACACAAT	58795393
<u>HE601631.2</u>	55394958	TTTCGGATGCACACAAT	55394974
<u>HE601631.2</u>	17951684	CATTTCCGGATGCACAC	17951669
<u>LR812602.1</u>	20725446	TTTCGGATGCACACAAT	20725462
<u>LR812577.1</u>	21923491	TTTCGGATGCACACAAT	21923507
<u>LR812577.1</u>	7565339	TTCGGATGCACACAA	7565353
<u>LR812552.1</u>	20433661	TTTCGGATGCACACAAT	20433677
<u>LR812552.1</u>	5522106	TTCGGATGCACACAA	5522120
<u>LR812514.1</u>	16160608	CATTTCCGGATGCACACA	16160624
<u>AC136507.19</u>	74014	CATTTCCGGATGCACACA	74030
<u>CR391919.5</u>	156569	TTTCGGATGCACACAAT	156585
<u>AC006247.12</u>	139248	TTTCGGATGCACACAAT	139264
<u>XM 038023528.1</u>	299	TTCGGATGCACACAAT	284
<u>XM 038023527.1</u>	304	TTCGGATGCACACAAT	289
<u>XM 038023526.1</u>	299	TTCGGATGCACACAAT	284
<u>CP065149.1</u>	2963910	TTCGGATGCACACAAT	2963895
<u>CP065161.1</u>	1067558	CATTTCCGGATGCACAC	1067543
<u>OA565374.1</u>	4441	ATTTTCGAATGCACACAATAA	4422
<u>XM 037692747.1</u>	307	CATTTCCGGATGCACAC	292
<u>XM 037667276.1</u>	646	TTTCGGATGCACACAA	631
<u>XM 037611556.1</u>	314	TTTCGGATGCACACAA	299
<u>XM 037610901.1</u>	627	TCGGATGCACACAATA	612
<u>XM 038048380.1</u>	3624	TTCGGATGCACACAAT	3639
<u>XM 038048379.1</u>	3064	TTCGGATGCACACAAT	3079
<u>XM 038017770.1</u>	1589	TTTCGGATGCACACAA	1604
<u>OA884554.1</u>	13328	TTTCGGATGCACACAA	13343
<u>LR899427.1</u>	14480734	CATTTCCGGATGCACAC	14480749
<u>CP046680.1</u>	41617088	GGATGCACACAATAA	41617102
<u>CP046680.1</u>	41620099	GGATGCACACAATAA	41620113

SEQUENZA N° 6/2 (T=2; C=3; G=0; A=1)

Query: **GCATTTCCGGATGCACACAATAA**

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
.....

SEQUENZA N° 7/1

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

SEQUENZA N° 7/1 (T=2; C=1; G=0; A=3)

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° 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
Sequenza 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	Spirometra erinaceieuropaei	38.2	38.2	86%	2.2	100.00%	2171
LR792826.1	S_erinaceieuropaei, scaffold SPER_scaffold0150729	Spirometra erinaceieuropaei	38.2	38.2	86%	2.2	100.00%	2171
LR761611.1	Digitaria exilis annotation	Digitaria exilis	36.2	36.2	81%	8.7	100.00%	27476960
AP018441.1	Digitaria exilis genome assembly, chromosome: 4A	Digitaria exilis	36.2	36.2	81%	8.7	100.00%	27476960
AP018439.1	Undibacterium sp. YM2 DNA, complete genome	Undibacterium sp. YM2	36.2	180	81%	8.7	100.00%	6484812
CP023747.1	Undibacterium sp. KW1 DNA, complete genome	Undibacterium sp. KW1	36.2	108	81%	8.7	100.00%	6557365
CP023697.1	Streptomyces nodosus strain ATCC 14899 chromosome, complete genome	Streptomyces nodosus	36.2	98.6	100%	8.7	95.45%	7772587
CP036291.1	Streptomyces prasinus strain ATCC 13879 chromosome, complete genome	Streptomyces prasinus	36.2	36.2	81%	8.7	100.00%	7647592
XM_030095728.1	Planctomycetes bacterium Pla175 chromosome, complete genome	Planctomycetes bacterium Pla175	36.2	66.4	81%	8.7	100.00%	6618662
XM_030095727.1	PREDICTED: Salarias fasciatus septin-7-like (LOC115391457), transcript variant X4, mRNA	jewelled blenny	36.2	36.2	81%	8.7	100.00%	4270
	PREDICTED: Salarias fasciatus septin-7-like (LOC115391457), transcript variant X4, mRNA	jewelled blenny	36.2	36.2	81%	8.7	100.00%	4273

	transcript variant X3, mRNA								
XM_030095726.1	PREDICTED: <i>Salarias fasciatus</i> septin-7-like (LOC115391457), transcript variant X2, mRNA	jewelled blenny	36.2	36.2	81%	8.7	100.00%		4282
XM_030095725.1	PREDICTED: <i>Salarias fasciatus</i> septin-7-like (LOC115391457), transcript variant X1, mRNA	jewelled blenny	36.2	36.2	81%	8.7	100.00%		4285
XM_030095379.1	PREDICTED: <i>Salarias fasciatus</i> septin-7-like (LOC115391256), transcript variant X5, mRNA	jewelled blenny	36.2	36.2	81%	8.7	100.00%		1397
XM_030095378.1	PREDICTED: <i>Salarias fasciatus</i> septin-7-like (LOC115391256), transcript variant X4, mRNA	jewelled blenny	36.2	36.2	81%	8.7	100.00%		1653
XM_030095377.1	PREDICTED: <i>Salarias fasciatus</i> septin-7-like (LOC115391256), transcript variant X3, mRNA	jewelled blenny	36.2	36.2	81%	8.7	100.00%		1656
XM_030095376.1	PREDICTED: <i>Salarias fasciatus</i> septin-7-like (LOC115391256), transcript variant X2, mRNA	jewelled blenny	36.2	36.2	81%	8.7	100.00%		1666
XM_030095375.1	PREDICTED: <i>Salarias fasciatus</i> septin-7-like (LOC115391256), transcript variant X1, mRNA	jewelled blenny	36.2	36.2	81%	8.7	100.00%		1669
LR597442.1	<i>Salarias fasciatus</i> genome assembly, chromosome: 7	jewelled blenny	36.2	72.4	81%	8.7	100.00%		33379828
CP032382.1	<i>Chryseolinea soli</i> strain KIS68-18, complete genome	<i>Chryseolinea soli</i>	36.2	66.4	81%	8.7	100.00%		8241389
CP031455.1	<i>Streptomyces olivoreticuli</i> subsp. <i>olivoreticuli</i> strain ATCC 31159 chromosome, complete genome	<i>Streptomyces olivoreticuli</i> subsp. <i>olivoreticuli</i>	36.2	66.4	81%	8.7	100.00%		8809793
CP025803.1	<i>Sulfitobacter</i> sp. SK011 chromosome, complete genome	<i>Sulfitobacter</i> sp. SK011	36.2	36.2	81%	8.7	100.00%		4190786
CP030989.1	<i>Arachis hypogaea</i> cultivar Shitouqi chromosome A07	peanut	36.2	36.2	81%	8.7	100.00%		83117376
CP017316.1	<i>Streptomyces rubrolavendulae</i> strain MJM4426, complete genome	<i>Streptomyces rubrolavendulae</i>	36.2	159	81%	8.7	100.00%		6543262
CP009313.1	<i>Streptomyces nodosus</i> strain ATCC 14899 genome	<i>Streptomyces nodosus</i>	36.2	98.6	100%	8.7	95.45%		7714110
CP005286.1	<i>Corynebacterium humireducens</i> NBRC 106098 = DSM 45392, complete genome	<i>Corynebacterium humireducens</i> NBRC 106098 = DSM 45392	36.2	66.4	81%	8.7	100.00%		2681312
XM_008910664.1	<i>Phytophthora parasitica</i> INRA-310 hypothetical protein mRNA	<i>Phytophthora parasitica</i> INRA-310	36.2	36.2	81%	8.7	100.00%		6776
XM_003660782.1	<i>Thermothelomyces thermophilus</i> ATCC 42464 uncharacterized protein (MYCTH_97448), partial mRNA	<i>Thermothelomyces thermophilus</i> ATCC 42464	36.2	36.2	81%	8.7	100.00%		1335
XM_036686787.1	<i>Fusarium subglutinans</i> uncharacterized protein (FSUBG_8384), partial mRNA	<i>Fusarium subglutinans</i>	36.2	36.2	100%	8.7	95.45%		1008
CP048289.1	<i>Streptacidiphilus</i> sp. P02-A3a chromosome, complete genome	<i>Streptacidiphilus</i> sp. P02-A3a	36.2	161	81%	8.7	100.00%		9330052
CP058334.1	<i>Natronomonas</i> sp. C90 chromosome, complete genome	<i>Natronomonas</i> sp. C90	36.2	126	81%	8.7	100.00%		3211682
CP003002.1	<i>Myceliophthora thermophila</i> ATCC 42464 chromosome 1, complete sequence	<i>Thermothelomyces thermophilus</i> ATCC 42464	36.2	36.2	81%	8.7	100.00%		10931058
XM_002998468.1	<i>Phytophthora infestans</i> T30-4 conserved hypothetical protein (PITG_06383) mRNA, complete cds	<i>Phytophthora infestans</i> T30-4	36.2	36.2	81%	8.7	100.00%		2769
CP002049.1	<i>Truepera radiovictrix</i> DSM 17093, complete genome	<i>Truepera radiovictrix</i> DSM 17093	36.2	36.2	81%	8.7	100.00%		3260398
CP053450.1	<i>Planctomycetes</i> bacterium isolate PNC22 chromosome	<i>Planctomycetes</i> bacterium	34.2	34.2	95%	35	95.24%		4124050
CP054038.1	<i>Microbacterium hominis</i> strain PA2F3 chromosome, complete genome	<i>Microbacterium hominis</i>	34.2	64.4	77%	35	100.00%		3908277
AP019869.1	<i>Marchantia polymorpha</i> subsp. <i>ruderalis</i> Tak-1 DNA, chromosome: 4	<i>Marchantia polymorpha</i> subsp. <i>ruderalis</i>	34.2	34.2	77%	35	100.00%		26988051
AP019867.1	<i>Marchantia polymorpha</i> subsp. <i>ruderalis</i> Tak-1 DNA, chromosome: 2	<i>Marchantia polymorpha</i> subsp. <i>ruderalis</i>	34.2	34.2	77%	35	100.00%		29643427
CP053625.1	<i>Apiotrichum mycotoxinovorans</i> strain CICC 1454 chromosome	<i>Apiotrichum mycotoxinovorans</i>	34.2	34.2	77%	35	100.00%		1638155

CP053621.1	Apiotrichum mycotoxinovorans strain CICC 1454 chromosome 2	Apiotrichum mycotoxinovorans	34.2	34.2	77%	35	100.00%	10254415
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 gelatinovorans strain NEB572 chromosome, complete genome	Thalassobius gelatinovorans	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 arthritis strain AUSMDU00012717 chromosome, complete genome	Nocardia arthritis	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 mycotoxinovorans	34.2	34.2	77%	35	100.00%	1646260
CP049824.1	Apiotrichum mycotoxinovorans strain GMU1709 chromosome IV	Apiotrichum mycotoxinovorans	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	Pandoraea fibrosis	34.2	34.2	77%	35	100.00%	5592065

	genome								
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 (LOC118336560), mRNA	striped sea-bass	34.2	34.2	77%	35	100.00%	2394
CP058693.1	Streptomyces sp. Rer75 chromosome	Streptomyces sp. Rer75	34.2	34.2	77%	35	100.00%	10474412

VISTA DI ALLINEAMENTO VERTICALE AGGANCIATO A QUERY CON LETTERE D'IDENTITA'

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

<u>CP025803.1</u>	2236918	AAACGGCGTCGACACCTC	2236901
<u>CP030989.1</u>	67044159	AACGGCGTCGACACCTCC	67044176
<u>CP017316.1</u>	12679	AACGGCGTCGACACCTCC	12696
<u>CP017316.1</u>	2170361	CGGCGTCGACACCTCC	2170376
<u>CP017316.1</u>	241740	GGCGTCGACACCTCC	241754
<u>CP017316.1</u>	2341321	GGCGTCGACACCTCC	2341307
<u>CP017316.1</u>	3640586	GGCGTCGACACCTCC	3640572
<u>CP009313.1</u>	4252661	CACAAACGGCGGCGACACCTCC	4252682
<u>CP009313.1</u>	3995346	CGGCGTCGACACCTCC	3995361
<u>CP009313.1</u>	1646973	AACGGCGTCGACACC	1646987
<u>CP005286.1</u>	1169796	AACGGCGTCGACACCTCC	1169813
<u>CP005286.1</u>	2111537	AACGGCGTCGACACC	2111523
<u>XM 008910664.1</u>	1121	AACGGCGTCGACACCTCC	1138
<u>XM 003660782.1</u>	1057	AACGGCGTCGACACCTCC	1074
<u>XM 036686787.1</u>	798	CACAAACGACGTCGACACCTCC	819
<u>CP048289.1</u>	2823686	AACGGCGTCGACACCTCC	2823703
<u>CP048289.1</u>	5599762	AACGGCGTCGACACCT	5599747
<u>CP048289.1</u>	9091151	CGGCGTCGACACCTCC	9091136
<u>CP048289.1</u>	728680	GGCGTCGACACCTCC	728694
<u>CP048289.1</u>	2908602	AACGGCGTCGACACC	2908588
<u>CP058334.1</u>	918529	AACGGCGTCGACACCTCC	918512
<u>CP058334.1</u>	858919	AACGGCGTCGACACC	858933
<u>CP058334.1</u>	1177828	ACGGCGTCGACACCT	1177842
<u>CP058334.1</u>	1665270	CGGCGTCGACACCTC	1665284
<u>CP003002.1</u>	9999161	AACGGCGTCGACACCTCC	9999144
<u>XM 002998468.1</u>	1754	CACAAACGGCGTCGACAC	1737
<u>CP002049.1</u>	836152	AACGGCGTCGACACCTCC	836135
<u>CP053450.1</u>	3761022	CACAAACGGCGTCGCCACCTC	3761002
<u>CP054038.1</u>	2316437	ACGGCGTCGACACCTCC	2316421
<u>CP054038.1</u>	719854	ACGGCGTCGACACCT	719868
<u>AP019869.1</u>	17134560	ACGGCGTCGACACCTCC	17134576
<u>AP019867.1</u>	23100797	CACAAACGGCGTCGACA	23100781
<u>CP053625.1</u>	1429148	AACGGCGTCGACACCTC	1429132
<u>CP053621.1</u>	314245	ACGGCGTCGACACCTCC	314229
<u>CP053032.1</u>	3358899	ACGGCGTCGACACCTCC	3358915
<u>CP052860.1</u>	5581283	AACGGCGTCGACACCTC	5581267
<u>CP052860.1</u>	6399354	CACAAACGGCGTCGA	6399368
<u>XM 033681904.1</u>	950	AACGGCGTCGACACCTC	934
<u>CP051181.1</u>	1971308	CACAAACGGCGTCGACA	1971292
<u>CP051010.1</u>	7300134	ACGGCGTCGACACCTCC	7300150
<u>CP051010.1</u>	6353150	CGGCGTCGACACCTCC	6353165
<u>CP051010.1</u>	2899305	AACGGCGTCGACACC	2899319
<u>CP051010.1</u>	9622143	AACGGCGTCGACACC	9622129
<u>CP051010.1</u>	10787014	AACGGCGTCGACACC	10787028
<u>CP050974.1</u>	7139722	ACGGCGTCGACACCTCC	7139738
<u>CP050974.1</u>	6222402	CGGCGTCGACACCTCC	6222417
<u>CP050974.1</u>	2945818	AACGGCGTCGACACC	2945832
<u>CP050974.1</u>	9426651	AACGGCGTCGACACC	9426637
<u>CP050974.1</u>	10760095	AACGGCGTCGACACC	10760109
<u>CP050975.1</u>	5759656	ACGGCGTCGACACCTCC	5759672
<u>CP050975.1</u>	3089795	CGGCGTCGACACCTCC	3089780
<u>CP050975.1</u>	5005684	CGGCGTCGACACCTCC	5005699
<u>CP050975.1</u>	2084240	AACGGCGTCGACACC	2084254
<u>CP046172.1</u>	8207102	ACGGCGTCGACACCTCC	8207086
<u>XM 032945539.1</u>	1495	ACGGCGTCGACACCTCC	1511
<u>CP049825.1</u>	207129	AACGGCGTCGACACCTC	207145
<u>CP049824.1</u>	9970787	ACGGCGTCGACACCTCC	9970803
<u>XM 032809184.1</u>	2446	ACAAACGGCGTCGACAC	2430

<u>XM 032809183.1</u>	2450	ACAAACGGCGTCGACAC	2434
<u>CP049199.1</u>	2654895	AACGGCGTCGACACCTC	2654911
<u>CP049199.1</u>	3239860	CGGCGTCGACACCTC	3239846
<u>CP041650.2</u>	7554117	ACGGCGTCGACACCTCC	7554133
<u>CP041650.2</u>	1161359	CGGCGTCGACACCTC	1161373
<u>CP041650.2</u>	2673368	AACGGCGTCGACACC	2673382
<u>CP041650.2</u>	3305153	AACGGCGTCGACACC	3305167
<u>CP041650.2</u>	9576160	AACGGCGTCGACACC	9576146
<u>CP041650.2</u>	11021158	AACGGCGTCGACACC	11021172
<u>CP048397.1</u>	4900581	ACGGCGTCGACACCTCC	4900597
<u>CP048397.1</u>	1645955	AACGGCGTCGACACCT	1645940
<u>CP048397.1</u>	1267077	AACGGCGTCGACACC	1267091
<u>CP041612.2</u>	7317910	ACGGCGTCGACACCTCC	7317926
<u>CP041612.2</u>	858628	CGGCGTCGACACCTC	858642
<u>CP041612.2</u>	2496028	AACGGCGTCGACACC	2496042
<u>CP041612.2</u>	3098102	AACGGCGTCGACACC	3098116
<u>CP041612.2</u>	9486014	AACGGCGTCGACACC	9486000
<u>CP041612.2</u>	10881723	AACGGCGTCGACACC	10881737
<u>CP041607.2</u>	7106408	ACGGCGTCGACACCTCC	7106424
<u>CP041607.2</u>	782328	CGGCGTCGACACCTC	782342
<u>CP041607.2</u>	2292309	AACGGCGTCGACACC	2292323
<u>CP041607.2</u>	2924127	AACGGCGTCGACACC	2924141
<u>CP041607.2</u>	9147840	AACGGCGTCGACACC	9147826
<u>CP041607.2</u>	10593155	AACGGCGTCGACACC	10593169
<u>CP041609.2</u>	7199676	ACGGCGTCGACACCTCC	7199692
<u>CP041609.2</u>	784863	CGGCGTCGACACCTC	784877
<u>CP041609.2</u>	2298219	AACGGCGTCGACACC	2298233
<u>CP041609.2</u>	2930021	AACGGCGTCGACACC	2930035
<u>CP041609.2</u>	9277199	AACGGCGTCGACACC	9277185
<u>CP041609.2</u>	10723930	AACGGCGTCGACACC	10723944
<u>CP041613.2</u>	7552518	ACGGCGTCGACACCTCC	7552534
<u>CP041613.2</u>	1277247	CGGCGTCGACACCTC	1277261
<u>CP041613.2</u>	1314664	GGCGTCGACACCTCC	1314650
<u>CP041613.2</u>	2673915	AACGGCGTCGACACC	2673929
<u>CP041613.2</u>	3279128	AACGGCGTCGACACC	3279142
<u>CP041613.2</u>	9933257	AACGGCGTCGACACC	9933243
<u>CP041613.2</u>	11319675	AACGGCGTCGACACC	11319689
<u>CP041604.2</u>	7471604	ACGGCGTCGACACCTCC	7471620
<u>CP041604.2</u>	1103167	CGGCGTCGACACCTC	1103181
<u>CP041604.2</u>	1122128	CGGCGTCGACACCTC	1122114
<u>CP041604.2</u>	1246344	CGGCGTCGACACCTC	1246358
<u>CP041604.2</u>	2626808	AACGGCGTCGACACC	2626822
<u>CP041604.2</u>	3229941	AACGGCGTCGACACC	3229955
<u>CP041604.2</u>	9585963	AACGGCGTCGACACC	9585949
<u>CP041604.2</u>	10977721	AACGGCGTCGACACC	10977735
<u>CP041602.2</u>	7193261	ACGGCGTCGACACCTCC	7193277
<u>CP041602.2</u>	877655	CGGCGTCGACACCTC	877669
<u>CP041602.2</u>	915072	GGCGTCGACACCTCC	915058
<u>CP041602.2</u>	2332815	AACGGCGTCGACACC	2332829
<u>CP041602.2</u>	2938025	AACGGCGTCGACACC	2938039
<u>CP041602.2</u>	9589792	AACGGCGTCGACACC	9589778
<u>CP041602.2</u>	10981344	AACGGCGTCGACACC	10981358
<u>CP041610.2</u>	7104883	ACGGCGTCGACACCTCC	7104899
<u>CP041610.2</u>	809048	CGGCGTCGACACCTC	809062
<u>CP041610.2</u>	2275886	AACGGCGTCGACACC	2275900
<u>CP041610.2</u>	2880087	AACGGCGTCGACACC	2880101
<u>CP041610.2</u>	9351182	AACGGCGTCGACACC	9351168
<u>CP041610.2</u>	10818242	AACGGCGTCGACACC	10818256

<u>AP022600.1</u>	6097870	ACGGCGTCGACACCTCC	6097854
<u>AP022600.1</u>	57054	CGGCGTCGACACCTCC	57039
<u>AP022600.1</u>	368925	CGGCGTCGACACCTCC	368940
<u>AP022561.1</u>	5804988	ACGGCGTCGACACCTCC	5804972
<u>AP022561.1</u>	4225043	CGGCGTCGACACCTC	4225057
<u>AP022561.1</u>	4288025	AACGGCGTCGACACC	4288039
<u>CP047385.1</u>	4262943	CAAACGGCGTCGACACC	4262927
<u>CP047386.1</u>	2084339	CAAACGGCGTCGACACC	2084355
<u>CP047183.1</u>	1084993	ACGGCGTCGACACCTCC	1085009
<u>CP047183.1</u>	236126	AACGGCGTCGACACC	236112
<u>CP043930.1</u>	7805105	AAACGGCGTCGACACCT	7805089
<u>XM 031649147.1</u>	331	ACGGCGTCGACACCTCC	347
<u>CP045643.1</u>	3681359	ACGGCGTCGACACCTCC	3681343
<u>CP045643.1</u>	4463765	CGGCGTCGACACCTCC	4463750
<u>CP045643.1</u>	7339390	AACGGCGTCGACACC	7339376
<u>XM 031503155.1</u>	5971	AAACGGCGTCGACACCT	5955
<u>XM 031478322.1</u>	7684	AAACGGCGTCGACACCT	7668
<u>CP023690.1</u>	3643320	ACGGCGTCGACACCTCC	3643336
<u>CP023690.1</u>	2238430	AACGGCGTCGACACC	2238444
<u>CP029194.1</u>	6347693	ACGGCGTCGACACCTCC	6347709
<u>CP029194.1</u>	5353612	CGGCGTCGACACCTCC	5353627
<u>CP043534.1</u>	1232091	ACGGCGTCGACACCTCC	1232075
<u>CP043448.1</u>	102288	ACGGCGTCGACACCTCC	102304
<u>XM 030681488.1</u>	153	ACGGCGTCGACACCTCC	169
<u>XM 030681487.1</u>	153	ACGGCGTCGACACCTCC	169
<u>XM 030588374.1</u>	154	ACGGCGTCGACACCTCC	170
<u>XM 030588373.1</u>	154	ACGGCGTCGACACCTCC	170
<u>XM 030588372.1</u>	154	ACGGCGTCGACACCTCC	170
<u>CP032580.1</u>	54132560	CACAAACGGCGTCGACA	54132576
<u>CP032571.1</u>	204763	ACGGCGTCGACACCTCC	204779
<u>XM 037413552.1</u>	19	ACGGCGTCGACACCTCC	3
<u>CP063451.1</u>	27415	CAAACGGCGTCGACACC	27399
<u>AP024094.1</u>	2513669	CAAACGGCGTCGACACC	2513653
<u>AP024094.1</u>	318121	ACGGCGTCGACACCT	318107
<u>CP061167.1</u>	110524	ACGGCGTCGACACCTCC	110540
<u>CP060496.1</u>	3316655	ACGGCGTCGACACCTCC	3316671
<u>CP048835.1</u>	2484608	AACGGCGTCGACACCTC	2484624
<u>CP048835.1</u>	2633466	AACGGCGTCGACACC	2633480
<u>CP041141.1</u>	111235	ACGGCGTCGACACCTCC	111251
<u>CP059696.1</u>	1534571	AACGGCGTCGACACCTC	1534587
<u>CP059696.1</u>	3230849	CGGCGTCGACACCTCC	3230834
<u>CP059696.1</u>	637027	ACGGCGTCGACACCT	637013
<u>CP059696.1</u>	1664580	AACGGCGTCGACACC	1664594
<u>XM 035673227.1</u>	851	ACGGCGTCGACACCTCC	835
<u>CP058693.1</u>	9175631	ACGGCGTCGACACCTCC	9175615

PRODOTTI DEGLI ALLINEAMENTI SIGNIFICATIVI (CDS FEATURE)

- **RecA-like DNA recombinase** [GenBank](#)
- **Hypothetical protein** [GenBank](#); [GenBank](#); [GenBank](#); [GenBank](#); [GenBank](#); [GenBank](#); [GenBank](#); [GenBank](#); [GenBank](#); [GenBank](#); [GenBank](#); [GenBank](#); [GenBank](#); [GenBank](#); [GenBank](#); [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](#); [GenBank](#); [GenBank](#); [GenBank](#); [GenBank](#); [GenBank](#); [GenBank](#); [GenBank](#); [GenBank](#); [GenBank](#); [GenBank](#); [GenBank](#); [GenBank](#); [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](#); [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](#)
- **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](#)
- **Sestrin-1-like** [GenBank](#)
- **Alpha-mannosidase 2** [GenBank](#); [GenBank](#)
- **TIGR03620 family F420-dependent LLM class oxidoreductase** [GenBank](#); [GenBank](#); [GenBank](#); [GenBank](#); [GenBank](#); [GenBank](#)
- **Esterase-like activity of phytase family protein** [GenBank](#); [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](#)

SEQUENZA N° 7/2

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

SEQUENZA N° 7/2 (T=2; C=1; G=0; A=3)

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° 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
Sequenza 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: **CTCTTTCG GAGACGCGATTATT**

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	wheat stem sawfly	38.2	38.2	86%	2.2	100.00%	18282

	uncharacterized LOC107271692 (LOC107271692), transcript variant X19, mRNA PREDICTED: Cephus cinctus uncharacterized LOC107271692 (LOC107271692), transcript variant X18, mRNA	wheat stem sawfly	38.2	38.2	86%	2.2	100.00%	18114
XM_025088961.1								
	uncharacterized LOC107271692 (LOC107271692), transcript variant X17, mRNA PREDICTED: Cephus cinctus uncharacterized LOC107271692 (LOC107271692), transcript variant X16, mRNA	wheat stem sawfly	38.2	38.2	86%	2.2	100.00%	19334
XM_015747960.2								
	uncharacterized LOC107271692 (LOC107271692), transcript variant X15, mRNA PREDICTED: Cephus cinctus uncharacterized LOC107271692 (LOC107271692), transcript variant X14, mRNA	wheat stem sawfly	38.2	38.2	86%	2.2	100.00%	16569
XM_025088960.1								
	uncharacterized LOC107271692 (LOC107271692), transcript variant X13, mRNA PREDICTED: Cephus cinctus uncharacterized LOC107271692 (LOC107271692), transcript variant X12, mRNA	wheat stem sawfly	38.2	38.2	86%	2.2	100.00%	16572
XM_025088959.1								
	uncharacterized LOC107271692 (LOC107271692), transcript variant X11, mRNA PREDICTED: Cephus cinctus uncharacterized LOC107271692 (LOC107271692), transcript variant X10, mRNA	wheat stem sawfly	38.2	38.2	86%	2.2	100.00%	15997
XM_025088958.1								
	uncharacterized LOC107271692 (LOC107271692), transcript variant X9, mRNA PREDICTED: Cephus cinctus uncharacterized LOC107271692 (LOC107271692), transcript variant X8, mRNA	wheat stem sawfly	38.2	38.2	86%	2.2	100.00%	17919
XM_015747954.2								
	uncharacterized LOC107271692 (LOC107271692), transcript variant X7, mRNA PREDICTED: Cephus cinctus uncharacterized LOC107271692 (LOC107271692), transcript variant X6, mRNA	wheat stem sawfly	38.2	38.2	86%	2.2	100.00%	17153
XM_025088957.1								
	uncharacterized LOC107271692 (LOC107271692), transcript variant X5, mRNA PREDICTED: Cephus cinctus uncharacterized LOC107271692 (LOC107271692), transcript variant X4, mRNA	wheat stem sawfly	38.2	38.2	86%	2.2	100.00%	19318
XM_025088956.1								
	uncharacterized LOC107271692 (LOC107271692), transcript variant X3, mRNA PREDICTED: Cephus cinctus uncharacterized LOC107271692 (LOC107271692), transcript variant X2, mRNA	wheat stem sawfly	38.2	38.2	86%	2.2	100.00%	17737
XM_015747952.2								
	uncharacterized LOC107271692 (LOC107271692), transcript variant X1, mRNA PREDICTED: Cephus cinctus uncharacterized LOC107271692 (LOC107271692), transcript variant X1, mRNA	wheat stem sawfly	38.2	38.2	86%	2.2	100.00%	19324
XM_025088955.1								
	complete genome Scleropages formosus genome assembly, chromosome: 6	wheat stem sawfly	38.2	38.2	86%	2.2	100.00%	18629
XM_025088954.1								
	Darwinula stevensoni	wheat stem sawfly	38.2	38.2	86%	2.2	100.00%	18643
XM_025088953.1								
		wheat stem sawfly	38.2	38.2	86%	2.2	100.00%	18648
XM_025088952.1								
		wheat stem sawfly	38.2	38.2	86%	2.2	100.00%	18291
XM_025088951.1								
		wheat stem sawfly	38.2	38.2	86%	2.2	100.00%	16623
XM_025088950.1								
		wheat stem sawfly	38.2	38.2	86%	2.2	100.00%	16889
XM_025088949.1								
		wheat stem sawfly	38.2	38.2	86%	2.2	100.00%	16641
XM_025088948.1								
		wheat stem sawfly	38.2	38.2	86%	2.2	100.00%	16647
XM_025088947.1								
	complete genome Scleropages formosus genome assembly, chromosome: 6	Paenibacillus sp. Y412MC10	38.2	38.2	86%	2.2	100.00%	7121665
CP001793.1								
		Asian bonytongue	36.2	66.4	86%	8.9	100.00%	38292110
LR584071.1								
		Darwinula stevensoni	36.2	72.4	81%	8.9	100.00%	32304
LR902823.1								

	PREDICTED: <i>Belonocnema treatae</i> uncharacterized LOC117174255 (LOC117174255), mRNA	<i>Belonocnema treatae</i>	36.2	36.2	81%	8.9	100.00%	3089
XM_033363171.1	<i>Pseudomonas fluorescens</i> strain NCTC9428 genome assembly, chromosome: 1	<i>Pseudomonas fluorescens</i>	36.2	36.2	81%	8.9	100.00%	6033966
LR134318.1 1/2 5/1 7/1 9/1	<i>Bos mutus</i> isolate yakQH1 chromosome 14	wild yak	36.2	36.2	81%	8.9	100.00%	81354091
CP027082.1 1/2 3/6 10/1	<i>Cryptococcus amyloletus</i> CBS 6039 chitin synthase 1 (L202_04556), mRNA	<i>Cryptococcus amyloletus</i> CBS 6039	36.2	36.2	81%	8.9	100.00%	3964
XM_019138661.1	PREDICTED: <i>Bombus impatiens</i> uncharacterized LOC100748080 (LOC100748080), transcript variant X10, mRNA	common eastern bumble bee	34.2	34.2	77%	35	100.00%	6059
XM_033324960.1	PREDICTED: <i>Bombus impatiens</i> uncharacterized LOC100748080 (LOC100748080), transcript variant X9, mRNA	common eastern bumble bee	34.2	34.2	77%	35	100.00%	6140
XM_033324959.1	PREDICTED: <i>Bombus impatiens</i> uncharacterized LOC100748080 (LOC100748080), transcript variant X8, mRNA	common eastern bumble bee	34.2	34.2	77%	35	100.00%	6143
XM_024372643.2	PREDICTED: <i>Bombus impatiens</i> uncharacterized LOC100748080 (LOC100748080), transcript variant X7, mRNA	common eastern bumble bee	34.2	34.2	77%	35	100.00%	6616
XM_033324958.1	PREDICTED: <i>Bombus impatiens</i> uncharacterized LOC100748080 (LOC100748080), transcript variant X6, mRNA	common eastern bumble bee	34.2	34.2	77%	35	100.00%	6111
XM_033324957.1	PREDICTED: <i>Bombus impatiens</i> uncharacterized LOC100748080 (LOC100748080), transcript variant X5, mRNA	common eastern bumble bee	34.2	34.2	77%	35	100.00%	6619
XM_024372641.2	PREDICTED: <i>Bombus impatiens</i> uncharacterized LOC100748080 (LOC100748080), transcript variant X4, mRNA	common eastern bumble bee	34.2	34.2	77%	35	100.00%	6195
XM_012394472.3	PREDICTED: <i>Bombus impatiens</i> uncharacterized LOC100748080 (LOC100748080), transcript variant X3, mRNA	common eastern bumble bee	34.2	34.2	77%	35	100.00%	7680
XM_024372640.2	PREDICTED: <i>Bombus impatiens</i> uncharacterized LOC100748080 (LOC100748080), transcript variant X2, mRNA	common eastern bumble bee	34.2	34.2	77%	35	100.00%	6671
XM_012394470.3	PREDICTED: <i>Bombus impatiens</i> uncharacterized LOC100748080 (LOC100748080), transcript variant X1, mRNA	common eastern bumble bee	34.2	34.2	77%	35	100.00%	7684
XM_012394469.3	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
MG592473.1	<i>Calothrix brevissima</i> NIES-22 DNA, nearly complete genome	<i>Calothrix brevissima</i> NIES-22	34.2	34.2	77%	35	100.00%	8527291
AP018207.1	<i>Pseudomonas</i> sp. Z003-0.4C(8344-21) genome assembly, chromosome: I	<i>Pseudomonas</i> sp. Z003-0.4C(8344-21)	34.2	34.2	77%	35	100.00%	6113481
LT629756.1 1/2 5/1 7/1 9/1	<i>Sphingobium baderi</i> strain DE-13, complete genome	<i>Sphingobium baderi</i>	34.2	34.2	77%	35	100.00%	4107398
CP013264.1	PREDICTED: <i>Orussus abietinus</i> zinc finger protein 236 (LOC105698248), mRNA	<i>Orussus abietinus</i>	34.2	34.2	77%	35	100.00%	1906
XM_012422314.1	<i>Agrobacterium vitis</i> VAR06-30 DNA, chromosome 2, complete sequence	<i>Agrobacterium vitis</i>	34.2	34.2	77%	35	100.00%	1162882
AP023273.1	PREDICTED: <i>Odontomachus brunneus</i> beta-galactosidase (LOC116846673), transcript variant X3, mRNA	<i>Odontomachus brunneus</i>	34.2	34.2	77%	35	100.00%	2422
XM_032820819.1	PREDICTED: <i>Odontomachus brunneus</i> beta-galactosidase (LOC116846673), transcript variant X2, mRNA	<i>Odontomachus brunneus</i>	34.2	34.2	77%	35	100.00%	2161
XM_032820818.1	PREDICTED: <i>Odontomachus brunneus</i> beta-galactosidase (LOC116846673), transcript	<i>Odontomachus brunneus</i>	34.2	34.2	77%	35	100.00%	2501
XM_032820817.1								

	variant X1, mRNA									
	PREDICTED: <i>Odontomachus brunneus</i> uncharacterized LOC116840469 (LOC116840469), transcript variant X2, mRNA	<i>Odontomachus brunneus</i>	34.2	34.2	95%	35	95.24%		7299	
XM_032807220.1	PREDICTED: <i>Odontomachus brunneus</i> uncharacterized LOC116840469 (LOC116840469), transcript variant X1, mRNA	<i>Odontomachus brunneus</i>	34.2	34.2	95%	35	95.24%		7326	
XM_032807219.1	<i>Deinococcus</i> sp. AJ005 plasmid p380k, complete sequence	<i>Deinococcus</i> sp. AJ005	34.2	34.2	77%	35	100.00%		380021	
CP044989.1	<i>Aquisphaera giovannonii</i> strain OJF2 chromosome, complete genome	<i>Aquisphaera giovannonii</i>	34.2	34.2	77%	35	100.00%		10370956	
CP042997.1	<i>Takifugu rubripes</i> genome assembly, chromosome: 11	torafugu	34.2	34.2	77%	35	100.00%		16367601	
LR584241.1	<i>Sphingobium yanoikuyae</i> strain S72 chromosome, complete genome	<i>Sphingobium yanoikuyae</i>	34.2	34.2	77%	35	100.00%		5532633	
CP023741.1	<i>Bathycoccus prasinus</i> predicted protein (Bathy05g03570), partial mRNA	<i>Bathycoccus prasinus</i>	34.2	34.2	77%	35	100.00%		2850	
XM_007513213.1	<i>Bathycoccus prasinus</i> genomic : chromosome_5	<i>Bathycoccus prasinus</i>	34.2	34.2	77%	35	100.00%		1019276	
FO082274.1	<i>Caenorhabditis remanei</i> hypothetical protein (CRE_26188) mRNA, complete cds	<i>Caenorhabditis remanei</i>	34.2	34.2	77%	35	100.00%		4470	
XM_003113858.1	<i>Candidatus Nitrospira defluvii</i> chromosome, complete genome	<i>Nitrospira defluvii</i>	34.2	34.2	77%	35	100.00%		4317083	
FP929003.1	<i>Agrobacterium vitis</i> S4 chromosome 2, complete sequence	<i>Agrobacterium vitis</i> S4	34.2	34.2	77%	35	100.00%		1283187	
CP000634.1	<i>Notodromas monacha</i> PREDICTED: <i>Pygocentrus nattereri</i> glutamate receptor, ionotropic, N-methyl D-aspartate 2A, a (grin2aa), mRNA	<i>Notodromas monacha</i>	32.2	32.2	72%	138	100.00%		85654	
OA882732.1	PREDICTED: <i>Syngnathus acus</i> heat shock protein b8 (hspb8), transcript variant X2, mRNA	red-bellied piranha	32.2	32.2	72%	138	100.00%		7232	
XM_037544960.1	PREDICTED: <i>Syngnathus acus</i> heat shock protein b8 (hspb8), transcript variant X1, mRNA	greater pipefish	32.2	32.2	72%	138	100.00%		853	
XM_037258044.1	<i>Macrobrachium nipponense</i> isolate FS-2020 chromosome 18	greater pipefish	32.2	32.2	72%	138	100.00%		980	
XM_037258043.1	<i>Macrobrachium nipponense</i> isolate FS-2020 chromosome 31	<i>Macrobrachium nipponense</i>	32.2	32.2	72%	138	100.00%		90060326	
CP062022.1	<i>Leptospira interrogans</i> serovar Bataviae strain 1489 chromosome 1, complete sequence	<i>Leptospira interrogans</i> serovar Bataviae	32.2	32.2	72%	138	100.00%		73203386	
CP062049.1	<i>Leptospira interrogans</i> serovar Icterohaemorrhagiae strain 898 chromosome 1, complete sequence	<i>Leptospira interrogans</i> serovar Icterohaemorrhagiae	32.2	32.2	72%	138	100.00%		4431091	
CP043893.1	<i>Leptospira interrogans</i> serovar Canicola strain 782 chromosome 1, complete sequence	<i>Leptospira interrogans</i> serovar Canicola	32.2	32.2	72%	138	100.00%		4280412	
CP043891.1	<i>Leptospira interrogans</i> serovar Bataviae strain 1548 chromosome, complete genome	<i>Leptospira interrogans</i> serovar Bataviae	32.2	32.2	72%	138	100.00%		4556406	
CP043884.1	<i>Leptospira interrogans</i> serovar Icterohaemorrhagiae strain Langkawi chromosome 1, complete sequence	<i>Leptospira interrogans</i> serovar Icterohaemorrhagiae	32.2	32.2	72%	138	100.00%		4354848	
CP043876.1	<i>Carposina sasakii</i> isolate BJYQ chromosome 17	<i>Carposina sasakii</i>	32.2	32.2	72%	138	100.00%		4367462	
CP053164.1	<i>Saccharopolyspora spinosa</i> strain CCTCC M206084 chromosome, complete genome	<i>Saccharopolyspora spinosa</i>	32.2	32.2	72%	138	100.00%		43742407	
CP061007.1	<i>Pyrenophora teres</i> f. <i>maculata</i> isolate P-A14 chromosome 6	<i>Pyrenophora teres</i> f. <i>maculata</i>	32.2	32.2	72%	138	100.00%		8876435	
CP060577.1	<i>Pyrenophora teres</i> f. <i>maculata</i> isolate FG0B10Ptm-1 chromosome 6	<i>Pyrenophora teres</i> f. <i>maculata</i>	32.2	32.2	72%	138	100.00%		2338523	
CP060561.1	<i>Pyrenophora teres</i> f. <i>maculata</i> isolate NZKF2 chromosome 6	<i>Pyrenophora teres</i> f. <i>maculata</i>	32.2	32.2	72%	138	100.00%		2295326	
CP060552.1	<i>Pyrenophora teres</i> f. <i>maculata</i> isolate DEN2.6 chromosome 6	<i>Pyrenophora teres</i> f. <i>maculata</i>	32.2	32.2	72%	138	100.00%		2316045	
CP060540.1	<i>Acomys russatus</i>	golden spiny mouse	32.2	32.2	72%	138	100.00%		2451318	
LR877232.1									61132863	

**genome assembly,
chromosome: 21**

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	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	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	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	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	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	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	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="MRMAVGSVKMQPPPCESPALARAVAAAAADSALRRSPSAREPEPEPEQPPAPPRPRLRDLPALLRSGTLRRKRKSVGGGRTLQRRISNPYLEHTSSQIYGENSSCAGRALRNIIIVQAADLIKDRVNLKGFYRRSCVSELVDWLLHCPFIQCRSTAI
GVWQLLLDMGILSSVDQHLYFQDTYVFYQFSSDECSYLYCEFEREEWQNGVKMLLQL
VPVIRTRAGLCELSHRKIEDSEESSDEILARLTSAVQRELAAVIALKAKKSAIEQDEE
SSDKHIPTAEADSDPDPQAGVMCKLQGREDIGRIELVQKLARENQCFLQTDTRKEQDKS
EQQDGEVTTVREQEQSVLVLRKVQRRGPAPPGGSADSDWRYVVVSGTPEKILEHLLND
LHLEEVQDKETETLLDDFLLTYTFVMTTDDLQALLRHYSAKKYQGKEENSVDVPRKR
KVLHLVLSQWISLYKDWLHEDEHSMFLKTIYRNVLDDVYEYPILEKELKEFQKILGMH
RRHTVDEYSPQRKNKALFHQFSLKENWLQHRGAVTETEEIFCHVYITEHSYVSVKAKV
SSTAQEVLRVVAEKIQHAEDLALVAVSFSGGKRELHPNDLAISKSLASGRIYVYRK
NLADTLNPFANEESQQRPLRILGMNTWDLALELMNFDWLSLFNSIHEQELIYFTFSRQ
GSGEHTANLSLLLQRCNEVQLWVATEILLCSQLGKRVQLVKKFIKIAAHCKAQRNLNS
FFAIVMGLNTASVSRLSQTWEKIPGKFKKLFSELESITDPSLNHKAYRDTFKMKPKPK
IPFMPLLLKDVTFIHEGNKTFLDNLVNFELHMIADTVRTLHRANQFGGDASPKHEH
HELKSYVTHLLVIDSQALFELSHRIEPR"

ORIGIN

```
1 cgctgccggg agaagccacg ggagctgagg ggcgcgcggc cgccgccggg gcatgaggat
61 ggccgtgggc tccgtcaaga tgcagccgcc gccgtgagag agcccgccgc tggccccggc
121 ggtggcggcg gcggcggcgg acagcgcctt ggcgcgcagc cccagcgcctt gcgagcccga
181 gcccagacc gagcagccgc ccgccccgcc ggcgcgcggg ctgcgggacc tgcgggcgct
241 gctgcggagc gggctcacgc tgcggaggaa ggcgagcgtc ggcggcggcg gccggactct
301 ccagagaagg atttccaatc catacctcga gcacacttct tcccagattt atggggagaa
```

361 ttcttcttgt gcaggaagag cactgaggaa tatcattatc gtccaggcag ctgatctgat
421 caaggacaga gtgaacctca agggatttta caggcgaagc tgtgttgat cagaactggt
481 cgactggctt ctggaacct 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 agacaggag agcagtgaca agcacatacc
901 cacggcagaa gctgacagtg atccagatcc tcaggcaggg gtgatgtgca agcttcaggg
961 aagagaggac atcgggcgca ttgagctggt ccagaagctg gcgcgggaga actgtcagtt
1021 tttgcagacg gacagaaaag aacaggacaa atctgaacag caggatggtg aggtgaccac
1081 agtcaggag caggagcaga gcgtcctggt gctgaggaag gtgcagcgca gaggcccggc
1141 cccccagga gggagcgtg acagcgattg gagatacgtg gtcgtgtccg ggacccccga
1201 gaagatcttg gagcatcttt tgaatgactt gcacctgga gaagtccagg acaaagaaac
1261 agagaccctc ctggatgact tcctcctcac ctacaccgtc ttcattgacca cggacgactt
1321 gtgccaggcg ctgctgagac actattctgc caagaagtat caaggcaaag aagaaaactc
1381 agatgttctt cgtcggaaac gcaaagtctt gcatcttgtc tcccagtgga tttctctgta
1441 caaagactgg ctacacgaag acgagcactc caaaatgtt ttaaagacca tatacaggaa
1501 cgttctggat gatgtctatg aataccaat acttgaaaaa gaattgaagg aatttcaaaa
1561 gatacttggg atgcaccgtc gtcacaccgt tgatgagtag tctccacaga ggaagaataa
1621 ggcccttttc caccaattca gtcttaagga gaactggctc cagcaccgag gcgctgtgac
1681 tgaacagag gaaatTTTTT gccatgtgta catcacggag cactcgtacg tcagcgtgaa
1741 ggcgaaagtt tccagtacgg ctcaggaggt cctaagagtc gtggcggaga agatccagca
1801 cgcggaagag gacctggccc tgggtggcgt ctcttctct gggggaaagc gtgaacttca
1861 tcccaatgat ctagecatct ccaaaccctt ggaagcatct ggtcggatat atgtctatcg
1921 gaaaaacctg gctgacactt tgaacccttt tgccgaaaac gaggagtccc agcaaaggcc
1981 gttgaggatc ttgggaatga aacttggga tcttgcctg gaattaatga attttgattg
2041 gagtctcttc aattcgattc acgagcaaga gctgatctac ttcacattca gcagacaagg
2101 cagtggggag cacaccgca acctcagcct cctgctccag cgctgcaatg aggtccagct
2161 ctgggtggcc acggagattc tgctctgcag ccagctgggc aagcgggtgc agctggtgaa
2221 gaagttcatc aagatcgccg cccactgcaa agcccagagg aacctgaact ctttctttgc
2281 tatcgtgatg ggtctcaaca cggcttctgt cagccgctg tcgcaaacct gggagaaaat
2341 ccctgggaaa ttaagaaac ttttctctga acttgaaagc ttaacagacc cttccctgaa
2401 ccacaaggcc tacagagata cttcaaaaa gatgaagccg cccaagatcc cttcatgcc
2461 cttactgctc aaagatgtga catttattca cgagggaaac aaaacatttt tggataatct
2521 cgtcaatttt gagaagctgc atatgattgc agacaccgtc cgaaccctga ggcaactgcag
2581 ggccaaccag tttggagggg acgctctcc gaaagagcac cacgagctga agtcctacgt
2641 cactcacctg ctcgatcag acagccagca ggccctgtt 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
EPEQPPAPPRPRLRDLPALLRSGTLRRKRKSVGGGRTLQRRISNPYLEHTSSQIYGE
NSSCAGRALRNIIIVQAADLIKDRVNLKGFYRRSCVSELVDWLLHCPFIQCRSTAI
GVWQLLLDMGILSSVDQHLYFQDTYVFYQFSSDECSYLYCEFEREEWQNGVKMLLQL
VPVIRTRAGLCELSHRKIEDSEESSDEILARLTSAVQRELAAVIALKAKKSAIEQDEE
SSDKHIPTAEADSDPDPQAGVMCKLQGREDIGRIELVQKLARENQCFLQTDNRKEQDKS
EQQDGEVTTVREQEQSVLVLRKVQRRGPAPPGGSADSDWRYVVVSGTPEKILEHLLND
LHLEEVQDKETETLLDDFLLTYTFVMTTDDLQALLRHYSAKKYQGKEENSVDVPRRKR
KVLHLVLSQWISLYKDWLHEDEHSMFLKTIYRNVLDDVYEYPILEKELKEFQKILGMH
RRHTVDEYSPQRKNKALFHQFSLKENWLQHRGAVTETEEIFCHVYITEHSYVSVKAKV
SSTAQEVLRVVAEKIQHAEDLALVAVSFSGGKRELHPNDLAIKSKSLEASGRIYVYRK
NLADTLNPFANEESQQRPLRILGMNTWDLALELMNFDWLSLFNSIHEQELIYFTFSRQ
GSGEHTANLSLLLQRCNEVQLWVATEILLCSQLGKRVQLVKKFIKIAAHCKAQRNLS
FFAIVMGLNTASVSRLSQTWEKIPGKFKKLFSELESITDPSLNHKAYRDTFKKMKPK
IPFMPLLLKDVTFIHEGNKTFLDNLVNFELHMIADTVRTL RHCRANQFGGDASPKHE
HELKSYVTHLLVIDSQALFELSHRIEPR"

```

ORIGIN

1 **cgcgtctccg aaagag**

//

VISTA DI ALLINEAMENTO VERTICALE AGGANCIATO A QUERY CON LETTERE D'IDENTITA'

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	TTTCGGAGACGCGATTATT	1357
LR134318.1	421823	TTTCGGAGACGCGATTATT	421806
CP027082.1	80817758	TTTCGGAGACGCGATTATT	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

<u>CP013264.1</u>	3247187	TCCTTTCGGAGACGCGAT	3247203
<u>XM 012422314.1</u>	432	TCGGAGACGCGATTATT	448
<u>AP023273.1</u>	71221	TCCTTTCGGAGACGCGAT	71237
<u>XM 032820819.1</u>	1673	CTTTCGGAGACGCGATT	1657
<u>XM 032820818.1</u>	1412	CTTTCGGAGACGCGATT	1396
<u>XM 032820817.1</u>	1123	CTTTCGGAGACGCGATT	1107
<u>XM 032807220.1</u>	6941	TCCTTTCGAGACGCGATTATT	6921
<u>XM 032807219.1</u>	6968	TCCTTTCGAGACGCGATTATT	6948
<u>CP044989.1</u>	123135	TCGGAGACGCGATTATT	123119
<u>CP042997.1</u>	7633838	CTTTCGGAGACGCGATT	7633822
<u>LR584241.1</u>	3301410	TCCTTTCGGAGACGCGAT	3301394
<u>CP023741.1</u>	1579636	TTCGGAGACGCGATTAT	1579620
<u>XM 007513213.1</u>	2073	CTCTTTCGGAGACGCGA	2057
<u>FO082274.1</u>	648929	CTCTTTCGGAGACGCGA	648913
<u>XM 003113858.1</u>	685	TCGGAGACGCGATTATT	669
<u>FP929003.1</u>	2288837	TCCTTTCGGAGACGCGAT	2288821
<u>CP000634.1</u>	709168	TCCTTTCGGAGACGCGAT	709152
<u>OA882732.1</u>	38100	CTCTTTCGGAGACGCG	38115
<u>XM 037544960.1</u>	1451	CTCTTTCGGAGACGCG	1466
<u>XM 037258044.1</u>	165	CTTTCGGAGACGCGAT	180
<u>XM 037258043.1</u>	166	CTTTCGGAGACGCGAT	181
<u>CP062022.1</u>	54352183	CGGAGACGCGATTATT	54352198
<u>CP062049.1</u>	25998022	CTCTTTCGGAGACGCG	25998037
<u>CP043893.1</u>	3895188	TCGGAGACGCGATTAT	3895203
<u>CP043891.1</u>	3755873	TCGGAGACGCGATTAT	3755888
<u>CP043884.1</u>	4031545	TCGGAGACGCGATTAT	4031560
<u>CP043880.1</u>	3830381	TCGGAGACGCGATTAT	3830396
<u>CP043876.1</u>	3835920	TCGGAGACGCGATTAT	3835935
<u>CP053164.1</u>	9752269	TTCGGAGACGCGATTA	9752284
<u>CP061007.1</u>	7161695	TTTCGGAGACGCGATT	7161710
<u>CP060577.1</u>	345982	TTCGGAGACGCGATTA	345997
<u>CP060561.1</u>	322497	TTCGGAGACGCGATTA	322512
<u>CP060552.1</u>	279310	TTCGGAGACGCGATTA	279325
<u>CP060540.1</u>	421508	TTCGGAGACGCGATTA	421523
<u>LR877232.1</u>	16749084	CTCTTTCGGAGACGCG	16749099
<u>OB736593.1</u>	165	CTCTTTCGGAGACGCG	150
<u>OB665008.1</u>	10526	CTCTTTCGGAGACGCG	10511
<u>LR901227.1</u>	24185	CTCTTTCGGAGACGCG	24170
<u>XM 037540769.1</u>	584	CTCTTTCGGAGACGAGATTA	565
<u>XR 005138601.1</u>	369	TTTCGGAGACGCGATT	354
<u>CP062018.1</u>	56490497	CGGAGACGCGATTATT	56490482
<u>XR 004986512.1</u>	2388	CTCTTTCGGAGACGCG	2373
<u>XM 036429202.1</u>	2345	CTCTTTCGGAGACGCG	2330
<u>XM 036429201.1</u>	2468	CTCTTTCGGAGACGCG	2453
<u>XM 036429200.1</u>	2495	CTCTTTCGGAGACGCG	2480
<u>XM 036429199.1</u>	2614	CTCTTTCGGAGACGCG	2599
<u>XM 036429198.1</u>	2617	CTCTTTCGGAGACGCG	2602
<u>CP061876.1</u>	4788579	TCTTTCGGAGACGCGA	4788564
<u>LR880646.1</u>	25246344	CTCTTTCGGAGACGCG	25246329
<u>XM 035929535.1</u>	458	CTCTTTCGGAGACGCG	443
<u>XM 035929534.1</u>	461	CTCTTTCGGAGACGCG	446
<u>LR862366.1</u>	31329587	CTCTTTCGGAGACGCG	31329572

PRODOTTI DEGLI ALLINEAMENTI SIGNIFICATIVI (CDS FEATURE)

- Epoxide hydrolase [GenBank](#); [GenBank](#)
- Uncharacterized protein LOC107271692 [GenBank](#); [GenBank](#); [GenBank](#); [GenBank](#); [GenBank](#); [GenBank](#); [GenBank](#); [GenBank](#); [GenBank](#); [GenBank](#); [GenBank](#); [GenBank](#); [GenBank](#); [GenBank](#); [GenBank](#); [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](#)

SEQUENZA N° 9/1

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

SEQUENZA N° 9/1 (T=2; C=0; G=1; A=3)

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° 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
Sequenza 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	Aquila chrysaetos chrysaetos genome assembly, chromosome: 16	Aquila chrysaetos chrysaetos	38.2	38.2	86%	2.2	100.00%	30610376
1/2 3/2 3/3 3/5	Danio kyathit genome assembly, chromosome: 9	Danio kyathit	38.2	102	86%	2.2	100.00%	70176753
LR812527.1	Mastacembelus armatus genome assembly, chromosome: 3	zig-zag eel	38.2	38.2	86%	2.2	100.00%	28130073
LR535835.1	Methylomusa anaerophila DNA, complete genome	Methylomusa anaerophila	38.2	38.2	86%	2.2	100.00%	4781198
AP018449.1	Macrobrachium nipponense isolate FS-2020 chromosome 46	Macrobrachium nipponense	38.2	68.4	86%	2.2	100.00%	48001745
CP062041.1	Danio kyathit genome assembly, chromosome: 7	Danio kyathit	38.2	68.4	86%	2.2	100.00%	89469171
LR812525.1	Epinephelus fuscoguttatus DNA, LG13, complete sequence	brown-marbled grouper	36.2	36.2	81%	8.9	100.00%	43720812
AP022687.1	Epinephelus fuscoguttatus DNA, LG2, complete sequence	brown-marbled grouper	36.2	66.4	100%	8.9	100.00%	52288159
AP022676.1	Prunus dulcis DNA, pseudomolecule Pd02	almond	36.2	36.2	81%	8.9	100.00%	24875702
AP019298.1	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_002182111.1	Scheffersomyces stipitis CBS 6054 Bud Emergence Mediator (BEM1), partial mRNA	Scheffersomyces stipitis CBS 6054	36.2	36.2	81%	8.9	100.00%	2123
XM_001387892.1	Mus musculus chromosome 16, clone RP23-129F22, complete sequence	house mouse	36.2	36.2	81%	8.9	100.00%	256720
AC110241.17	Mus musculus chromosome 16, clone	house mouse	36.2	36.2	81%	8.9	100.00%	215198
AC113180.5								

	RP23-366B13, complete sequence PREDICTED: Prunus dulcis myosin-12 (LOC117617764), transcript variant X3, mRNA	almond	36.2	36.2	81%	8.9	100.00%	4872
XM_034347292.1								
	PREDICTED: Prunus dulcis myosin-12 (LOC117617764), transcript variant X2, mRNA	almond	36.2	36.2	81%	8.9	100.00%	4955
XM_034347291.1								
	PREDICTED: Prunus dulcis myosin-12 (LOC117617764), transcript variant X1, mRNA	almond	36.2	36.2	81%	8.9	100.00%	5000
XM_034347290.1								
	Pseudochaenichthys georgianus genome assembly, chromosome: 12	South Georgia icefish	36.2	36.2	81%	8.9	100.00%	30946200
LR792557.1								
	Acidianus brierleyi strain DSM 1651 genome, complete genome	Acidianus brierleyi	36.2	36.2	81%	8.9	100.00%	2947244
CP029289.2								
	Lateolabrax maculatus chromosome Lm23	spotted sea bass	36.2	96.6	95%	8.9	100.00%	28603024
CP027284.1								
	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_022004535.1								
	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_022004534.1								
	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_022004533.1								
	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_022004532.1								
	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_022004531.1								
	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_022004530.1								
	PREDICTED: Prunus avium myosin-12 (LOC110760740), mRNA	sweet cherry	36.2	36.2	81%	8.9	100.00%	4996
XM_021963057.1								
	PREDICTED: Prunus persica myosin- 12 (LOC18786033), transcript variant X3, mRNA	peach	36.2	36.2	81%	8.9	100.00%	5238
XM_020557865.1								
	PREDICTED: Prunus persica myosin- 12 (LOC18786033), transcript variant X2, mRNA	peach	36.2	36.2	81%	8.9	100.00%	5161
XM_020557864.1								
	PREDICTED: Prunus persica myosin- 12 (LOC18786033), transcript variant X1, mRNA	peach	36.2	36.2	81%	8.9	100.00%	5203
XM_020557863.1								
	PREDICTED: Prunus mume myosin- 12-like (LOC103331946), partial mRNA	Japanese apricot	36.2	36.2	81%	8.9	100.00%	2283
XM_008234624.2								
	Arthroderma uncinatum phosphoserine aminotransferase (GIQ15_02998), partial mRNA	Arthroderma uncinatum	34.2	34.2	77%	35	100.00%	1335
XM_033550415.1								
	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_004496365.1								
	PREDICTED: Parus major EPM2A glucan phosphatase, laforin (EPM2A), transcript variant X3, misc_RNA	Great Tit	34.2	34.2	77%	35	100.00%	2012
XR_004496364.1								
	PREDICTED: Parus major ubiquitin specific peptidase 25 (USP25), transcript variant X3, mRNA	Great Tit	34.2	34.2	77%	35	100.00%	3744
XM_033518452.1								
	PREDICTED: Parus major ubiquitin specific peptidase 25 (USP25), transcript variant X2, mRNA	Great Tit	34.2	34.2	77%	35	100.00%	4346
XM_033518447.1								
	PREDICTED: Parus major ubiquitin specific peptidase 25 (USP25), transcript variant X1, mRNA	Great Tit	34.2	34.2	77%	35	100.00%	4556
XM_033518444.1								
	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_033052136.1								
	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_033052135.1								
	PREDICTED: Tyto alba alba ubiquitin specific peptidase 25 (USP25), partial mRNA	Tyto alba alba	34.2	34.2	77%	35	100.00%	2487
XM_032989801.1								

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 ubiquitin specific peptidase 25 (USP25), transcript variant X2, mRNA	zebra finch	34.2	34.2	77%	35	100.00%	3423
XM_030262523.2	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 coralgroup	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	Pipistrellus pipistrellus genome assembly, chromosome: 11	common pipistrelle	34.2	66.4	77%	35	100.00%	69386144
1/2 2/2 3/2 3/3 3/5 6/1 7/2 8/1 10/1 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 coralgroup	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	Pseudomonas stutzeri strain ATCC 14405 chromosome, complete genome	Pseudomonas stutzeri	34.2	66.4	77%	35	100.00%	4639098
1/2 5/1 7/1 7/2 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
CONSRM 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	

VISTA DI ALLINEAMENTO VERTICALE AGGANCIATO A QUERY CON LETTERE D'IDENTITA'

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

PRODOTTI DEGLI ALLINEAMENTI SIGNIFICATIVI (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](#) [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](#) [GenBank](#) [GenBank](#) [GenBank](#) [GenBank](#) [GenBank](#)
[GenBank](#) [GenBank](#) [GenBank](#) [GenBank](#) [GenBank](#) [GenBank](#) [GenBank](#) [GenBank](#) [GenBank](#) [GenBank](#) [GenBank](#) [GenBank](#)
[GenBank](#) [GenBank](#) [GenBank](#) [GenBank](#) [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](#)

SEQUENZA N° 10/1

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

SEQUENZA N° 10/1 (T=2; C=1; G=3; A=0)

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° 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
Sequenza 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 clupeioides 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,	house mouse	36.2	36.2	81%	8.9	100.00%	179708

	clone RP23-366A20, complete sequence								
CP050606.1	Canis lupus familiaris breed Labrador retriever chromosome 07a	dog	36.2	126	86%	8.9	100.00%	80755643	
CP050603.1	Canis lupus familiaris breed Labrador retriever chromosome 17a	dog	36.2	68.4	86%	8.9	100.00%	64150251	
CP050637.1	Canis lupus familiaris breed Labrador retriever chromosome 07b	dog	36.2	126	86%	8.9	100.00%	80755819	
CP050614.1	Canis lupus familiaris breed Labrador retriever chromosome 17b PREDICTED: Xiphophorus hellerii titin-like (LOC116723448), mRNA	dog	36.2	68.4	86%	8.9	100.00%	64169108	
XM_032568351.1	PREDICTED: Myripristis murdjan coiled-coil alpha-helical rod protein 1-like (LOC115371980), transcript variant X2, 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 X1, mRNA	pinecone soldierfish	36.2	36.2	81%	8.9	100.00%	3464	
XM_030069653.1	Streptopelia turtur genome assembly, chromosome: 6	pinecone soldierfish	36.2	36.2	81%	8.9	100.00%	3470	
LR594558.1	Streptopelia turtur genome assembly, chromosome: Z	Streptopelia turtur	36.2	36.2	81%	8.9	100.00%	40337708	
LR594555.1	PREDICTED: Denticeps clupeoides minichromosome maintenance 10 replication initiation factor (mcm10), transcript variant X3, mRNA	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 X2, 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 X1, mRNA	denticle herring	36.2	36.2	81%	8.9	100.00%	2837	
XM_028955182.1	PREDICTED: Xiphophorus couchianus titin-like (LOC114148751), mRNA	denticle herring	36.2	36.2	81%	8.9	100.00%	2840	
XM_028024223.1	Arachis hypogaea cultivar Shitouqi chromosome A07	Monterrey platyfish	36.2	36.2	81%	8.9	100.00%	91226	
CP030989.1	PREDICTED: Xiphophorus maculatus titin-like (LOC102219802), mRNA	peanut	36.2	36.2	81%	8.9	100.00%	83117376	
XM_023337148.1	PREDICTED: Acanthaster planci basement membrane-specific heparan sulfate proteoglycan core protein-like (LOC110979181), transcript variant X6, 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 X5, mRNA	crown-of-thorns starfish	36.2	36.2	81%	8.9	100.00%	11632	
XM_022234772.1	PREDICTED: Acanthaster planci basement membrane-specific	crown-of-thorns starfish	36.2	36.2	81%	8.9	100.00%	11888	
XM_022234771.1		crown-of-thorns starfish	36.2	36.2	81%	8.9	100.00%	11891	

XM_022234770.1	heparan sulfate proteoglycan core protein-like (LOC110979181), transcript variant X3, mRNA 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	heparan sulfate proteoglycan core protein-like (LOC110979181), transcript variant X2, mRNA 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	heparan sulfate proteoglycan core protein-like (LOC110979181), transcript variant X1, mRNA PREDICTED: Poecilia reticulata titin (LOC103461835), mRNA	crown-of-thorns starfish	36.2	36.2	81%	8.9	100.00%	13158
XM_017309405.1	PREDICTED: Poecilia formosa titin (ttn), mRNA	guppy	36.2	36.2	81%	8.9	100.00%	93950
XM_016662277.1	PREDICTED: Cyprinodon variegatus titin (ttn), mRNA	Amazon molly	36.2	36.2	81%	8.9	100.00%	103364
XM_015391676.1	PREDICTED: Poecilia latipinna titin (ttn), transcript variant X18, mRNA	sheepshead minnow	36.2	36.2	81%	8.9	100.00%	95566
XM_015043247.1	PREDICTED: Poecilia latipinna titin (ttn), transcript variant X17, mRNA	sailfin molly	36.2	36.2	81%	8.9	100.00%	88892
XM_015043246.1	PREDICTED: Poecilia latipinna titin (ttn), transcript variant X16, mRNA	sailfin molly	36.2	36.2	81%	8.9	100.00%	89735
XM_015043245.1	PREDICTED: Poecilia latipinna titin (ttn), transcript variant X15, mRNA	sailfin molly	36.2	36.2	81%	8.9	100.00%	91521
XM_015043244.1	PREDICTED: Poecilia latipinna titin (ttn), transcript variant X14, mRNA	sailfin molly	36.2	36.2	81%	8.9	100.00%	91535
XM_015043243.1	PREDICTED: Poecilia latipinna titin (ttn), transcript variant X13, mRNA	sailfin molly	36.2	36.2	81%	8.9	100.00%	91797
XM_015043241.1	PREDICTED: Poecilia latipinna titin (ttn), transcript variant X12, mRNA	sailfin molly	36.2	36.2	81%	8.9	100.00%	92154
XM_015043240.1	PREDICTED: Poecilia latipinna titin (ttn), transcript variant X11, mRNA	sailfin molly	36.2	36.2	81%	8.9	100.00%	92691
XM_015043239.1	PREDICTED: Poecilia latipinna titin (ttn), transcript variant X10, mRNA	sailfin molly	36.2	36.2	81%	8.9	100.00%	93168
XM_015043238.1	PREDICTED: Poecilia latipinna titin (ttn), transcript variant X9, mRNA	sailfin molly	36.2	36.2	81%	8.9	100.00%	93444
XM_015043237.1	PREDICTED: Poecilia latipinna titin (ttn), transcript variant X8, mRNA	sailfin molly	36.2	36.2	81%	8.9	100.00%	93714
XM_015043236.1	PREDICTED: Poecilia latipinna titin (ttn), transcript variant X7, mRNA	sailfin molly	36.2	36.2	81%	8.9	100.00%	93801
XM_015043235.1	PREDICTED: Poecilia latipinna titin (ttn), transcript variant X6, mRNA	sailfin molly	36.2	36.2	81%	8.9	100.00%	94494
XM_015043234.1	PREDICTED: Poecilia latipinna titin (ttn), transcript variant X5, mRNA	sailfin molly	36.2	36.2	81%	8.9	100.00%	94614
XM_015043233.1	PREDICTED: Poecilia latipinna titin (ttn), transcript variant X4, mRNA	sailfin molly	36.2	36.2	81%	8.9	100.00%	94758
XM_015043232.1	PREDICTED: Poecilia latipinna titin (ttn), transcript variant X3, mRNA	sailfin molly	36.2	36.2	81%	8.9	100.00%	94782
XM_015043231.1	PREDICTED: Poecilia latipinna titin (ttn), transcript variant X2, mRNA	sailfin molly	36.2	36.2	81%	8.9	100.00%	94848
XM_015043230.1	PREDICTED: Poecilia latipinna titin (ttn), transcript variant X1, mRNA	sailfin molly	36.2	36.2	81%	8.9	100.00%	95256
XM_015043229.1	PREDICTED: Poecilia latipinna titin (ttn), transcript variant X0, mRNA	sailfin molly	36.2	36.2	81%	8.9	100.00%	95361

XM_014984134.1	titin (ttn), transcript variant X1, mRNA PREDICTED: Poecilia mexicana	Poecilia mexicana	36.2	36.2	81%	8.9	100.00%	103334
CP011894.1	Ovis canadensis canadensis isolate 43U chromosome 9 sequence Poecilia reticulata genome assembly, chromosome: 2 PREDICTED: Fundulus heteroclitus titin, tandem duplicate 2 (ttn.2), mRNA	Ovis canadensis canadensis	36.2	161	100%	8.9	100.00%	94836416
LR880646.1	guppy	guppy	36.2	36.2	81%	8.9	100.00%	45373629
XM_036132592.1	mummichog	mummichog	36.2	36.2	81%	8.9	100.00%	96129
LS423614.1	Anas platyrhynchos genome assembly, chromosome: 4 Mus musculus chromosome 7, clone RP24-78M9, complete sequence	mallard	36.2	66.4	86%	8.9	100.00%	74522197
AC158296.6	house mouse	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 PREDICTED: Pan troglodytes phosphatidylinositol 4-kinase alpha (PI4KA), transcript variant X6, misc_RNA	chimpanzee	34.2	34.2	77%	35	100.00%	6434
XR_002941146.1	chimpanzee	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 PREDICTED: Pan troglodytes phosphatidylinositol 4-kinase alpha (PI4KA), transcript variant X1, mRNA	chimpanzee	34.2	34.2	77%	35	100.00%	6873
XM_024352560.1	chimpanzee	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 PREDICTED: Pongo abelii phosphatidylinositol 4-kinase alpha (PI4KA), transcript variant X4, mRNA	wild yak	34.2	94.6	81%	35	100.00%	79990276
XM_024240188.1	PREDICTED: Pongo abelii phosphatidylinositol 4-kinase alpha (PI4KA), transcript variant X3, 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 X2, misc_RNA PREDICTED: Pongo abelii phosphatidylinositol 4-kinase alpha (PI4KA), transcript variant X1, misc_RNA	Sumatran orangutan	34.2	34.2	77%	35	100.00%	6800
XR_002912889.1	Sumatran orangutan	Sumatran orangutan	34.2	34.2	77%	35	100.00%	7240
XR_002912888.1	PREDICTED: Salvelinus alpinus zinc finger protein 420 (LOC111979690), transcript variant X2, mRNA PREDICTED: Salvelinus alpinus zinc finger protein 420 (LOC111979690), transcript variant X1, mRNA	Sumatran orangutan	34.2	34.2	77%	35	100.00%	7285
XM_024010326.1	Arctic char	Arctic char	34.2	34.2	77%	35	100.00%	3488
XM_024010325.1	PREDICTED: Aotus nancymaae phosphatidylinositol 4-kinase alpha (PI4KA), mRNA	Arctic char	34.2	34.2	77%	35	100.00%	3491
XM_012466500.1	Ma's night monkey	Ma's night monkey	34.2	34.2	77%	35	100.00%	6754
AC236497.1	Xenopus laevis clone CH219-	African clawed frog	34.2	34.2	77%	35	100.00%	140175

212B10, complete sequence

[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](#)
[6/1](#)

phosphatidylinositol 4-kinase
alpha (PI4KA), transcript variant
2, mRNA

human

34.2 34.2 77% 35 100.00% 6658

[NM_001362863.2](#)
[1/2](#) [2/2](#) [3/2](#) [3/3](#) [5/1](#)
[6/1](#)

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](#)

Pipistrellus
pipistrellus genome
assembly,
chromosome: 12

common pipistrelle

34.2 128 100% 35 100.00% 58291996

[LR862361.1](#)

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

CONSRM 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

CONSRM 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	

VISTA DI ALLINEAMENTO VERTICALE AGGANCIATO A QUERY CON LETTERE D'IDENTITA'

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

<u>CP050637.1</u>	12932850	CCCTGGTCACTCTCT	12932864
<u>CP050637.1</u>	26351875	CTCCCTGGTCACTCT	26351889
<u>CP050637.1</u>	57087982	CTGGTCACTCTCTTA	57087968
<u>CP050614.1</u>	58962803	CCTGGTCACTCTCTTATT	58962786
<u>CP050614.1</u>	8705602	CCCTGGTCACTCTCTT	8705587
<u>XM 032568351.1</u>	89446	TCTCCCTGGTCACTCTCT	89429
<u>XM 030069654.1</u>	1793	TCTCCCTGGTCACTCTCT	1776
<u>XM 030069653.1</u>	1799	TCTCCCTGGTCACTCTCT	1782
<u>LR594558.1</u>	40292066	CCTGGTCACTCTCTTATT	40292049
<u>LR594555.1</u>	52712341	CCTGGTCACTCTCTTATT	52712324
<u>LR594555.1</u>	54643947	CTCCCTGGTCACTCT	54643961
<u>XM 028955184.1</u>	926	TCTCCCTGGTCACTCTCT	909
<u>XM 028955183.1</u>	923	TCTCCCTGGTCACTCTCT	906
<u>XM 028955182.1</u>	926	TCTCCCTGGTCACTCTCT	909
<u>XM 028024223.1</u>	89495	TCTCCCTGGTCACTCTCT	89478
<u>CP030989.1</u>	57446801	TCCCTGGTCACTCTCTTA	57446784
<u>XM 023337148.1</u>	82268	TCTCCCTGGTCACTCTCT	82251
<u>XM 022234773.1</u>	2170	CTCCCTGGTCACTCTCTT	2153
<u>XM 022234772.1</u>	2426	CTCCCTGGTCACTCTCTT	2409
<u>XM 022234771.1</u>	2429	CTCCCTGGTCACTCTCTT	2412
<u>XM 022234770.1</u>	3441	CTCCCTGGTCACTCTCTT	3424
<u>XM 022234769.1</u>	3693	CTCCCTGGTCACTCTCTT	3676
<u>XM 022234768.1</u>	3696	CTCCCTGGTCACTCTCTT	3679
<u>XM 017309405.1</u>	92231	TCTCCCTGGTCACTCTCT	92214
<u>XM 016662277.1</u>	101644	TCTCCCTGGTCACTCTCT	101627
<u>XM 015391676.1</u>	93812	TCTCCCTGGTCACTCTCT	93795
<u>XM 015043247.1</u>	87178	TCTCCCTGGTCACTCTCT	87161
<u>XM 015043246.1</u>	88021	TCTCCCTGGTCACTCTCT	88004
<u>XM 015043245.1</u>	89807	TCTCCCTGGTCACTCTCT	89790
<u>XM 015043244.1</u>	89821	TCTCCCTGGTCACTCTCT	89804
<u>XM 015043243.1</u>	90083	TCTCCCTGGTCACTCTCT	90066
<u>XM 015043241.1</u>	90440	TCTCCCTGGTCACTCTCT	90423
<u>XM 015043240.1</u>	90977	TCTCCCTGGTCACTCTCT	90960
<u>XM 015043239.1</u>	91454	TCTCCCTGGTCACTCTCT	91437
<u>XM 015043238.1</u>	91730	TCTCCCTGGTCACTCTCT	91713
<u>XM 015043237.1</u>	92000	TCTCCCTGGTCACTCTCT	91983
<u>XM 015043236.1</u>	92087	TCTCCCTGGTCACTCTCT	92070
<u>XM 015043235.1</u>	92780	TCTCCCTGGTCACTCTCT	92763
<u>XM 015043234.1</u>	92900	TCTCCCTGGTCACTCTCT	92883
<u>XM 015043233.1</u>	93044	TCTCCCTGGTCACTCTCT	93027
<u>XM 015043232.1</u>	93068	TCTCCCTGGTCACTCTCT	93051
<u>XM 015043231.1</u>	93134	TCTCCCTGGTCACTCTCT	93117
<u>XM 015043230.1</u>	93542	TCTCCCTGGTCACTCTCT	93525
<u>XM 015043229.1</u>	93647	TCTCCCTGGTCACTCTCT	93630
<u>XM 014984134.1</u>	101620	TCTCCCTGGTCACTCTCT	101603
<u>CP011894.1</u>	26322777	TCCCTGGTCACTCACTTATT	26322796
<u>CP011894.1</u>	73003361	TCTCCCTGGTCACTCTCT	73003344
<u>CP011894.1</u>	74063983	CCCTGGTCACTCTCTT	74063968
<u>CP011894.1</u>	73665434	CTCCCTGGTCACTCT	73665448
<u>CP011894.1</u>	13298756	CCCTGGTCACTCTCT	13298742
<u>LR880646.1</u>	41540390	TCTCCCTGGTCACTCTCT	41540373
<u>XM 036132592.1</u>	94364	TCTCCCTGGTCACTCTCT	94347
<u>LS423614.1</u>	55414446	CTCCCTGGTCACTCTCTT	55414429
<u>LS423614.1</u>	7212898	TCTCCCTGGTCACTC	7212884
<u>AC158296.6</u>	181325	TCCCTGGTCACTCTCTTA	181308
<u>XR 002941147.1</u>	5873	TCTCCCTGGTCACTCTC	5889
<u>XR 002941146.1</u>	6104	TCTCCCTGGTCACTCTC	6120
<u>XM 024352563.1</u>	6152	TCTCCCTGGTCACTCTC	6168

<u>XM 024352562.1</u>	6159	TCTCCCTGGTCACTCTC	6175
<u>XM 024352561.1</u>	6174	TCTCCCTGGTCACTCTC	6190
<u>XR 002941145.1</u>	6225	TCTCCCTGGTCACTCTC	6241
<u>XM 024352560.1</u>	6225	TCTCCCTGGTCACTCTC	6241
<u>XM 024353570.1</u>	2728	TCTCCCTGGTCACTCTC	2744
<u>CP027083.1</u>	64175254	CTCCCTGGTCACTCTCT	64175270
<u>CP027083.1</u>	33522521	TCCCTGGTCACTCTC	33522535
<u>CP027083.1</u>	74038647	CCTGGTCACTCTCTT	74038633
<u>XM 024240188.1</u>	6191	TCTCCCTGGTCACTCTC	6207
<u>XM 024240187.1</u>	6237	TCTCCCTGGTCACTCTC	6253
<u>XR 002912889.1</u>	6677	TCTCCCTGGTCACTCTC	6693
<u>XR 002912888.1</u>	6722	TCTCCCTGGTCACTCTC	6738
<u>XM 024010326.1</u>	738	TCTCCCTGGTCACTCTC	754
<u>XM 024010325.1</u>	738	TCTCCCTGGTCACTCTC	754
<u>XM 012466500.1</u>	6185	TCTCCCTGGTCACTCTC	6201
<u>AC236497.1</u>	14566	TCTCCCTGGTCACTCTC	14582
<u>AP023477.1</u>	70373543	CTCCCTGGTCACTCTCT	70373559
<u>AP023477.1</u>	34546277	CTCCCTGGTCACTCT	34546291
<u>NM 001362862.2</u>	6097	TCTCCCTGGTCACTCTC	6113
<u>NM 001362863.2</u>	6124	TCTCCCTGGTCACTCTC	6140
<u>FR799580.1</u>	185176	CTCCCTGGTCACTCTCT	185192
<u>XM 024464007.1</u>	472	CTCCCTCGTCACTCTCTTATT	452
<u>CP013203.1</u>	5136801	CCTGGTCACTCTCTTAT	5136785
<u>XM 012733809.1</u>	2275	TCTCCCTGGTCACTCTC	2259
<u>XM 004694238.2</u>	2267	TCTCCCTGGTCACTCTC	2251
<u>CP055242.1</u>	6706558	TCTCGTGGTCACTCTCTTAT	6706538
<u>AP023482.1</u>	22644416	TCTCCCAGGTCACTCTCTT	22644434
<u>AP023482.1</u>	16683251	TCTCCCTGGTCACTCTC	16683235
<u>AP023482.1</u>	44075993	CCCTGGTCACTCTCTT	44075978
<u>AP023482.1</u>	15944981	CTCCCTGGTCACTCT	15944967
<u>AP023479.1</u>	48016079	CTCCCTGTTCACTCTCTTATT	48016059
<u>LR877229.1</u>	10280687	CTCCCTGGTCACTCTCT	10280671
<u>LR877229.1</u>	12079709	TGGTCACTCTCTTAT	12079695
<u>LR877229.1</u>	47555154	CTCCCTGGTCACTCT	47555140
<u>LR877198.1</u>	67955308	TCCCTGGTCACTCTCTT	67955292
<u>LR877198.1</u>	13949408	TCTCCCTGGTCACTC	13949394
<u>LR877198.1</u>	39660837	CCCTGGTCACTCTCT	39660823
<u>LR877198.1</u>	87550219	CCTGGTCACTCTCTT	87550205
<u>LR862368.1</u>	30728468	TCCCTGGTCACTCTATTATT	30728449
<u>LR862368.1</u>	46682100	TCTCCCTGGTCACTCTC	46682084
<u>LR862368.1</u>	30678003	TCCCTGGTCACTCTCT	30677988
<u>LR862368.1</u>	32170032	CCTGGTCACTCTCTT	32170018
<u>LR862361.1</u>	13011192	TCTCCCTGGTCTCTCTCTT	13011174
<u>LR862361.1</u>	6633788	CTGGTCACTCTCTTATT	6633772
<u>LR862361.1</u>	60936043	TCTCCCTGGTCACTCTC	60936027
<u>LR862361.1</u>	22931834	CTCCCTGGTCACTCT	22931848
<u>LR862361.1</u>	23032086	CTCCCTGGTCACTCT	23032100
<u>LR862361.1</u>	50687521	TCTCCCTGGTCACTC	50687507
<u>AC279687.1</u>	88398	CTCCCTGGTCACTCTCT	88382
<u>CP002352.1</u>	2034124	TCTCCCTGGTCACTCTC	2034108
<u>AP023475.1</u>	53280683	TCCCTGGTCTCTCTCTTATT	53280702
<u>AP023475.1</u>	53740717	TCCCTGGTCACTCTC	53740731
<u>AP023475.1</u>	81848063	CCCTGGTCACTCTCT	81848077
<u>AP023476.1</u>	71003346	TCTCCCTGGTCACTC	71003360
<u>AP023476.1</u>	5618759	CTGGTCACTCTCTTA	5618745
<u>AP023476.1</u>	53953669	CCCTGGTCACTCTCT	53953655

CONCLUSIONI

CONSIDERAZIONI FINALI

In questo lavoro, avvalendoci della Teoria Tricromatica dell'Equilibrio dei Sistemi e del suo programma computerizzato, è stata modificata la sequenza di 22 basi di *RNA* del *Gene RdRp* del *SARS-CoV-2* utilizzata come controllo positivo del tampone rinofaringeo. Sono stati individuati in parte e chiariti meglio i legami tra l'**RNA** del *SARS-CoV-2* e il genoma del *Pipistrello* e del *Pangolino*, entrambi considerati implicati nell'infezione da Covid-19. Questi legami saranno approfonditi meglio nella **Parte Seconda** di questo documento. Ampio spazio è stato dedicato all'analisi di una sequenza dalle caratteristiche nuove (*misc_feature*) del *Virus dell'Immunodeficienza Umana (HIV-1)* identificata dalla modificazione della sequenza di 22 basi di *RNA* del *Gene RdRp* del *SARS-CoV-2*. Lo stesso trattamento è stato riservato a diverse sequenze dell'*Influenzavirus A* e, soprattutto, alle numerose sequenze dei *Bacilli Paranthracis* e *Anthraxis*. Riguardo quest'ultimo batterio, essendo considerato nella categoria di massima priorità degli agenti infettivi associati a rischio potenziale d'uso come arma biologica, sono auspicabili approfondimenti ulteriori da parte della comunità scientifica internazionale al fine di neutralizzarne gli aspetti genomici comuni individuati (o ulteriormente individuabili) al *SARS-CoV-2*. Per rendere possibile la replicazione dei risultati mostrati e per permettere a tutti gli scienziati di fare ulteriori autonome indagini, è stata insegnata la procedura utilizzata per la modificazione di sequenze di DNA o RNA. Le modificazioni della sequenza di *RNA* del *Gene RdRp* del *SARS-CoV-2* e la metodologia proposta in questo documento aprono prospettive diverse per avviare studi finalizzati e trattamenti innovativi della pandemia da Covid 19 e per una riorganizzazione più funzionale delle banche dati del DNA e dell'RNA.

FINE
DELLA
PARTE PRIMA



www.ttesystems.eu

Corrispondenza: *nunzio.bonaventura@libero.it*

13 Aprile 2021