DNA or RNA Sequence Analysis and Modification Through the
TRICROMATIC THEORY
OF EQUILIBRIUM
OF SYSTEMS

Chapter II°

*Comparison of the Complete Analyses
of 4 Sequences
of the Macaca Nemestrina organism
obtained by the Significant Alignments
of Insulin A Chain*

Author: *Nunzio Bonaventura*
It Technictian: *Vincenzo Viggiano*

12 Agosto 2019

INDEX CHAPTER SECOND

INTRODUCTION

In this Chapter, the results of the *Comparison of the Complete Analyses of 4 Sequences of the organism Macaca Nemestrina* will be discussed.
The *Complete Analyses of Sequences* were carried out through the application of the **TRICROMATIC THEORY OF EQUILIBRIUM OF SYSTEMS** or so called **T.T.E.S.** (for further info, please check out the website www.ttesystems.eu).

Before starting the reading of this Chapter, readers are referred to read and familiarise themselves with the Introduction and paragraphs 1.1 e 1.2 of the Chapter I°- First Part [(DNA or RNA Sequence Analysis and Modification through T_T_E_S_ (Chapter I° - First Part)] and the Introduction and Paragraphs 1.1, 1.2, 1.3, 1.41, 1.42 e 1.43 of Chapter I°- Second Part [(DNA or RNA Sequence Analysis and Modification through T_T_E_S_ (Chapter I° - Second Part)]

The *Complete Analyses* of the **4 Sequences** were obtained by the *significant alignments* of the *Insulin A Chain* (see Paragraf 1.1 of Chapter I°- Second Part):

Sequences producing significant alignments:							
Selected seq	Description	Max score	Total score	Query cover	E value	Ident	Accession
XM_011721319.1 2/1 3/1 8/1 17/1	PREDICTED: Macaca nemestrina insulin (INS), transcript variant X4, mRNA	110	110	100%	6e-21	98%	XM_011721319.1
XM_011721318.1 2/1 3/1 8/1 17/1	PREDICTED: Macaca nemestrina insulin (INS), transcript variant X3, mRNA	110	110	100%	6e-21	98%	XM_011721318.1
XM_011721317.1 2/1 3/1 8/1 17/1	PREDICTED: Macaca nemestrina insulin (INS), transcript variant X2, mRNA	110	110	100%	6e-21	98%	XM_011721317.1
XM_011721316.1 2/1 3/1 8/1 17/1	PREDICTED: Macaca nemestrina insulin (INS), transcript variant X1, mRNA	110	110	100%	6e-21	98%	XM_011721316.1

ATTENTION: The BLAST research related to the above- mentioned *significant alignments* was carried out in date **9/04/2019**. As advised by the people responsible of the BLAST DATABASES, future eventual variations of *significant alignments* should be ascribed to regular updates and/or modifications due to the mobile parts of this intricate DATABASES).

Here, we have selected the following **4 mRna sequences** since their **four** products are "**insulin isoform**":

1) XM_011721319.1 Product = "insulin isoform X2";
2) XM_011721318.1 Product = "insulin isoform X1";
3) XM_011721317.1 Product = "insulin isoform X1";
4) XM_011721316.1 Product = "insulin isoform X1".

These **4 mRna sequences** share many nitrogenous bases amongst themselves.

In this Chapter, the **potentiality of the graphs** elaborated by putting in the Excel program the data obtained by the **T.T.E.S.** software will be explored.

The goal is to highlight the **validity**, **reliability** and **sensibility** of the calculations made for the elaboration and the graphic representation of the data.

The **validity** of an *instrument* is the degree of precision through which the instrument measures what it sets itself to measure. In the case here considered, **validity** is proven if the relationships among the (Dna or Rna) sequences, the calculations made for the elaboration of all the graphs and their factual rappresentations are *effective*, *not due to chances.*

The **reliability**, intended as **internal consistency**, is the degree of precision that the instrument can get, notwithstanding the presence of external factors or distrurbances.
In the case here considered, **reliability**, is proven if the relationships of Dna (or Rna) sequences, the calculations made for the elaboration of all the graphs and their rapresentations can be still considered **consistent,** notwistanding the *lenght of the sequences, the different types of "trends"*, the *"quality" of the single nitrogenous bases* (concerning the last two aspects, please see paragraf 2.2) and other parameters, such as whether the *sequences belong to the same organism or different ones, or different species*.

The **sensibility** of an *instrument* is its capacity of discriminating and differentiate different phenomena, and highlighting which are the different aspects amongst those phenomena.

Hence, from the ***Comparison of the Complete Analyses of the 4 Sequences if the organism Macaca Nemestrina*** considered here in this chapter, we expect that:
1) the sequence XM_011721319.1 (Product = "insulin isoform X2") is to be visualised in a **slightly different manner,** compared to the other three sequences (Product = "insulin isoform X1");
2) the sequences XM_011721316.1, XM_011721317.1 and XM_011721318.1 are to be graphed in a **very similar manner**, although the "*quantity*" and "*quality" of the single nitrougenous bases* are *in part different* in each of these sequences.

CHAPTER II°

Complete Analysis
of the
Sequence XM_011721319.1

PREDICTED: Macaca nemestrina insulin (INS), transcript variant X4, mRNA

2.1 CHARACTERISTICS OF THE SEQUENCE XM_011721319.1
PREDICTED: Macaca nemestrina insulin (INS), transcript variant X4, mRNA

Here, out of the numerous results obtained by the **BLAST** (*Basic Local Alignment Search Tool* (1)) research carried out on **Insulin A Chain** (see Paragraf 1.1 of Chapter I°- Second Part, ["DNA or RNA Sequence Analysis and Modification through T_T_E_S_ (Chapter I° - Second Part)"], we highlight the *significant alignment* with the **Macaca nemestrina** *organism*'s **mRNA** (**SEQUENCE XM_011721319.1**).

PREDICTED: Macaca nemestrina insulin (INS), transcript variant X4, mRNA
Sequence ID: XM_011721319.1 2/1 3/1 8/1 17/1
Product="insulin isoform X2"
Length=297   Number of Matches: 1
Range 1: 232 to 294 **GenBank   Graphics   FASTA**

Score	Expect	Identities	Gaps	Strand
110 bits(121)	7e-21	62/63(98%)	0/63(0%)	Plus/Plus

```
Query  1    GGCATCGTGGAGCAGTGCTGCACCAGCATCTGTTCCCTCTACCAGCTGGAGAACTACTGC  60
            ||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Sbjct  232  GGCATCGTGGAGCAGTGCTGCACCAGCATCTGCTCCCTCTACCAGCTGGAGAACTACTGC  291

Query  61   AAC  63
            |||
Sbjct  292  AAC  294
```

PREDICTED: Macaca nemestrina insulin (INS), transcript variant X4, mRNA (NCBI Reference Sequence: XM_011721319.1)

```
LOCUS       XM_011721319             297 bp   mRNA    linear   PRI 24-APR-2018
DEFINITION  PREDICTED: Macaca nemestrina insulin (INS), transcript variant X4,
            mRNA.
ACCESSION   XM_011721319
VERSION     XM_011721319.1
DBLINK      BioProject: PRJNA279145
KEYWORDS    RefSeq.
SOURCE      Macaca nemestrina (pig-tailed macaque)
  ORGANISM  Macaca nemestrina
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini;
            Catarrhini; Cercopithecidae; Cercopithecinae; Macaca.
COMMENT     MODEL REFSEQ: This record is predicted by automated computational
            analysis. This record is derived from a genomic sequence
            (NW_012013911.1) 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             :: Macaca nemestrina Annotation Release
                                           101
            Annotation Version          :: 101
            Annotation Pipeline         :: NCBI eukaryotic genome annotation
                                           pipeline
            Annotation Software Version :: 8.0
            Annotation Method           :: Best-placed RefSeq; Gnomon
            Features Annotated          :: Gene; mRNA; CDS; ncRNA
            ##Genome-Annotation-Data-END##
```

```
FEATURES             Location/Qualifiers
     source          1..297
                     /organism="Macaca nemestrina"
                     /mol_type="mRNA"
                     /isolate="M95218"
                     /db_xref="taxon:9545"
                     /chromosome="Unknown"
                     /sex="female"
                     /tissue_type="blood"
     gene            1..297
                     /gene="INS"
                     /note="Derived by automated computational analysis using
                     gene prediction method: Gnomon. Supporting evidence
                     includes similarity to: 1 mRNA, 1 Protein, and 17%
                     coverage of the annotated genomic feature by RNAseq
                     alignments"
                     /db_xref="GeneID:105469786"
     CDS             1..297
                     /gene="INS"
                     /codon_start=1
                     /product="insulin isoform X2"
                     /protein_id="XP_011719621.1"
                     /db_xref="GeneID:105469786"
                     /translation="MALWMRLLPLLALLALWGPDPAPAFVNQHLCGSHLVEALYLVCG
                     ERGFFYTPKTRREAEDPQGSLQPLALEGSLQKRGIVEQCCTSICSLYQLENYCN"
     STS             1..296
                     /gene="INS"
                     /standard_name="PMC123023P3"
                     /db_xref="UniSTS:270424"
     STS             114..295
                     /gene="INS"
                     /standard_name="Ins1"
                     /db_xref="UniSTS:267003"
     STS             139..290
                     /gene="INS"
                     /standard_name="PMC24644P6"
                     /db_xref="UniSTS:265494"
ORIGIN
        1 atggccctgt ggatgcgcct cttgcccctg ctggcgctgc tggccctctg gggacctgac
       61 ccggcccccgg cctttgtgaa ccagcacctg tgcggctccc acctggtgga agctctctac
      121 ctggtgtgcg gggagcgagg cttcttctac acacccaaga cccgccggga ggcagaggac
      181 cctcagggca gcctgcagcc cttggcgctg gagggggtccc tgcagaagcg cggcatcgtg
      241 gagcagtgct gcaccagcat ctgctccctc taccagctgg agaactactg caactag
//
```

The information on the characteristics of the above- mentioned **SEQUENCE XM_011721319.1** were directly acquired from the **NCBI** [*National Center for Biotechnology Information* (2)] website.

(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/

**2.2 GRAPHIC RESULTS OF THE ANALYSIS OF THE SEQUENCE XM_011721319.1
PREDICTED: Macaca nemestrina insulin (INS), transcript variant X4, mRNA**

All the graphs presented in this paragraf are referred to the **entire sequence of bases** analysed.

The **8 Principal Code Profile** (see the graph in Pct. 1) constitutes a very general synthesis of the entire sequence.

That is to say that it *SYNTHETISES IN AN UNSPECIFIC MANNNER* the joint contribution of the single bases (to a minor extent) and that of the *"Trend"* (to a greater extent) in *characterising* the sequence.

Such a **Profile** is not limited in any way by the sequence lenght; however, the longer is the sequence, the lesser is the capacity of highlighting *peculiar aspects of the entire sequence synthesis.*

Such a graph is very useful to highlight the general characteristics of the sequence (above all, it does so in relation to the role carried out by the "Trend").

Attention: *SIMILAR GRAPHS might rappresent DIFFERENT SEQUENCES, that HOWEVER have SIMILAR GENERAL CHARACTERISTICS.*
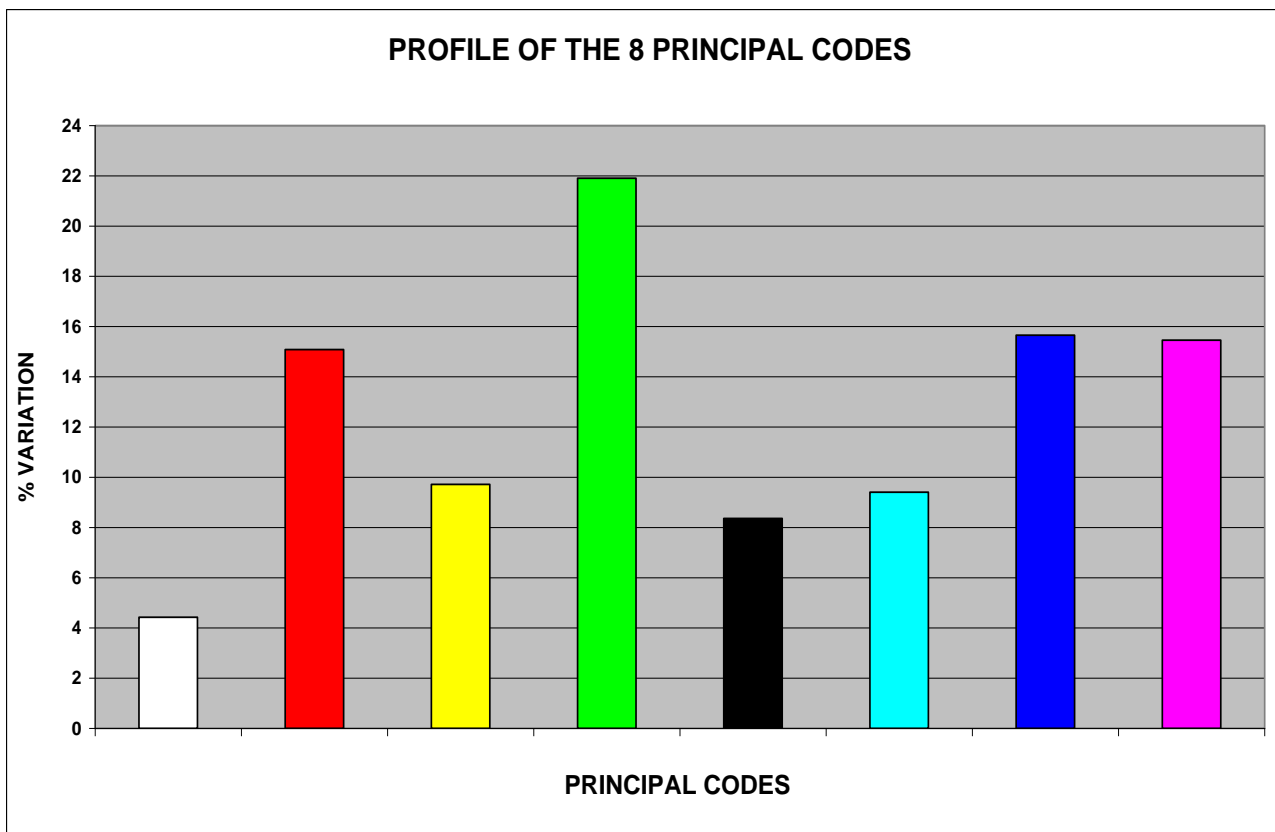
Pct. 1

9

The **Distribution of the Variation Percentage of the 8 Principal Codes** (see graph in Pct. 2) is a graph that <u>clearly highlights precise aspects of the "Trend"</u> of the entire sequence.

Such a **Profile** <u>is not limited by the sequence lenght</u>; however, the longer is the sequence, the lesser is the capacity of the graph of making perceptibly discriminable the *peculiar aspects of the "Trend"*.

Such a graph is very useful to highlight the *characteristics of the sequence "Trend"*.

<u>Attention</u>: ***SIMILAR GRAPHS might indicate DIFFERENT SEQUENCES, that HOWEVER have SIMILAR "TREND" ASPECTS***.
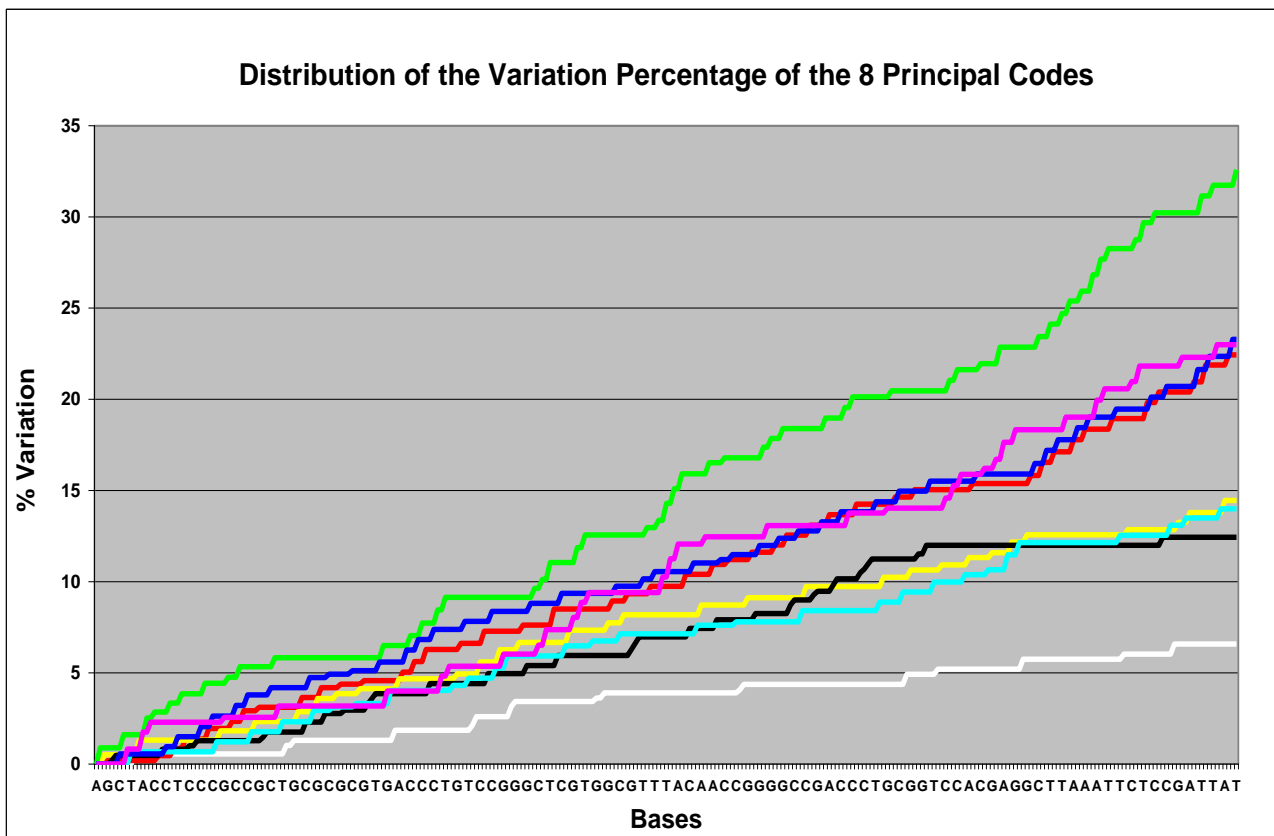
Pct. 2

The graph of the **64 Total Code Tonalities** (see graph in Pct. 3) highlights the <u>single</u> <u>bases</u> of the entire sequence, while offering very little information on its "Trend".

Such a **Profile** <u>is not limited by the lenght</u> of the sequence; however, the longer is the sequence, the lesser is the capacity of the graph of making the *single bases* perceptibly discriminable.

Such a graph is very useful to highlight the "*quality*" (Tonality and % of Variation) of the *single bases* of the *sequence*.

<u>Attention:</u> *SIMILAR GRAPHS rapresents SIMILAR SEQUENCES.*
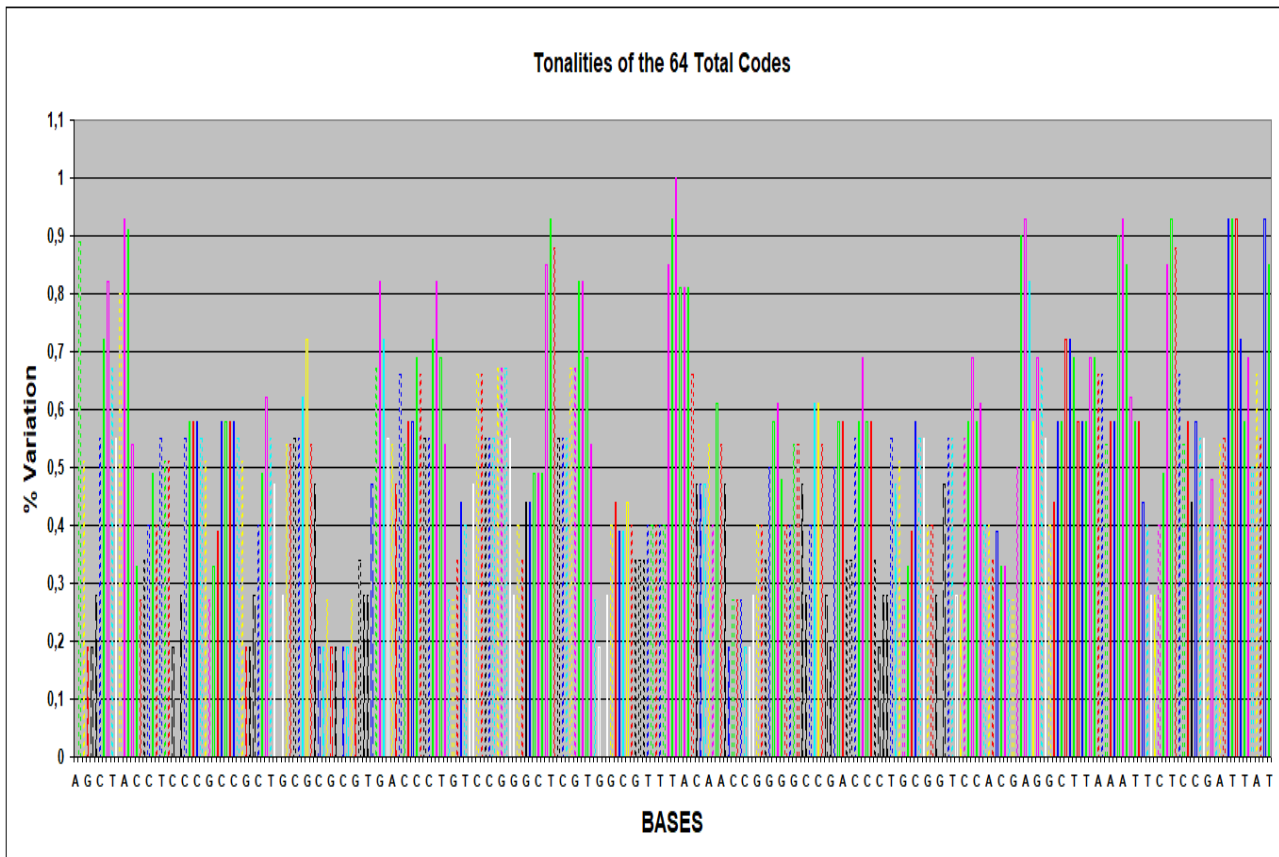
Pct. 3

The **64 Total Code Profile** shown in Pct. 4 is a <u>very specific synthesis</u> of the entire sequence.

That is to say that it *SYNTHETISES IN A SPECIFIC MANNER* and in the best way the joint contribution of the *single bases* (to a minor extent) and that of the *"Trend"* (to a greater extent) in *characterising* the sequence.

Such a **Profile** <u>is not limited</u> in any way by the <u>lenght</u> of the sequence; however, the longer is the sequence, the lesser is the capacity of highlighting *every peculiar aspects of the synthesis of the entire sequence.*

Such a graph is very useful to quickly highlight the *identifying characteristics* of a *sequence.*

<u>Attention</u>: *SIMILAR GRAPHS might identify DIFFERENT SEQUENCES, that HOWEVER have SPECIFIC SIMILAR CHARACTERISTICS*.
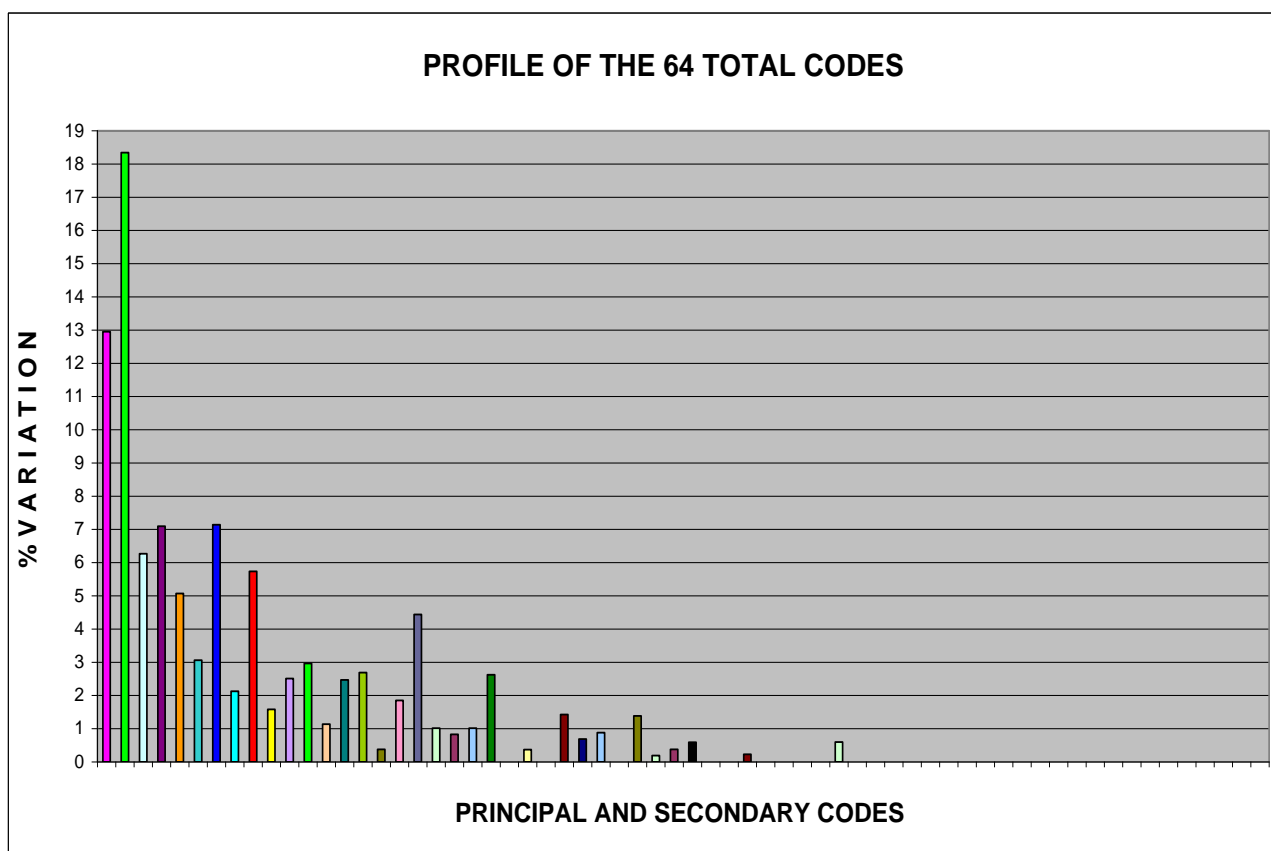
Pct. 4

Complete Analisys
of the
Sequence XM_011721318.1

PREDICTED: Macaca nemestrina insulin (INS), transcript variant X3, mRNA

2.3 CHARACTERISTICS OF THE SEQUENCE XM_011721318.1
PREDICTED: Macaca nemestrina insulin (INS), transcript variant X3, mRNA

Here, out of the numerous results obtained by the **BLAST** (*Basic Local Alignment Search Tool* (1)) research carried out **Insulin A Chain** (see Paragraf 1.1 of Chapter I°- Second Part, ["DNA or RNA Sequence Analysis and Modification through T_T_E_S_ (Chapter I° - Second Part)"], we highlith the *significant alignment* with the **Macaca nemestrina** *organism's* **mRNA** (**SEQUENCE XM_011721318.1**).

PREDICTED: Macaca nemestrina insulin (INS), transcript variant X3, mRNA
Sequence ID: XM_011721318.1   2/1 3/1 8/1 17/1
Product="insulin isoform X1"
Length=532   Number of Matches: 1
Range 1: 395 to 457 **GenBank**   **Graphics**   **FASTA**

Score	Expect	Identities	Gaps	Strand
110 bits(121)	7e-21	62/63(98%)	0/63(0%)	Plus/Plus

```
Query  1    GGCATCGTGGAGCAGTGCTGCACCAGCATCTGTTCCCTCTACCAGCTGGAGAACTACTGC  60
            ||||||||||||||||||||||||||||||||||| |||||||||||||||||||||||||
Sbjct  395  GGCATCGTGGAGCAGTGCTGCACCAGCATCTGCTCCCTCTACCAGCTGGAGAACTACTGC  454

Query  61   AAC  63
            |||
Sbjct  455  AAC  457
```

PREDICTED: Macaca nemestrina insulin (INS),   transcript variant X3, mRNA   (NCBI Reference Sequence: XM_011721318.1)

```
LOCUS       XM_011721318            532 bp    mRNA    linear   PRI 24-APR-2018
DEFINITION  PREDICTED: Macaca nemestrina insulin (INS), transcript variant X3,
            mRNA.
ACCESSION   XM_011721318
VERSION     XM_011721318.1
DBLINK      BioProject: PRJNA279145
KEYWORDS    RefSeq.
SOURCE      Macaca nemestrina (pig-tailed macaque)
  ORGANISM  Macaca nemestrina
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini;
            Catarrhini; Cercopithecidae; Cercopithecinae; Macaca.
COMMENT     MODEL REFSEQ:  This record is predicted by automated computational
            analysis. This record is derived from a genomic sequence
            (NW_012013911.1) 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           :: Macaca nemestrina Annotation Release
                                         101
            Annotation Version        :: 101
            Annotation Pipeline       :: NCBI eukaryotic genome annotation
                                         pipeline
            Annotation Software Version :: 8.0
            Annotation Method         :: Best-placed RefSeq; Gnomon
            Features Annotated        :: Gene; mRNA; CDS; ncRNA
            ##Genome-Annotation-Data-END##

FEATURES             Location/Qualifiers
     source          1..532
                     /organism="Macaca nemestrina"
```

```
                    /mol_type="mRNA"
                    /isolate="M95218"
                    /db_xref="taxon:9545"
                    /chromosome="Unknown"
                    /sex="female"
                    /tissue_type="blood"
     gene           1..532
                    /gene="INS"
                    /note="Derived by automated computational analysis using
                    gene prediction method: Gnomon. Supporting evidence
                    includes similarity to: 2 mRNAs, 10 Proteins, and 9%
                    coverage of the annotated genomic feature by RNAseq
                    alignments"
                    /db_xref="GeneID:105469786"
     STS            8..254
                    /gene="INS"
                    /standard_name="GDB:181496"
                    /db_xref="UniSTS:155248"
     CDS            128..460
                    /gene="INS"
                    /codon_start=1
                    /product="insulin isoform X1"
                    /protein_id="XP_011719620.1"
                    /db_xref="GeneID:105469786"
                    /translation="MALWMRLLPLLALLALWGPDPAPAFVNQHLCGSHLVEALYLVCG
                    ERGFFYTPKTRREAEDPQVGQVELGGGPGAGSLQPLALEGSLQKRGIVEQCCTSICSL
                    YQLENYCN"
     STS            128..459
                    /gene="INS"
                    /standard_name="PMC123023P3"
                    /db_xref="UniSTS:270424"
     STS            241..458
                    /gene="INS"
                    /standard_name="Ins1"
                    /db_xref="UniSTS:267003"
     STS            266..453
                    /gene="INS"
                    /standard_name="PMC24644P6"
                    /db_xref="UniSTS:265494"
     STS            315..532
                    /gene="INS"
                    /standard_name="GDB:179433"
                    /db_xref="UniSTS:155046"
ORIGIN
        1 gggacaggct gcatcagaag aggccagcaa gcaggtctgt tccaagggcc ttcacgtcag
       61 gtgggctcag ggctgcccca cttgggggtt ccagggtggc tggaccccag gtcactgtcc
      121 ttccgccatg gccctgtgga tgcgcctctt gccccctgctg gcgctgctgg ccctctgggg
      181 acctgacccg gccccggcct ttgtgaacca gcacctgtgc ggctcccacc tggtggaagc
      241 tctctacctg gtgtgcgggg agcgaggctt cttctacaca cccaagaccc gccgggaggc
      301 agaggaccct caggtggggc aggtggagct gggcgggggc cctggcgcag gcagcctgca
      361 gcccttggcg ctggaggggt ccctgcagaa gcgcggcatc gtggagcagt gctgcaccag
      421 catctgctcc ctctaccagc tggagaacta ctgcaactag atgcggcccg caggcggccc
      481 acaccctcca cctcctgcac caagagagat cgaataaagc ccttgaacca gc
//
```

The information on the characteristics of the above- mentioned **SEQUENCE XM_011721318.1** were directly acquired by the **NCBI** [*National Center for Biotechnology Information* (2)] website.

(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/

2.4 GRAPHIC RESULTS OF THE ANALYSIS OF THE SEQUENCE XM_011721318.1 PREDICTED: Macaca nemestrina insulin (INS), transcript variant X3, mRNA

All the graphs presented in this paragraph are referred to the **entire sequence of bases** analised.
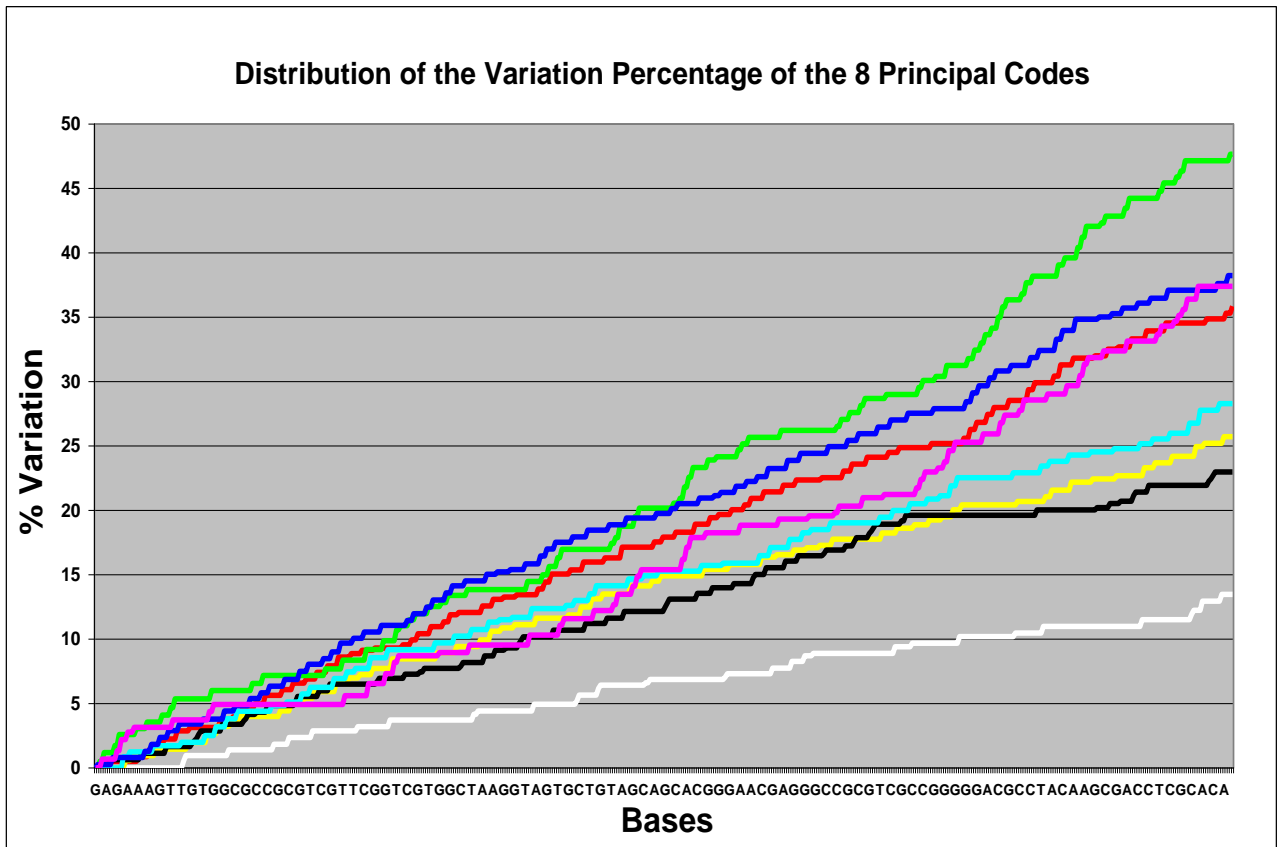
To avoid redundant explanations, here the reader who is interested in correctly interpreting the following graphs, is referred to what already explained in *paragraf 1.1* .

The **8 Principal Code Profile** (see graph in Pct. 5) constitutes a <u>very general synthesis</u> of the entire sequence.

**PROFILE OF THE 8 PRINCIPAL CODES**
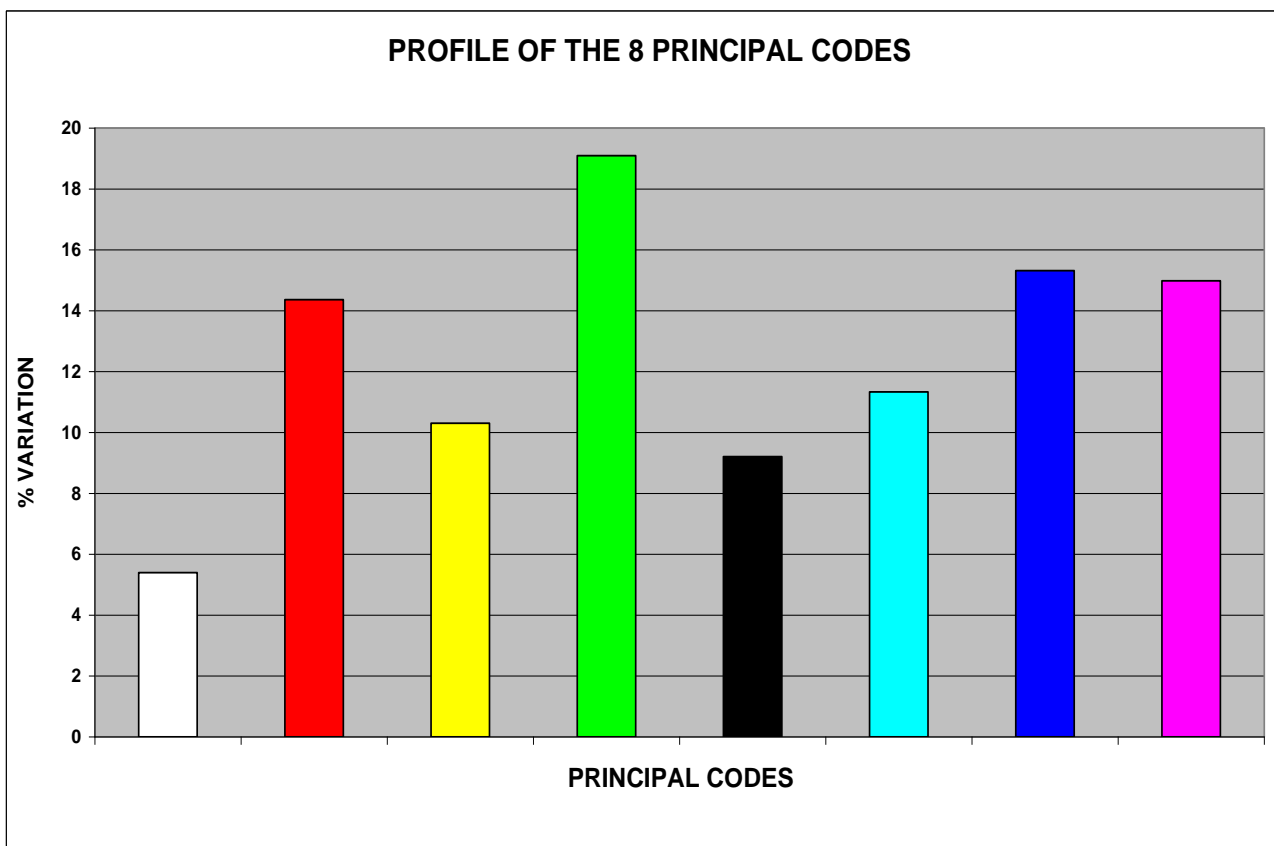
Pct. 5

The **Distribution of the Variation Percentage of the 8 Principal Codes** (see graph in Pct. 6) is a graph that <u>highlights, in a very clear manner, precise aspects of the</u> <u>"Trend"</u> of the entire sequence.

Pct. 6

The graph of the **64 Total Code Tonalities** (see the graph in Pct. 7) highlights the single bases of the entire sequence, while it offers less clear information on its "Trend".

Pct. 7

The **64 Total Code Profile** shown in Pct. 8 constitutes a <u>very specific synthesis</u> of the entire sequence.
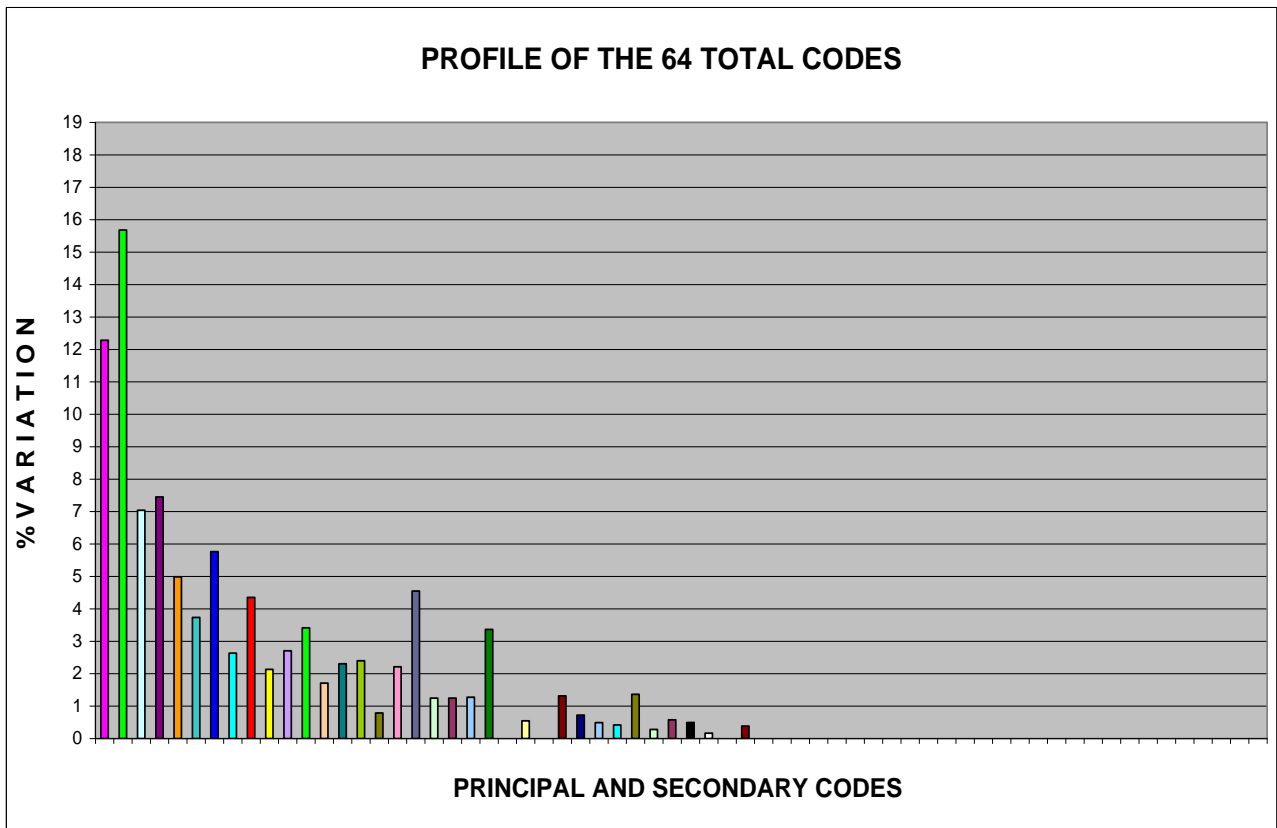
Pct. 8

Complete Analysis
of the
Sequence XM_011721317.1

PREDICTED: Macaca nemestrina insulin (INS), transcript variant X2, mRNA

2.5 CHARACTERISTICS OF THE SEQUENCE XM_011721317.1
PREDICTED: Macaca nemestrina insulin (INS), transcript variant X2, mRNA

Here, out of the numerous results obtained by the **BLAST** (*Basic Local Alignment Search Tool* (1)) research carried out on **Insulin A Chain** (see the Paragraf 1.1 of Chapter I°- Second Part, ["DNA or RNA Sequence Analysis and Modification through T_T_E_S_ (Chapter I° - Second Part)"], we highlight the *significant alignment* with the **Macaca nemestrina** *organism's* **mRNA** (**SEQUENCE XM_011721317.1**).

PREDICTED: Macaca nemestrina insulin (INS), transcript variant X2, mRNA
Sequence ID: XM_011721317.1   2/1 3/1 8/1 17/1
Product="insulin isoform X1"
Length=482   Number of Matches: 1
Range 1: 345 to 407 **GenBank   Graphics   FASTA**

Score	Expect	Identities	Gaps	Strand
110 bits(121)	7e-21	62/63(98%)	0/63(0%)	Plus/Plus

```
Query   1    GGCATCGTGGAGCAGTGCTGCACCAGCATCTGTTCCCTCTACCAGCTGGAGAACTACTGC   60
             |||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||
Sbjct   345  GGCATCGTGGAGCAGTGCTGCACCAGCATCTGCTCCCTCTACCAGCTGGAGAACTACTGC   404

Query   61   AAC   63
             |||
Sbjct   405  AAC   407
```

PREDICTED: Macaca nemestrina insulin (INS),   transcript variant X2, mRNA   (NCBI Reference Sequence: XM_011721317.1)

```
LOCUS       XM_011721317            482 bp    mRNA    linear   PRI 24-APR-2018
DEFINITION  PREDICTED: Macaca nemestrina insulin (INS), transcript variant X2,
            mRNA.
ACCESSION   XM_011721317
VERSION     XM_011721317.1
DBLINK      BioProject: PRJNA279145
KEYWORDS    RefSeq.
SOURCE      Macaca nemestrina (pig-tailed macaque)
  ORGANISM  Macaca nemestrina
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini;
            Catarrhini; Cercopithecidae; Cercopithecinae; Macaca.
COMMENT     MODEL REFSEQ:  This record is predicted by automated computational
            analysis. This record is derived from a genomic sequence
            (NW_012013911.1) 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             :: Macaca nemestrina Annotation Release
                                           101
            Annotation Version          :: 101
            Annotation Pipeline         :: NCBI eukaryotic genome annotation
                                           pipeline
            Annotation Software Version :: 8.0
            Annotation Method           :: Best-placed RefSeq; Gnomon
            Features Annotated          :: Gene; mRNA; CDS; ncRNA
            ##Genome-Annotation-Data-END##
FEATURES             Location/Qualifiers
     source          1..482
                     /organism="Macaca nemestrina"
                     /mol_type="mRNA"
```

```
                     /isolate="M95218"
                     /db_xref="taxon:9545"
                     /chromosome="Unknown"
                     /sex="female"
                     /tissue_type="blood"
     gene            1..482
                     /gene="INS"
                     /note="Derived by automated computational analysis using
                     gene prediction method: Gnomon. Supporting evidence
                     includes similarity to: 3 mRNAs, 10 Proteins, and 10%
                     coverage of the annotated genomic feature by RNAseq
                     alignments"
                     /db_xref="GeneID:105469786"
     STS             8..204
                     /gene="INS"
                     /standard_name="GDB:181496"
                     /db_xref="UniSTS:155248"
     CDS             78..410
                     /gene="INS"
                     /codon_start=1
                     /product="insulin isoform X1"
                     /protein_id="XP_011719619.1"
                     /db_xref="GeneID:105469786"
                     /translation="MALWMRLLPLLALLALWGPDPAPAFVNQHLCGSHLVEALYLVCG
                     ERGFFYTPKTRREAEDPQVGQVELGGGPGAGSLQPLALEGSLQKRGIVEQCCTSICSL
                     YQLENYCN"
     STS             78..409
                     /gene="INS"
                     /standard_name="PMC123023P3"
                     /db_xref="UniSTS:270424"
     STS             191..408
                     /gene="INS"
                     /standard_name="Ins1"
                     /db_xref="UniSTS:267003"
     STS             216..403
                     /gene="INS"
                     /standard_name="PMC24644P6"
                     /db_xref="UniSTS:265494"
     STS             265..482
                     /gene="INS"
                     /standard_name="GDB:179433"
                     /db_xref="UniSTS:155046"
ORIGIN
        1 gggacaggct gcatcagaag aggccagcaa gcaggtctgt tccaagggcc ttcacgtcag
       61 gtcactgtcc ttccgccatg gccctgtgga tgcgcctctt gcccctgctg gcgctgctgg
      121 ccctctgggg acctgacccg gccccggcct ttgtgaacca gcacctgtgc ggctcccacc
      181 tggtggaagc tctctacctg gtgtgcgggg agcgaggctt cttctacaca cccaagaccc
      241 gccgggaggc agaggaccct caggtggggc aggtggagct gggcgggggc cctggcgcag
      301 gcagcctgca gcccttggcg ctggaggggt ccctgcagaa gcgcggcatc gtggagcagt
      361 gctgcaccag catctgctcc ctctaccagc tggagaacta ctgcaactag atgcggcccg
      421 caggcggccc acaccctcca cctcctgcac caagagagat cgaataaagc ccttgaacca
      481 gc
//
```

The information on the characteristics of the above-mentioned **SEQUENCE XM_011721317.1** were directly acquired by the **NCBI** [*National Center for Biotechnology Information* (2)] website.

(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/

2.6 GRAPHIC RESULTS OF THE ANALYSIS OF SEQUENCE XM_011721317.1
PREDICTED: Macaca nemestrina insulin (INS), transcript variant X2, mRNA

**All the graphs** presented in this paragraf are referred to the **entire sequence of bases** analysed.

To avoid redundant explanations, here the reader who is interested in correctly interpreting the following graphs, is referred to what already explained in *paragraf 1.1* .

The **8 Principal Code Profile** (see graph in Pct. 9) constitutes a very general synthesis of the entire sequence.
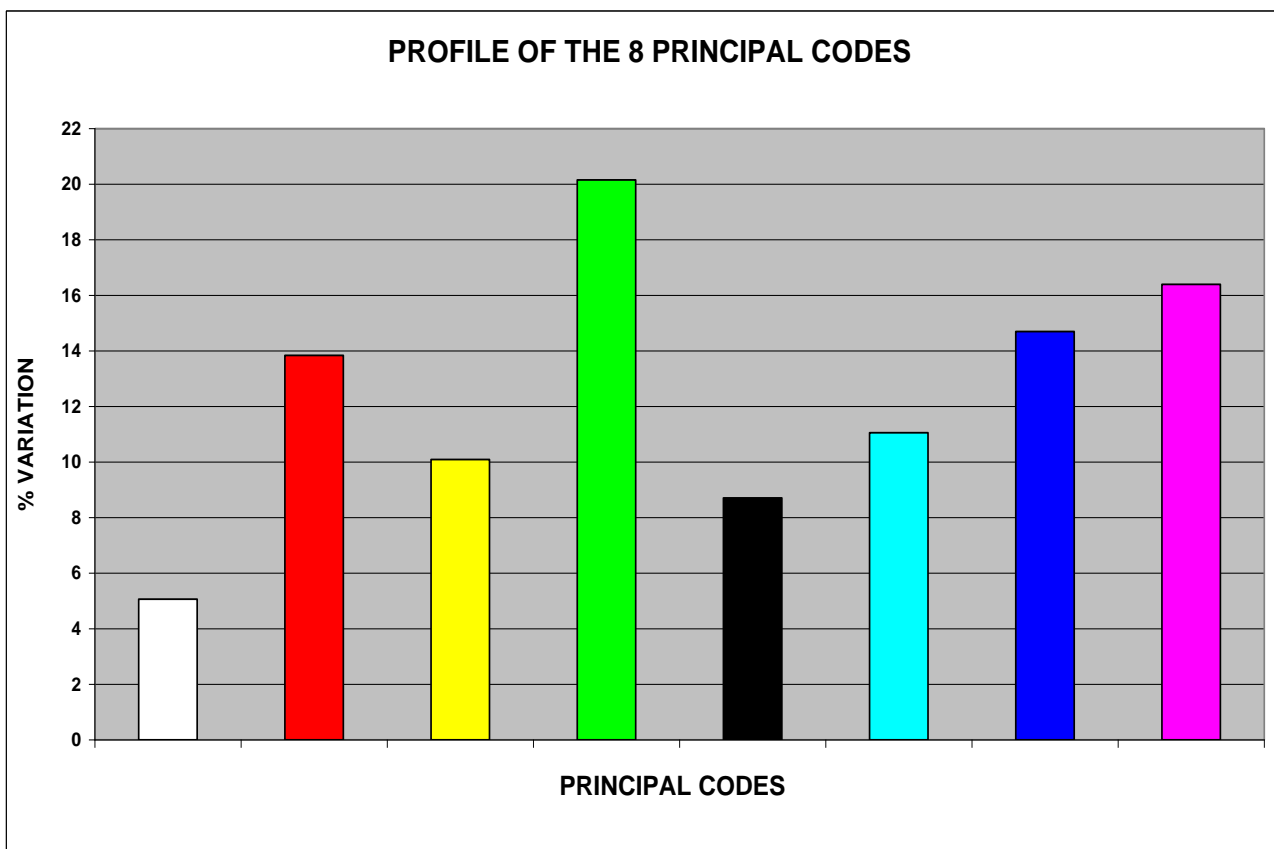
Pct. 9

The **Distribution of the Variation Percentage of the 8 Principal Codes** (see the graph in Pct. 10) is a graph that <u>clearly highlights precise aspects of the "Trend"</u> of the entire sequence.

Distribution of the Variation Percentage of the 8 Principal Codes
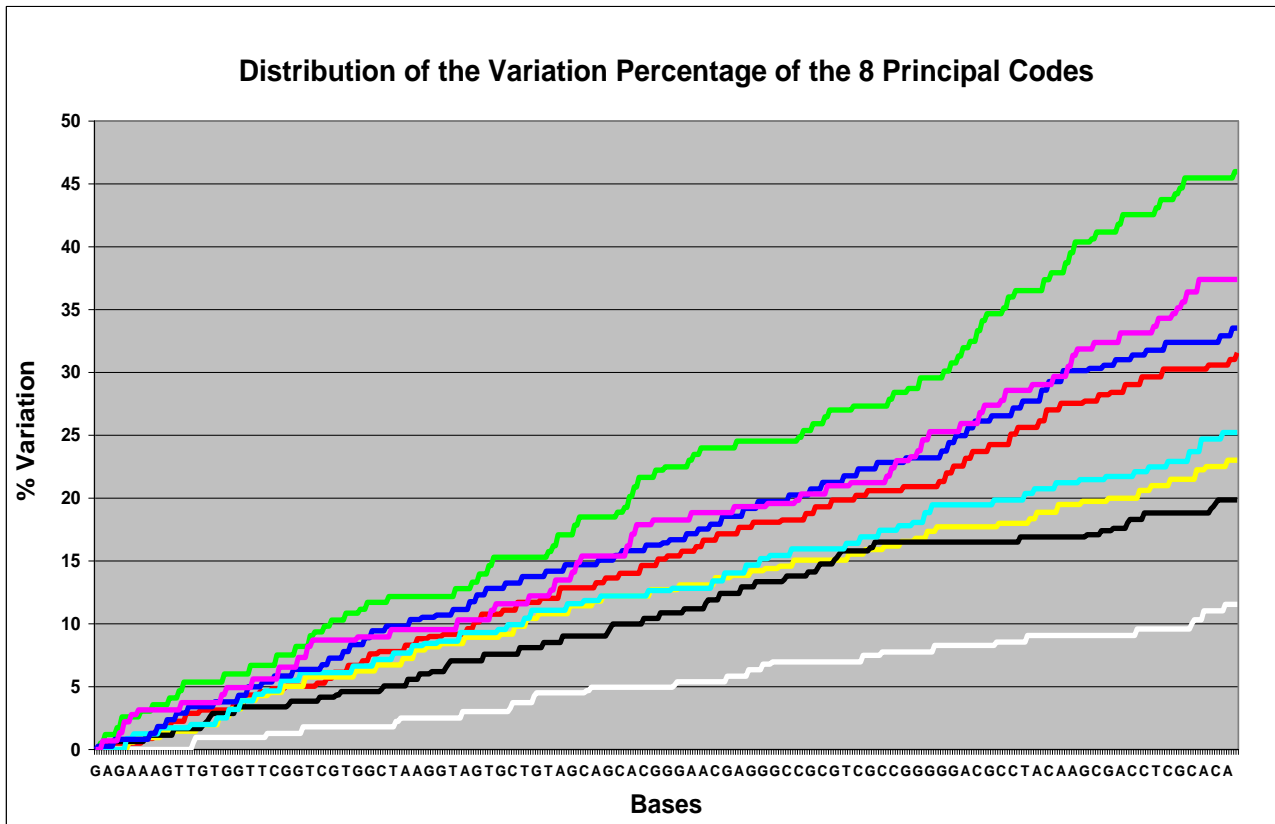
Pct. 10

The graph of the **64 Total Code Tonalities** (see the graph in Pct. 11) highlights the
single bases of the entire sequence, while offering less clear information on its
Trend".
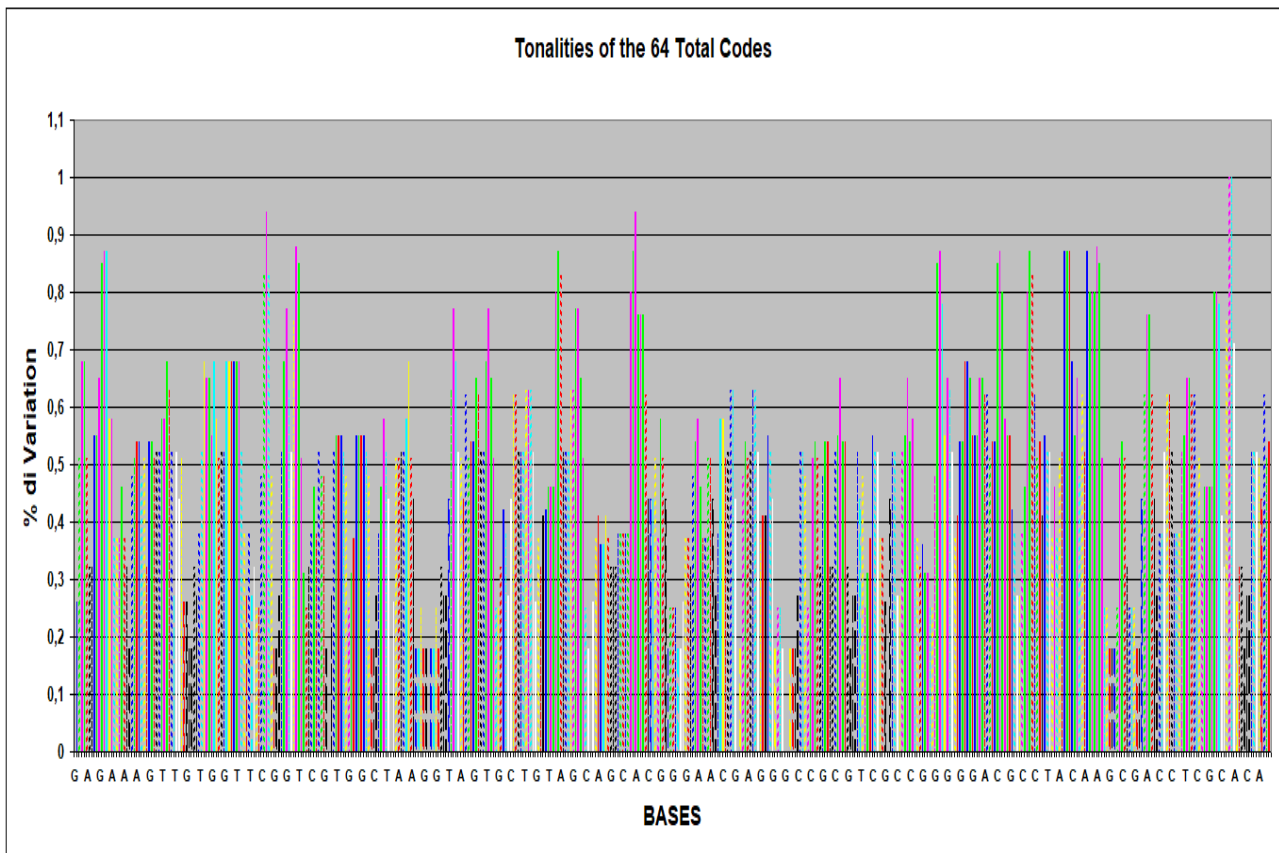
Pct. 11

The **64 Total Code Profile** shown in Pct. 12 constitutes <u>a very specific synthesis</u> of the entire sequence.

PROFILE OF THE 64 TOTAL CODES
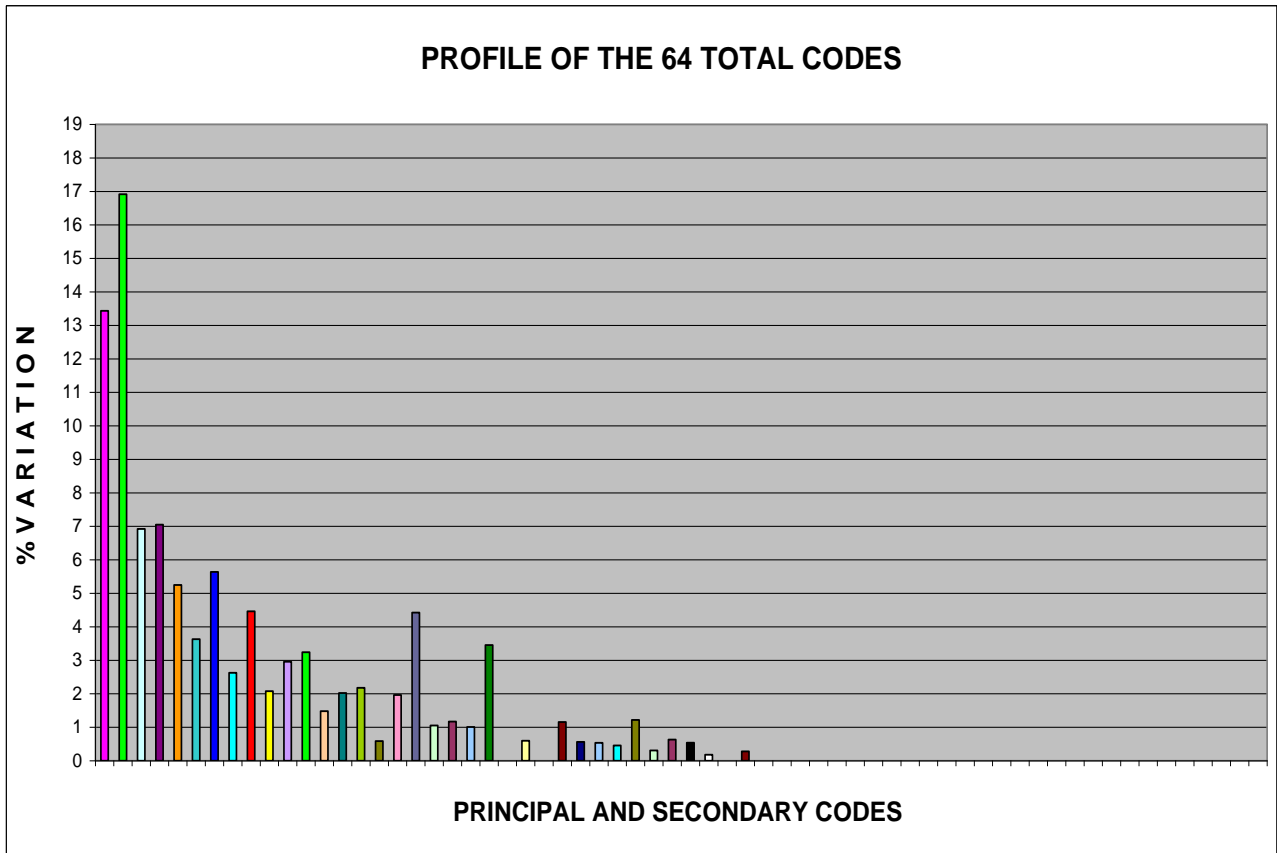
Pct. 12

Complete Analysis
of the
Sequence XM_011721316.1

PREDICTED: Macaca nemestrina insulin (INS), transcript variant X1, mRNA

2.7 CHARACTERISTICS OF THE SEQUENCE XM_011721316.1
PREDICTED: Macaca nemestrina insulin (INS), transcript variant X1, mRNA

Here, out of the numerous results obtained by the **BLAST** (*Basic Local Alignment Search Tool* (1)) research carried out on **Insulin Chain A** (see paragraf 1.1 of Chapter I°- Second Part, ["DNA or RNA Sequence Analysis and Modifcation through T_T_E_S_ (Chapter I° - Second Part)"], we highlight the *significant alignment* with the **Macaca nemestrina** *organism's* **mRNA** (**SEQUENCE XM_011721316.1**) .

PREDICTED: Macaca nemestrina insulin (INS), transcript variant X1, mRNA
Sequence ID: XM_011721316.1 2/1 3/1 8/1 17/1
Product="insulin isoform X1"
Length=454   Number of Matches: 1
Range 1: 319 to 381 **GenBank** **Graphics** **FASTA**

Score	Expect	Identities	Gaps	Strand
110 bits(121)	7e-21	62/63(98%)	0/63(0%)	Plus/Plus

```
Query  1    GGCATCGTGGAGCAGTGCTGCACCAGCATCTGTTCCCTCTACCAGCTGGAGAACTACTGC  60
            |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Sbjct  319  GGCATCGTGGAGCAGTGCTGCACCAGCATCTGCTCCCTCTACCAGCTGGAGAACTACTGC  378

Query  61   AAC  63
            |||
Sbjct  379  AAC  381
```

PREDICTED: Macaca nemestrina insulin (INS), transcript variant X1, mRNA   (NCBI Reference Sequence: XM_011721316.1)

```
LOCUS       XM_011721316            454 bp    mRNA    linear   PRI 24-APR-2018
DEFINITION  PREDICTED: Macaca nemestrina insulin (INS), transcript variant X1,
            mRNA.
ACCESSION   XM_011721316
VERSION     XM_011721316.1
DBLINK      BioProject: PRJNA279145
KEYWORDS    RefSeq.
SOURCE      Macaca nemestrina (pig-tailed macaque)
  ORGANISM  Macaca nemestrina
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini;
            Catarrhini; Cercopithecidae; Cercopithecinae; Macaca.
COMMENT     MODEL REFSEQ:  This record is predicted by automated computational
            analysis. This record is derived from a genomic sequence
            (NW_012013911.1) 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            :: Macaca nemestrina Annotation Release
                                          101
            Annotation Version         :: 101
            Annotation Pipeline        :: NCBI eukaryotic genome annotation
                                          pipeline
            Annotation Software Version :: 8.0
            Annotation Method          :: Best-placed RefSeq; Gnomon
            Features Annotated         :: Gene; mRNA; CDS; ncRNA
            ##Genome-Annotation-Data-END##
```

```
FEATURES             Location/Qualifiers
     source          1..454
                     /organism="Macaca nemestrina"
                     /mol_type="mRNA"
                     /isolate="M95218"
                     /db_xref="taxon:9545"
                     /chromosome="Unknown"
                     /sex="female"
                     /tissue_type="blood"
     gene            1..454
                     /gene="INS"
                     /note="Derived by automated computational analysis using
                     gene prediction method: Gnomon. Supporting evidence
                     includes similarity to: 5 mRNAs, 10 Proteins, and 11%
                     coverage of the annotated genomic feature by RNAseq
                     alignments"
                     /db_xref="GeneID:105469786"
     STS             8..178
                     /gene="INS"
                     /standard_name="GDB:181496"
                     /db_xref="UniSTS:155248"
     CDS             52..384
                     /gene="INS"
                     /codon_start=1
                     /product="insulin isoform X1"
                     /protein_id="XP_011719618.1"
                     /db_xref="GeneID:105469786"
                     /translation="MALWMRLLPLLALLALWGPDPAPAFVNQHLCGSHLVEALYLVCG
                     ERGFFYTPKTRREAEDPQVGQVELGGGPGAGSLQPLALEGSLQKRGIVEQCCTSICSL
                     YQLENYCN"
     STS             52..383
                     /gene="INS"
                     /standard_name="PMC123023P3"
                     /db_xref="UniSTS:270424"
     STS             165..382
                     /gene="INS"
                     /standard_name="Ins1"
                     /db_xref="UniSTS:267003"
     STS             190..377
                     /gene="INS"
                     /standard_name="PMC24644P6"
                     /db_xref="UniSTS:265494"
ORIGIN
        1 gggacaggct gcatcagaag aggccagcaa gcaggtcact gtccttccgc catggccctg
       61 tggatgcgcc tcttgcccct gctggcgctg ctggccctct ggggacctga cccggccccg
      121 gcctttgtga accagcacct gtgcggctcc cacctggtgg aagctctcta cctggtgtgc
      181 ggggagcgag gcttcttcta cacacccaag acccgccggg aggcagagga ccctcaggtg
      241 gggcaggtgg agctgggcgg gggccctggc gcaggcagcc tgcagccctt ggcgctggag
      301 gggtccctgc agaagcgcgg catcgtggag cagtgctgca ccagcatctg ctccctctac
      361 cagctggaga actactgcaa ctagatgcgg cccgcaggcg gcccacaccc tccacctcct
      421 gcaccaagag agatcgaata aagcccttga acca
//
```

The information on the characteristics of the above-mentioned **SEQUENCE XM_011721316.1** were directly acquired by the **NCBI** [*National Center for Biotechnology Information* (2)] website.

(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/

2.8 GRAPHIC RESULTS OF THE ANALYSIS OF THE SEQUENCE XM_011721316.1
PREDICTED: Macaca nemestrina insulin (INS), transcript variant X1, mRNA

All the graphs presented in this paragraph are referred to the **entire sequence of bases** analised.

To avoid redundant explanations, here the reader who is interested in correctly interpreting the following graphs, is referred to what already explained in *paragraf 1.1* .

The **8 Principal Code Profile** (see the graph in Pct. 13) constitutes a <u>very general synthesis</u> of the entire sequence.
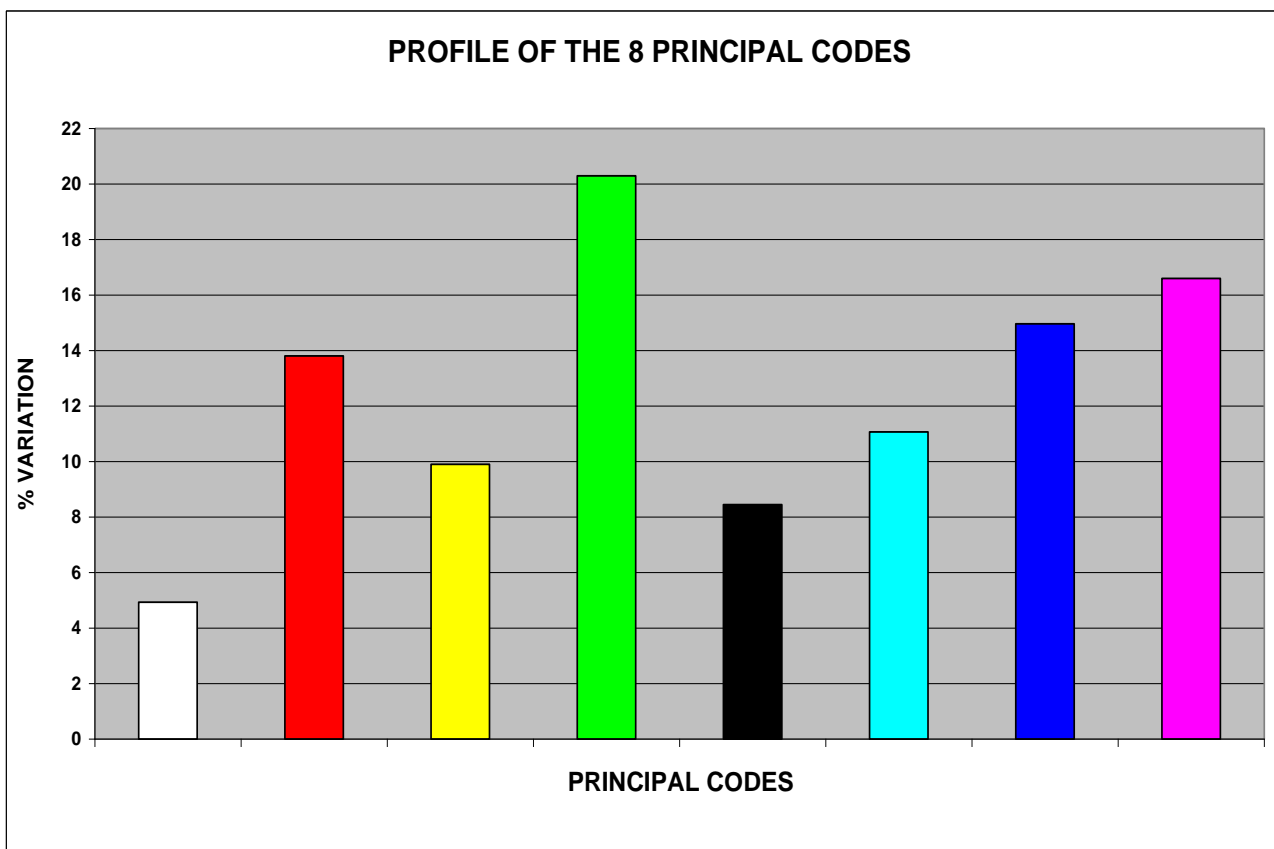
Pct. 13

The **Distribution of the Variation Percentage of the 8 Principal Codes** (see the graph in Pct. 14) is a graph that <u>clearly highlights precise aspects of the "Trend"</u> of the entire sequence.
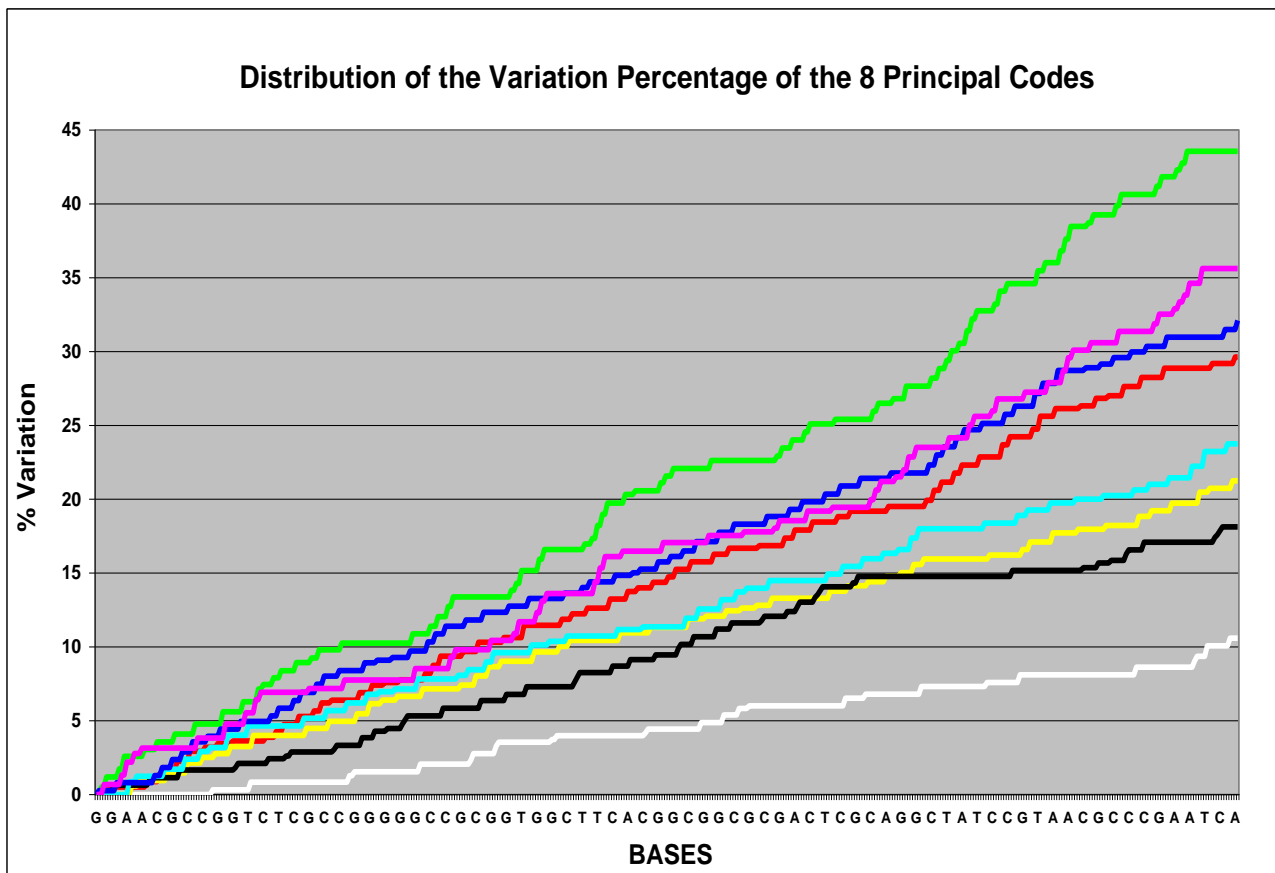
Pct. 14

The graph of the **64 Total Code Tonalities** (see the graph in Pct. 15) highlights the <u>single bases</u> of the entire sequence, while offering very little information on its "Trend".
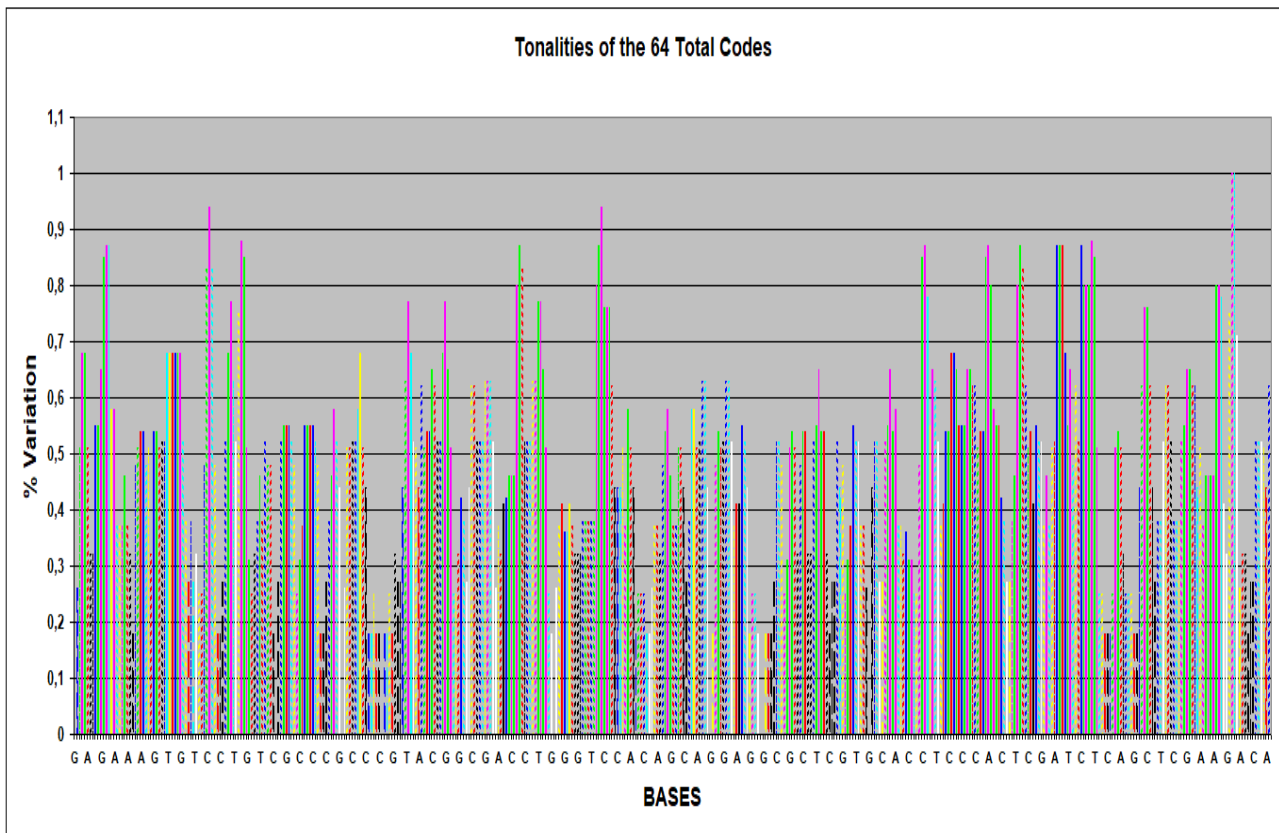
Pct. 15

The **64 Total Code Profile** shown in Pct. 16 constitutes a <u>very specific synthesis </u>of the entire sequence.

PROFILE OF THE 64 TOTAL CODES

%VARIATION — PRINCIPAL AND SECONDARY CODES
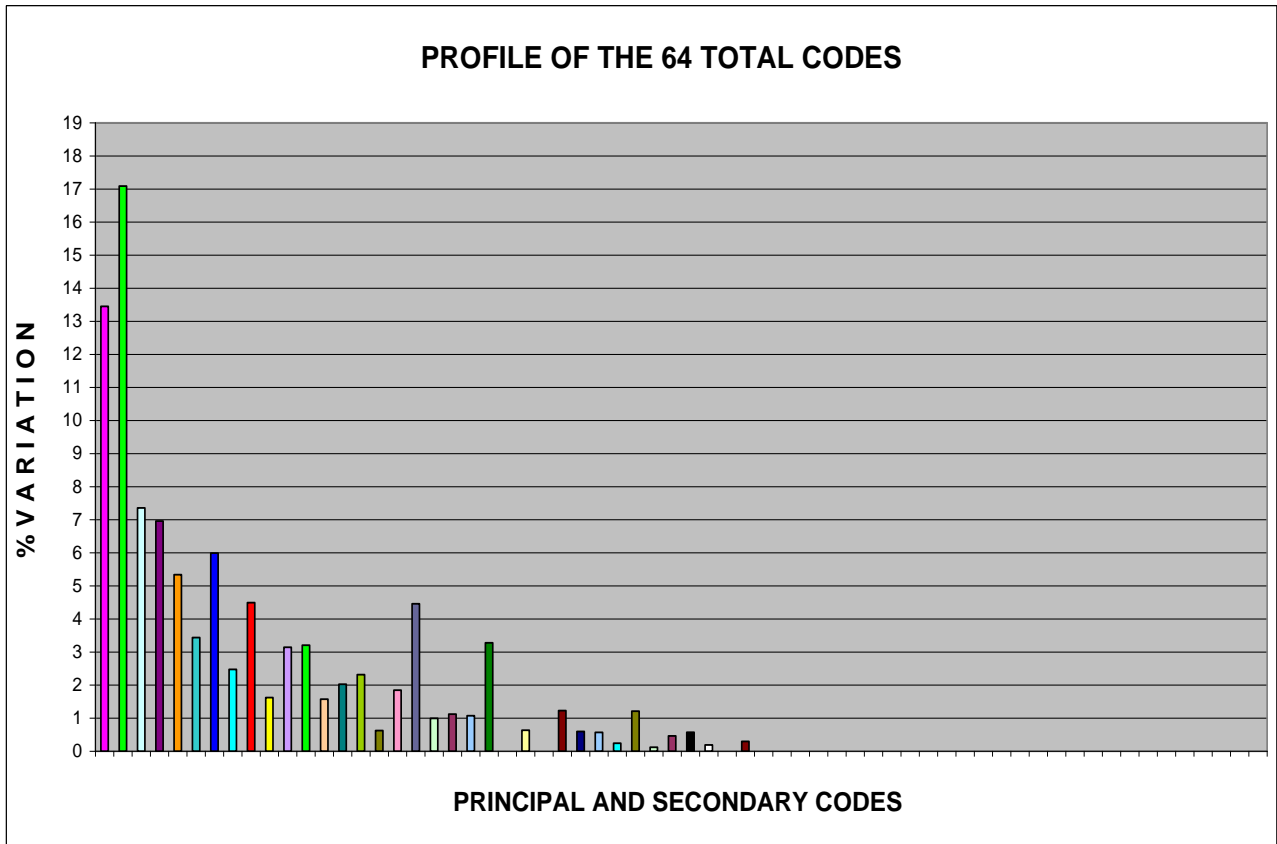
Pct. 16

Comparison and Comment on the Results of the Complete Analyses of the Sequences
XM_011721319.1
XM_011721316.1
XM_011721317.1
XM_011721318.1

2.9 COMPARISON OF THE BASES OF THE SEQUENCES
XM_011721319.1, XM_011721316.1, XM_011721317.1 and XM_011721318.1

In the following page, the **bases** of the *four mRNA sequences* will be compared.

The bases of the sequence **XM_011721319.1** are **presented in all** the other sequences and are reported in **BLACK** on WHITE background.

Differently, the bases added progressively to the sequence **XM_011721319.1** and those that follow it, were highlighted in **COLOURS**.

The bases added to the sequence **XM_011721319.1** and those present in the sequence **XM_011721316.1** were highlighted in GREY and PURPLE .

The bases added to the sequence **XM_011721316.1** and those present in the sequence **XM_011721317.1** were highlighted in RED .

Finally, the bases added to the sequence **XM_011721317.1** and **present only** in the sequence **XM_011721318.1** were highlighted in YELLOW.

XM_011721319.1 PREDICTED: Macaca nemestrina insulin (INS), transcript variant X4, mRNA

Product="**insulin isoform X2**"   Length=**297**

```
ATGGCCCTGTGGATGCGCCTCTTGCCCCTGCTGGCGCTGCTGGCCCTCTGGGGACCTGACCCGGCCCCGG
CCTTTGTGAACCAGCACCTGTGCGGCTCCCACCTGGTGGAAGCTCTCTACCTGGTGTGCGGGGAGCGAGG
CTTCTTCTACACACCCAAGACCCGCCGGGAGGCAGAGGACCCTCAGGGCAGCCTGCAGCCCTTGGCGCTG
GAGGGGTCCCTGCAGAAGCGCGGCATCGTGGAGCAGTGCTGCACCAGCATCTGCTCCCTCTACCAGCTGG
AGAACTACTGCAACTAG
```

XM_011721316.1 PREDICTED: Macaca nemestrina insulin (INS), transcript variant X1, mRNA

Product="**insulin isoform X1**"   Length=**454**

```
GGGACAGGCTGCATCAGAAGAGGCCAGCAAGCAGGTCACTGTCCTTCCGCCATGGCCCTGTGGATGCGCC
TCTTGCCCCTGCTGGCGCTGCTGGCCCTCTGGGGACCTGACCCGGCCCCGGCCTTTGTGAACCAGCACCT
GTGCGGCTCCCACCTGGTGGAAGCTCTCTACCTGGTGTGCGGGGAGCGAGGCTTCTTCTACACACCCAAG
ACCCGCCGGGAGGCAGAGGACCCTCAGGTGGGGCAGGTGGAGCTGGGCGGGGGCCCTGGCGCAGGCAGCC
TGCAGCCCTTGGCGCTGGAGGGGTCCCTGCAGAAGCGCGGCATCGTGGAGCAGTGCTGCACCAGCATCTG
CTCCCTCTACCAGCTGGAGAACTACTGCAACTAGATGCGGCCCGCAGGCGGCCCACACCCTCCACCTCCT
GCACCAAGAGAGATCGAATAAAGCCCTTGAACCA
```

XM_011721317.1 PREDICTED: Macaca nemestrina insulin (INS), transcript variant X2, mRNA

Product="**insulin isoform X1**"   Length=**482**

```
GGGACAGGCTGCATCAGAAGAGGCCAGCAAGCAGGTCTGTTCCAAGGGCCTTCACGTCAGGT
CACTGTCCTTCCGCCATGGCCCTGTGGATGCGCCTCTTGCCCCTGCTGGCGCTGCTGGCCCTCTGGGGGAC
CTGACCCGGCCCCGGCCTTTGTGAACCAGCACCTGTGCGGCTCCCACCTGGTGGAAGCTCTCTACCTGGT
GTGCGGGGAGCGAGGCTTCTTCTACACACCCAAGACCCGCCGGGAGGCAGAGGACCCTCAGGTGGGGCAG
GTGGAGCTGGGCGGGGGCCCTGGCGCAGGCAGCCTGCAGCCCTTGGCGCTGGAGGGGTCCCTGCAGAAGC
GCGGCATCGTGGAGCAGTGCTGCACCAGCATCTGCTCCCTCTACCAGCTGGAGAACTACTGCAACTAGAT
GCGGCCCGCAGGCGGCCCACACCCTCCACCTCCTGCACCAAGAGAGATCGAATAAAGCCCTTGAACCAGC
```

XM_011721318.1 PREDICTED: Macaca nemestrina insulin (INS), transcript variant X3, mRNA

Product="**insulin isoform X1**"   Length=**532**

```
GGGACAGGCTGCATCAGAAGAGGCCAGCAAGCAGGTCTGTTCCAAGGGCCTTCACGTCAGGT
GGGCTCAGGGCTGCCCCACTTGGGGGGTTCCAGGGTGGCTGGACCCCAGGTCACTGTCCTTCCGCC
ATGGCCCTGTGGATGCGCCTCTTGCCCCTGCTGGCGCTGCTGGCCCTCTGGGGACCTGACCCGGCCCCGG
CCTTTGTGAACCAGCACCTGTGCGGCTCCCACCTGGTGGAAGCTCTCTACCTGGTGTGCGGGGAGCGAGG
CTTCTTCTACACACCCAAGACCCGCCGGGAGGCAGAGGACCCTCAGGTGGGGCAGGTGGAGCTGGGCGGG
GGCCCTGGCGCAGGCAGCCTGCAGCCCTTGGCGCTGGAGGGGTCCCTGCAGAAGCGCGGCATCGTGGAGC
AGTGCTGCACCAGCATCTGCTCCCTCTACCAGCTGGAGAACTACTGCAACTAGATGCGGCCCGCAGGCGG
CCCACACCCTCCACCTCCTGCACCAAGAGAGATCGAATAAAGCCCTTGAACCAGC
```

2.10 COMPARISON OF AND COMMENT ON THE GRAPHIC RESULTS OF THE ANALYSES OF THE SEQUENCES XM_011721319.1, XM_011721316.1, XM_011721317.1 and XM_011721318.1

In the next page (see the Pictures17, 18, 19 and 20) the **8 Principal Code Profiles** of the four sequences taken as case- study in this chapter will be compared.

As it has been stated in paragraf 2.2, the **8 Principal Code Profile** *SYNTHETISES IN AN UNSPECIFIC MANNER* the joint contribution of the *single bases* (to a minor extent) and that of the *"Trend"* (to a greater extent) in *characterising* the sequence.

Such a graph is very useful to compare the *general characteristics* of the four analysed sequences (above all, in relation to the role carried out by the "Trend").

At page 42, in **only one graph** (see Pct. 21), the four **8 Principal Code Profiles** (those related to the Pictures 17, 18, 19 and 20) were compared.

For the graph to be immediately clear, in Picture 21, the **Sequence XM_011721319.1** has been denominated **9**; the **Sequence XM_011721316.1** has been denominated **6**; the **Sequence XM_011721317.1** has been denominated **7**; and, finally, the **Sequence XM_011721318.1** has been denominated **8**.

As it can be ascertained by obeserving the Pct. 21, *the sequences XM_011721316.1 and XM_011721317.1 are much more similar, than the other sequences when compared to each other.*

Notwistanding such a similarity of the *sequences XM_011721316.1 and XM_011721317.1*, the differences with the other sequences are minimal, as one should expect from a graph which *synthetises in an unspecific manner the characteristics of the sequences.*

In conclusion, the comparisons of the **8 Principal Code Profiles** of the four analysed sequences, highlight that the *sequences XM_011721316.1* (Pct. 18) and *XM_011721317.1* (Pct. 19) are those that have **much more similar** *general characteristics* (above all, in reference to the role carried out by the "Trend") than the other sequences when compared to each other. Conversely, the comparisons of the **8 Principal Code Profiles** of the four analysed sequences, highlight that the *sequences XM_011721319.1* (Pct. 17) and *XM_011721318.1* (Pct. 20) are those that have **much less similar** *general characteristics* (above all, in reference to the role carried out by the "Trend") than the other sequences when compared to each other.

At page 43 (see the Pictures 22, 23, 24 and 25) the **Profiles of the Distribution of the Variation Percentage of the 8 Principal Codes** of the four analysed sequences were compared.

Such a graph is very useful to compare the *characteristics of the "Trends"* of the four analysed sequences.

As it is possible to ascertain by observing the Pictures 22, 23, 24 and 25, *the sequences XM_011721316.1* (Pct. 23) *and XM_011721317.1* (Pct. 24) *are much more similar than other sequences when compared to each other.*

While this result was quite predictable, the differences noticed between the *sequence XM_011721318.1* (Pct. 25) and respectively the *sequences XM_011721316.1* (Pct. 23) *and XM_011721317.1* (Pct. 24) are striking.

The reason behind those differences cannot be but ascribed to the "weight" (in influencing the "Trend") of the bases highlghted in <mark>YELLOW</mark> at page 36 and **present only** in the *sequence XM_011721318.1*.

On the contrary, the same relevance, in influencing the "Trend" and in generating differences among sequences, seems not to be taken on by the bases highlighted in <mark>RED</mark> at page 36, **present** in the *sequence XM_011721317.1* and **absent** in the *sequence XM_011721316.1*.

In conclusion, the comparisons of the **Profiles of the Distribution of the Variation Percentage of the 8 Principal Codes** of the four analysed sequences, highlight that the *sequences* **XM_011721316.1** (Pct. 23) and **XM_011721317.1** (Pct. 24) are those that have **much more similar** *characteristics of "Trends"* than other sequences when compared to each other. Conversely, the comparisons of the **Profiles of the Distribution of the Variation Percentage of the 8 Principal Codes** of the four analysed sequences, highlight that the *sequences XM_011721319.1* (Pct. 22) and *XM_011721318.1* (Pct. 25) are those that have **much less similar** *characteristics of "Trends"* than other sequences when compared to each other.

At page 44 (see the Pictures 26, 27, 28 and 29) the **Profiles of 64 Total Code Tonalities** of the four analysed sequences were compared.

Such a graph is very useful to compare the "*quality*" (Tonality and % of Variation) of the *single bases* of the four analysed sequences.

Being the longest of the four, the *sequence XM_011721318.1* contains **all the bases** that are also present in the other three sequences.

For such a reason, the graph in Pct. 29, compared to the **Profiles of the 64 Total Code Tonalities**, represents the graph of reference for the comparison with all the others.

As it can be ascertain (even if the visive discrimination is much harder than the other types of graphic profiles) by observing the Pictures 26, 27, 28 and 29, the *major differences,* as expected, are noted between the sequence *XM_011721319.1* (Pct. 26) and the sequence *XM_011721318.1* (Pct. 29). Such a result is understandable, being the sequence *XM_011721319.1* the one with the fewest number of bases.

On the contrary, the difference between the sequences *XM_011721316.1* (Pct. 27) *and XM_011721317.1* (Pct. 28) is much smaller.

In conclusion, being the difference of the number of bases minimal between the sequences *XM_011721316.1* (Pct. 27) *e XM_011721317.1* (Pct. 28), the comparisons of the **Profiles of the 64 Total Code Tonalities**, highlight that the *sequences XM_011721316.1* and *XM_011721317.1* are those that have "*qualities*" (Tonality e % of Variation) **much more similar** than other sequences when compared to each other.

Conversely, being the difference of the number of bases at a maximum between the *sequences XM_01172131**9**.1* (Pct. 26) *e XM_01172131**8**.1* (Pct. 29), the comparisons of the **Profiles of the 64 Total Code Tonalities**, highlight that the *sequences XM_01172131**9**.1* and *XM_01172131**8**.1* are those that have "*qualities*" (Tonality e % of Variation) **much less similar** than other sequences when compared to each other.

At page 45 (see the Pictures 30, 31, 32 and 33) the **64 Total Code Profiles** were compared. Such a graph *SYINTHETISES IN AN UNSPECIFIC MANNER* and in the best possible way the joint contribution of the *single bases* (to a minor extent) and that of the *"Trend"* (to a major extent) in *characterising* the sequence.

Such a graph is very useful to compare the *identifying characteristics* of the four analysed sequences.

A pagina 46, in **only one graph** (see the Pct. 34), the four **64 Total Code Profiles** (those related to the Pictures 30, 31, 32 and 33) were compared.

In Picture 34, to make visually discernable the histograms of each sequence graph, it has been assigned to the **Sequence XM_01172131**9**.1** (already denominated **9**) the **BLACK** colour; to the **Sequence XM_01172131**6**.1** (already denominated **6**) the **YELLOW** colour; to the **Sequence XM_01172131**7**.1** (already denominated **7**) the **RED** colour; and finally, to the **Sequence XM_01172131**8**.1** (already denominated **8**) the **WHITE** colour.

Surprisingly, it can be ascertained by obeserving the Pct. 34, the differences among the four sequences *__seem__* truly **minimal**.

Such a result was not very predictable because the comparisons among the **Profiles of the Distribution of the Variation Percentage of the 8 Principal Codes** (see Pictures 22, 23, 24 and 25) and, above all, those among the **Profiles of the 64 Total Code Tonalities** (see Pictures 26, 27, 28 and 29) highlighted important differences among sequences. To clarify better the result of the graph comparisons of the four **64 Total Code Profiles** (those related to Pictures 30, 31, 32 and 33), two more graphs were elaborated. These are shown in Pictures 35 and 36.

Pct. 35 compares the profiles of the sequences **XM_01172131**6**.1** (Pct. 31) and **XM_01172131**7**.1** (Pct. 32). As it can be ascertained by observing Pct. 35, the differences between these two sequences are **substantially neglegible.** Such a piece of information confirms that the results already highlighted by other types of graphs, that is to say that *the sequences XM_01172131**6**.1 and XM_01172131**7**.1 are much more similar* (also in relation to the 64 Total Codes) *than other sequences when compared to each other.*

On the contrary, Pct. 36 compares the profiles of the sequences **XM_01172131**9**.1** (Pct. 30) and **XM_01172131**8**.1** (Pct. 33). As it can be ascertain by observing the Pct. 36, the differences between these two sequences are **more noticeable**, even if smaller than what one would have expected. Hence, this data also confirms the results that other types of graphs have already highlighted.

In conclusion, the comparisons of the **Profiles of the 64 Total Codes** of the four analysed sequences, highlight that:
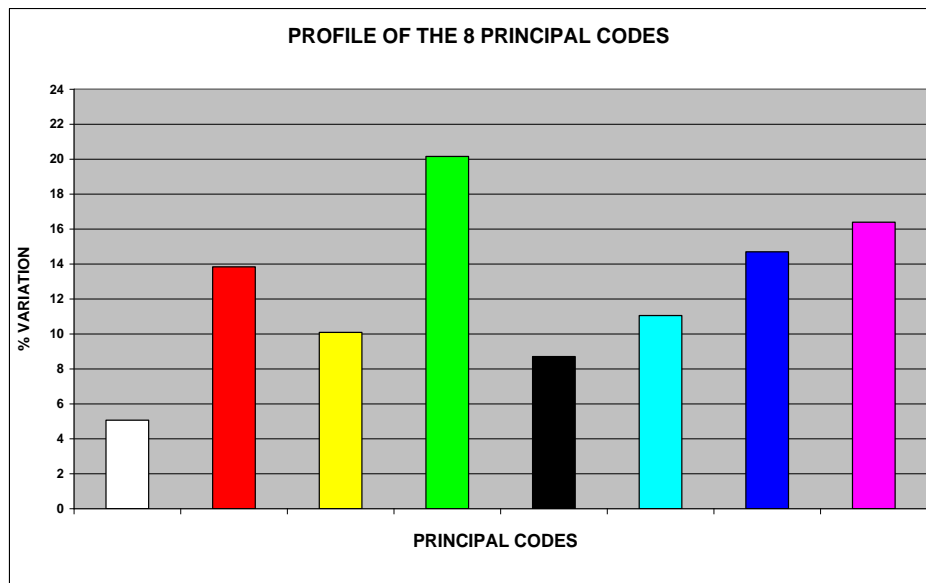
1) the *sequences XM_011721316.1* and *XM_011721317.1* are those that have **much more similar** *identifying characteristics* than other sequences when compared to each other;

2) the *sequences XM_011721319.1* and *XM_011721318.1* are those that have **much less similar** *identifying characteristics* than other sequences when compared to each other;

3) the **Profile of the 64 Total Codes** seems to be capable, more than expected, of *SYNTHETISING* and sharply picking the *SPECIFIC SIMILAR CHARACTERISTICS* (in this case, the specific similar characteristic was represented by their *products* being "*insulin isoform*") of sequences that can have a different **quantity** and "**quality**" of bases. Such a conclusion concerning the **Profile of the 64 Total Codes** has anyway been verified also by the analyses of the graphic results considered in Chapter Fist (First and Second Part), even if there, they were not really emphasised. Indeed, the 19 new sequences obtained by manipulating the Insulin A Chain (original sequence) had more than 73 % of different bases (different for typology and position taken in the sequence) from the original sequence, but the **Profiles of the 64 Total Codes** of the *new obtained sequences* **did not differ very much** from those of the Insulin Chain.
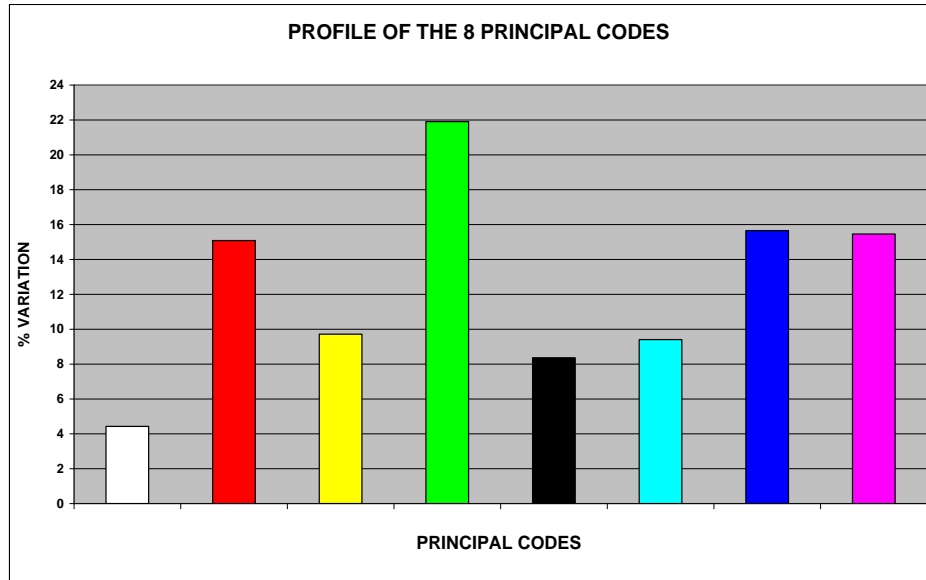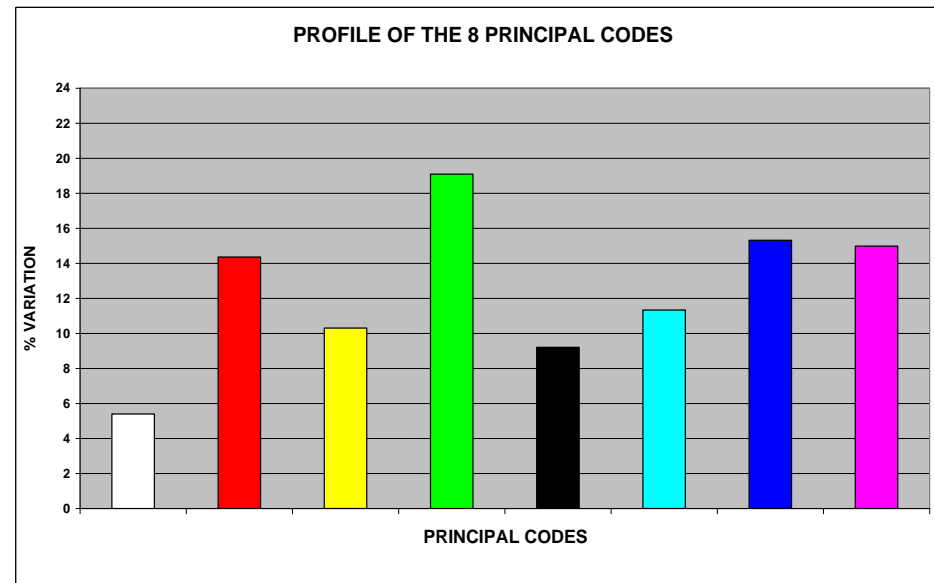
Pct. 17 (Sequence XM_011721319.1)
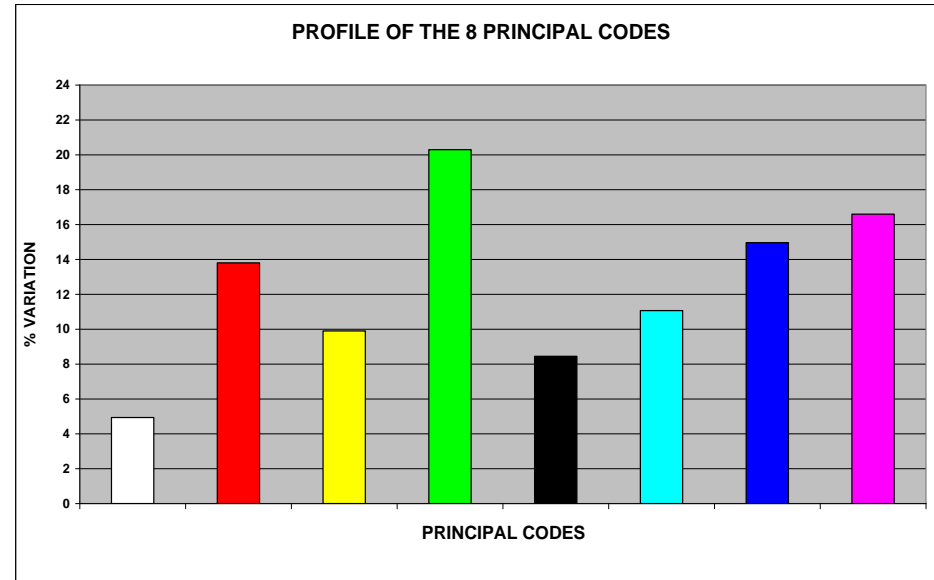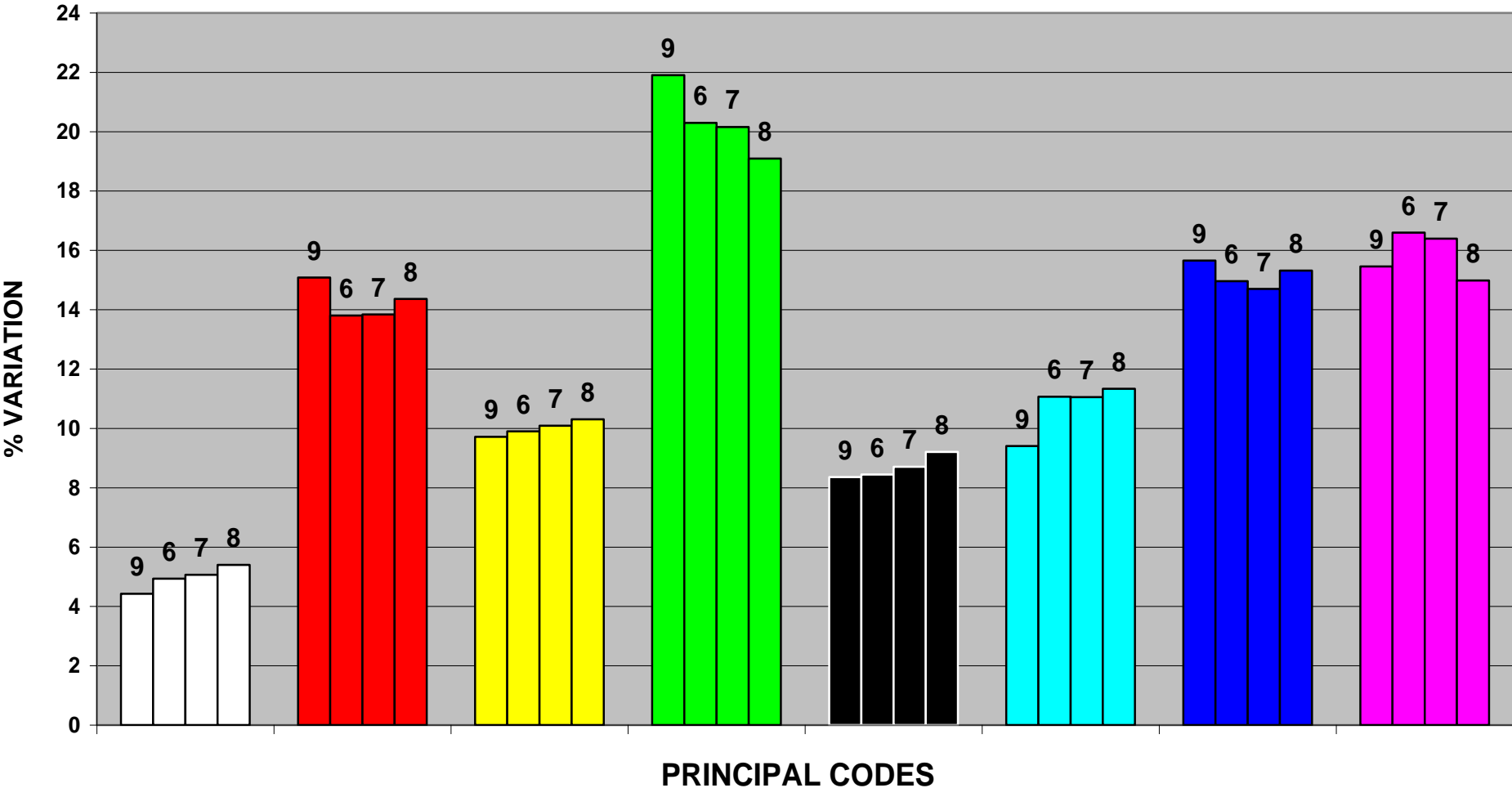

Pct. 18 (Sequence XM_011721316.1)
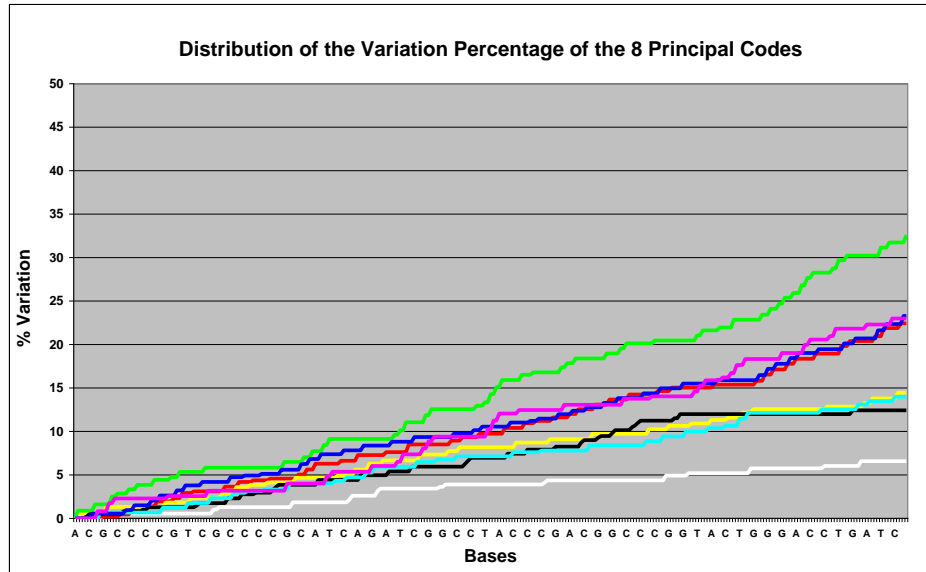

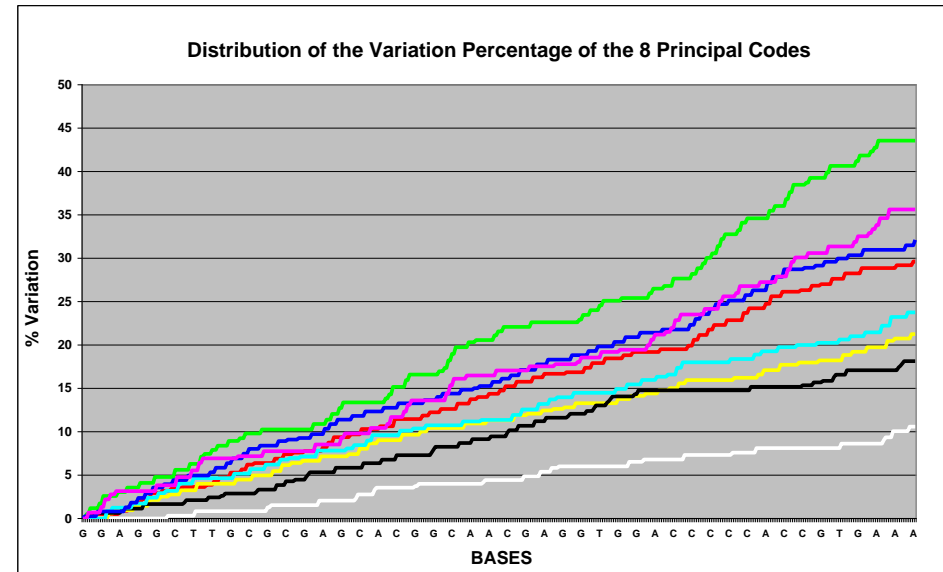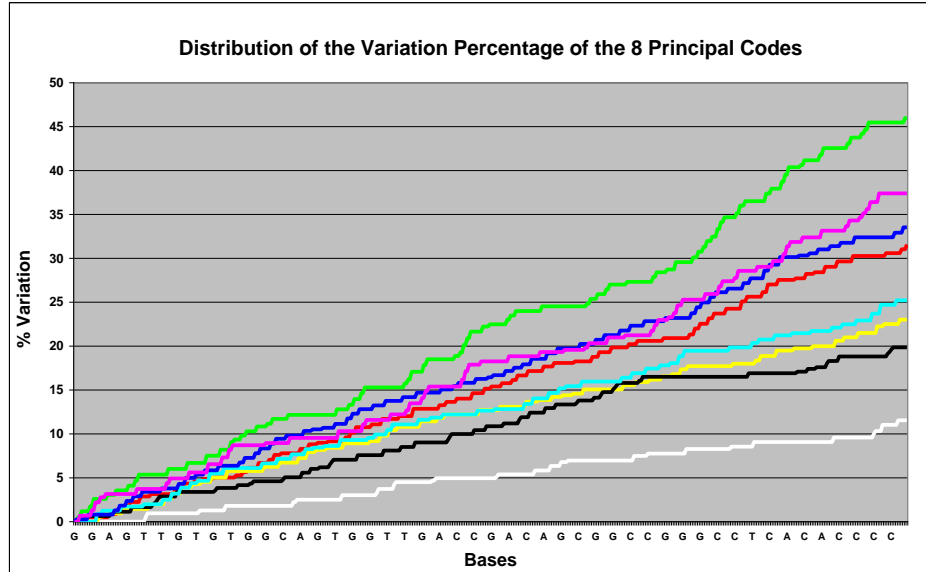**Pct. 19 (Sequence XM_011721317.1)**

Pct. 20 (Sequence XM_011721318.1)

41

PROFILE OF THE 8 PRINCIPAL CODES

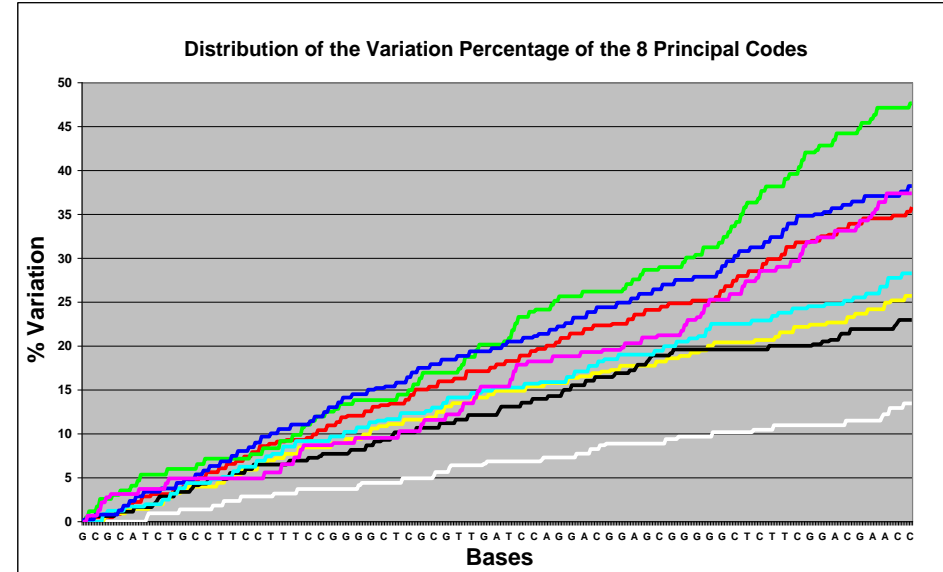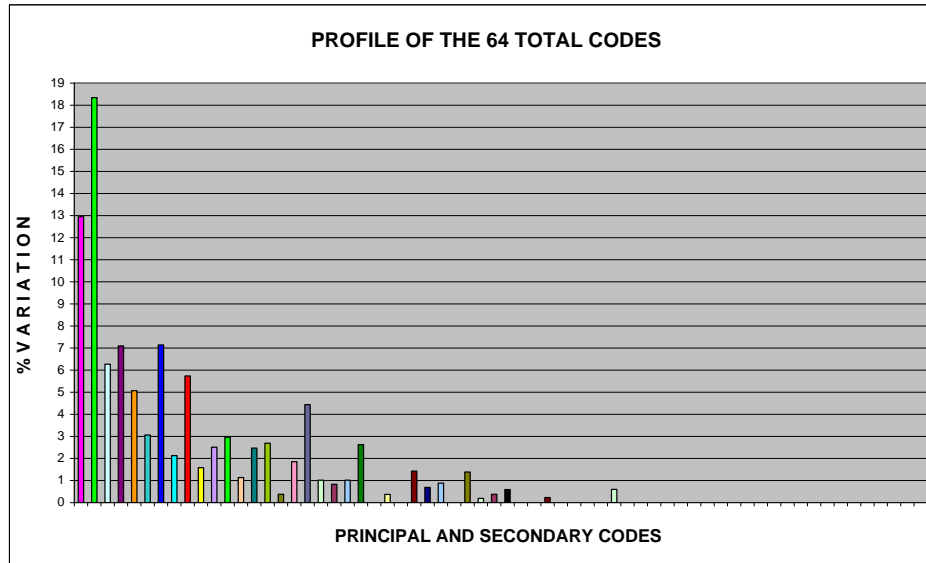

PRINCIPAL CODES

% VARIATION

Pct. 21

42

Distribution of the Variation Percentage of the 8 Principal Codes

Pct. 22 (Sequence XM_0117213199.1)

Pct. 23 (Sequence XM_0117213166.1)

Pct. 24 (Sequence XM_0117213177.1)

Pct. 25 (Sequence XM_0117213188.1)

Pct. 26 (Sequence XM_011721319.1)

Pct. 27 (Sequence XM_011721316.1)

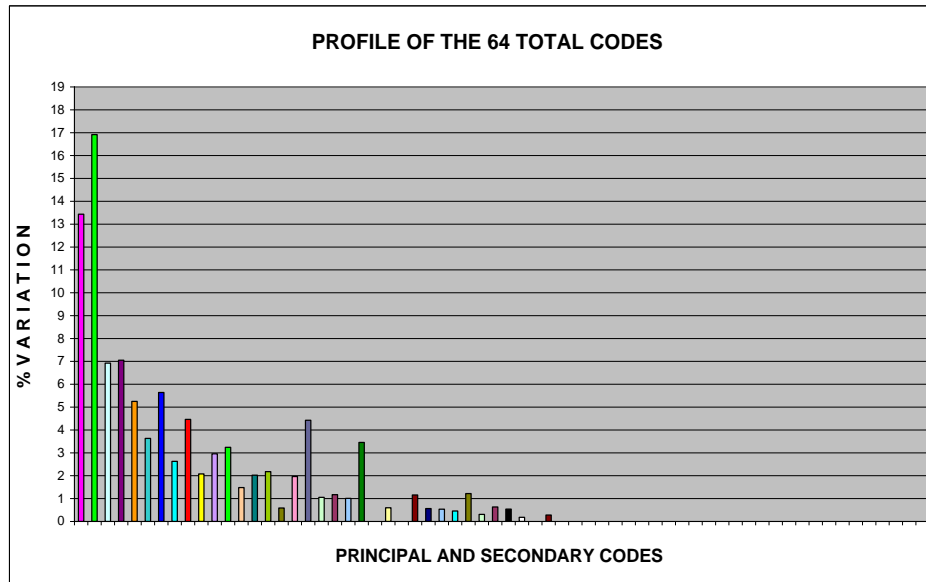**Pct. 28 (Sequence XM_011721317.1)**
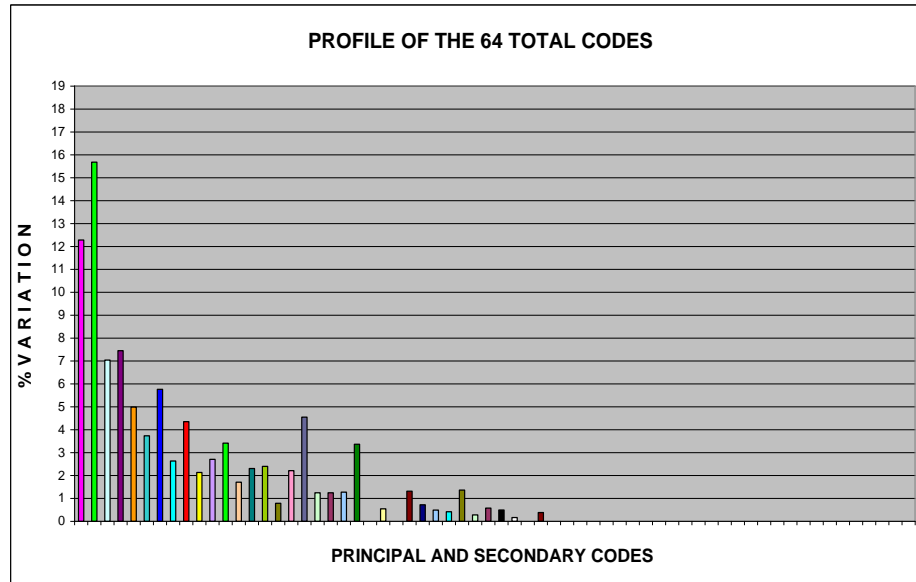

Pct. 29 (Sequence XM_011721318.1)

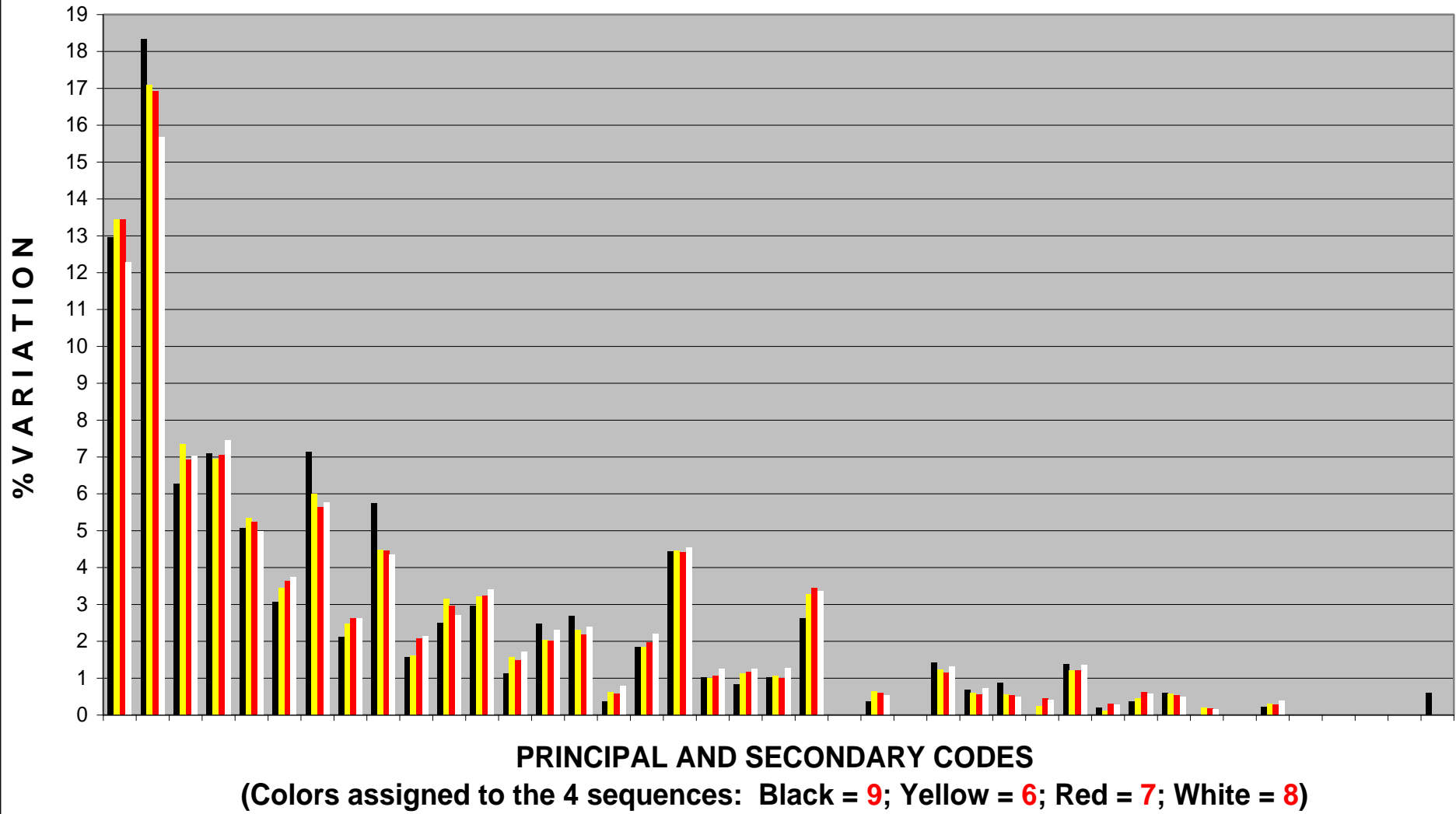**Pct. 30 (Sequence XM_011721319.1)**
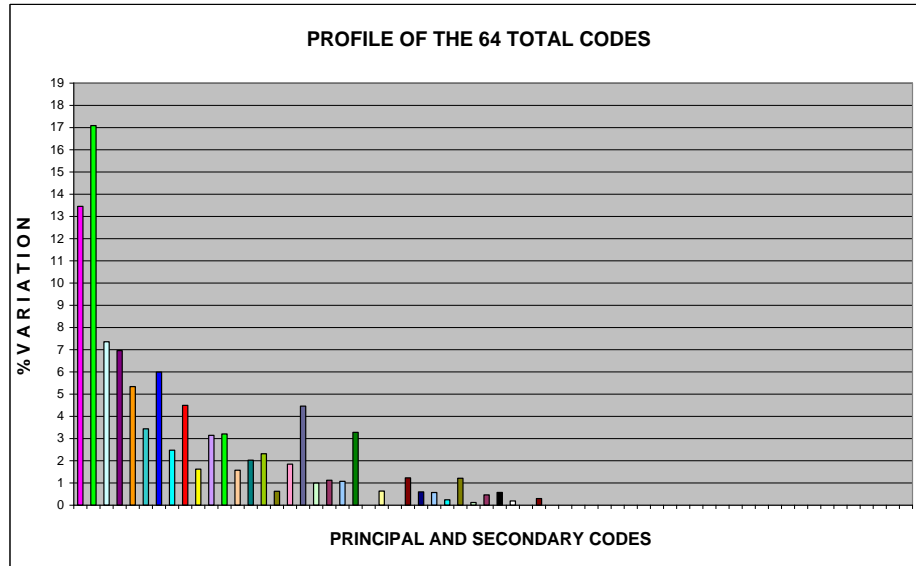
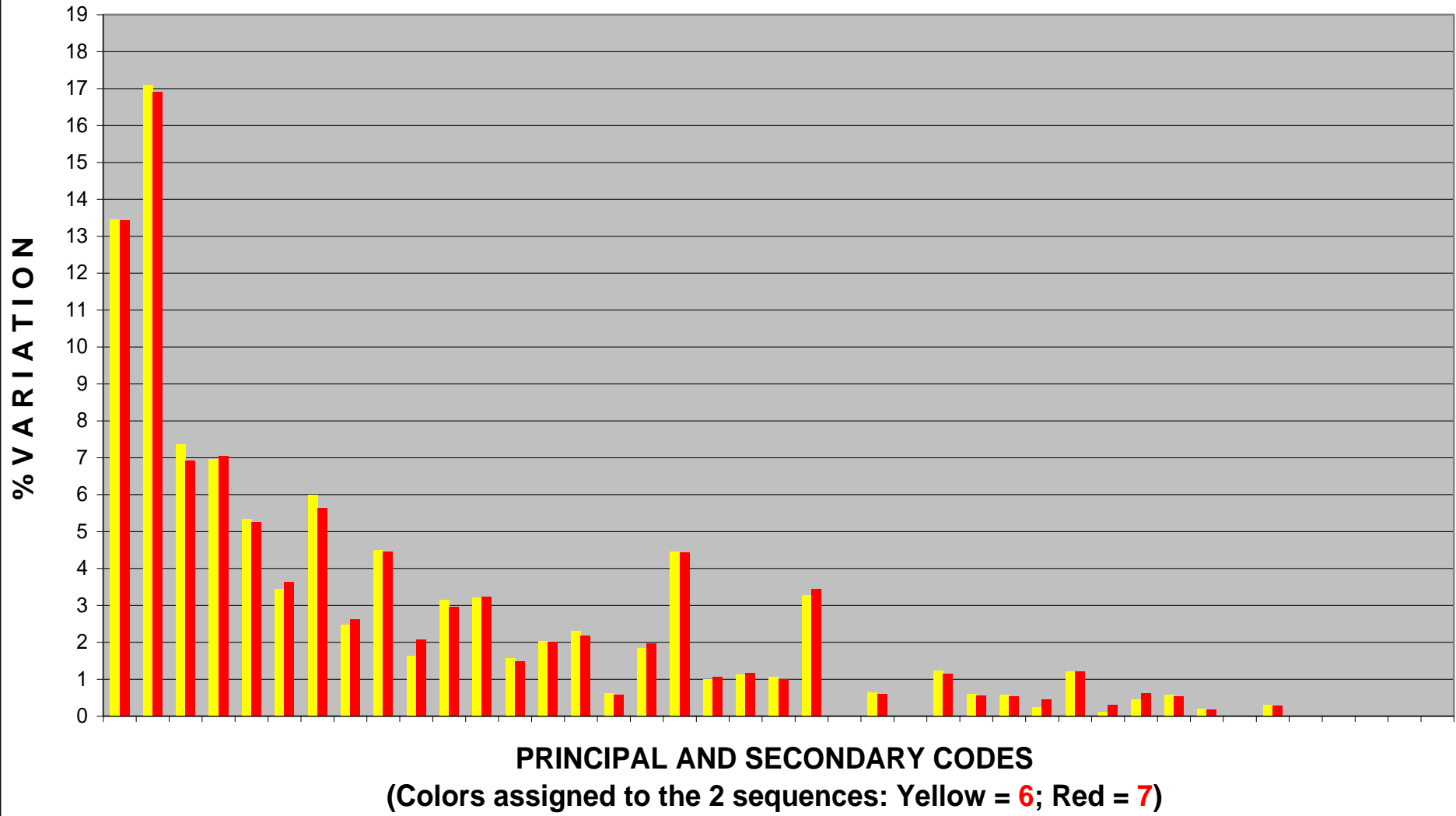**Pct. 31 (Sequence XM_011721316.1)**

Pct. 32 (Sequence XM_011721317.1)

Pct. 33 (Sequence XM_011721318.1)

45

PROFILE OF THE 64 TOTAL CODES

%VARIATION

PRINCIPAL AND SECONDARY CODES
(Colors assigned to the 4 sequences: Black = 9; Yellow = 6; Red = 7; White = 8)

Pct. 34

PROFILE OF THE 64 TOTAL CODES

PRINCIPAL AND SECONDARY CODES
(Colors assigned to the 2 sequences: Yellow = 6; Red = 7)

Pct. 35

PROFILE OF THE 64 TOTAL CODES

%VARIATION

PRINCIPAL AND SECONDARY CODES
(Colors assigned to the 2 sequences: Black = 9; White = 8)
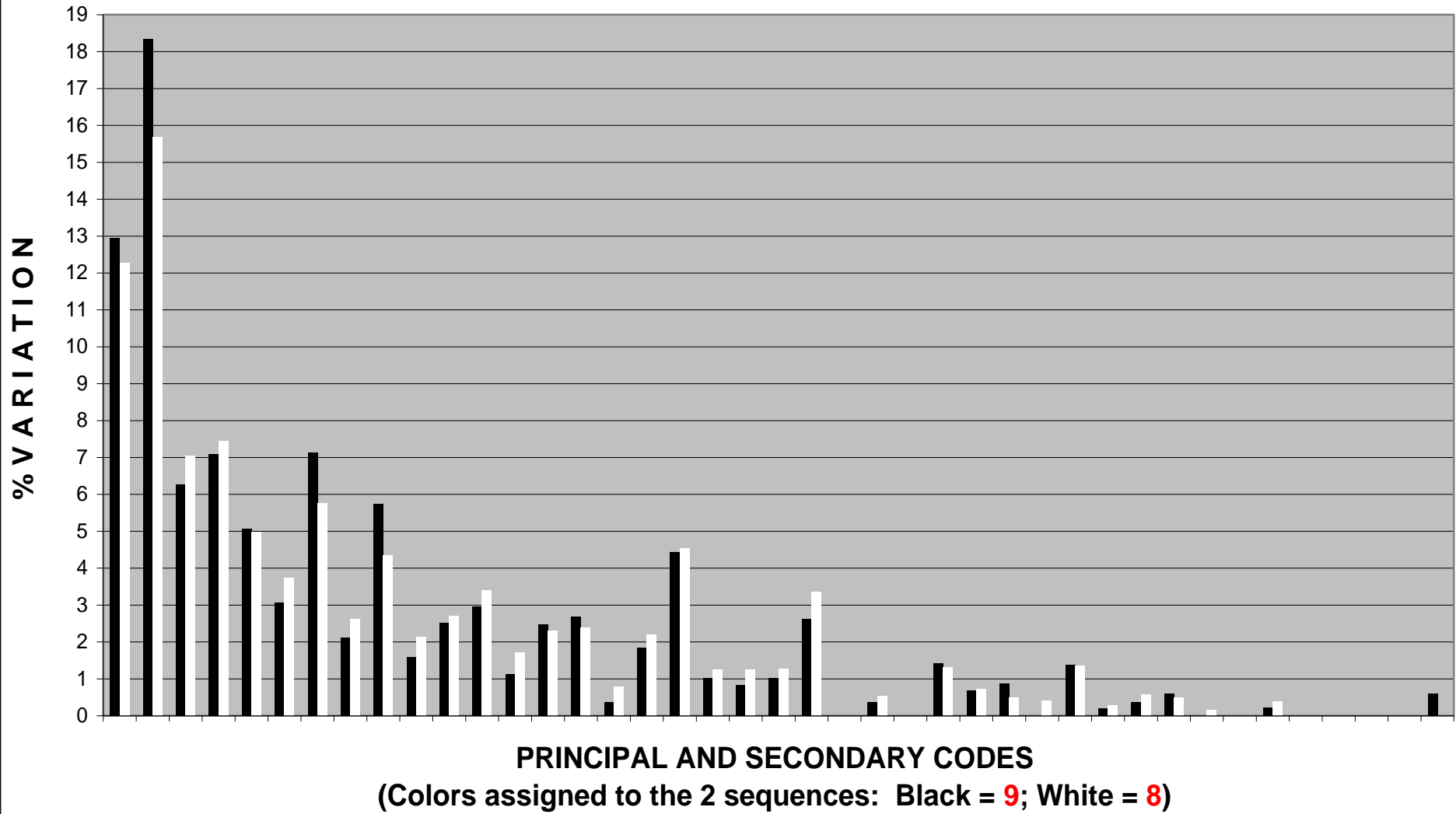
Pct. 36

General Discussion of the Results and Conclusion

2.11 GENERAL DISCUSSION OF THE RESULTS AND CONCLUSION

In this Chapter, we have explored the **potentiality of the graphs** elaborated with the data obtained by the **T.T.E.S.** software and the Excel program.

The goal was to highlight the **validity**, **reliability** and **sensibility** of the calculations made for the generation and the graphic representation of the data.

The graphic results of this chapter, along with those contained in the previous Chapter (First and Second Part), seem to confirm the **validity** of the calculation made.
The relationships among the **mRna sequences** (case study in this Chapter), the calculations made for the generation of all the graphs and their representations were proven *effective*, *not due to chances.*

Furthermore, **reliability**, intended as **internal consistency**, seems to be mantained; indeed, the relationships among the **mRna sequences**, the calculations made for the generation of all the graphs and their representations can still be considered as **consistent,** notwithstanding the *lenght of the sequences, the different types of "trends"* and the *"quality" of the single nitrogenous bases.*

Indeed, the **sensibility** of the graphs, intended as their capacity of discriminating the 4 **mRna sequences** and highlighting the aspects that differ among them, that is to say the different *lenghts* of the analysed sequences, the *different types of "trends"*, the different *"quality" of the single nitrogenous bases* has been vastly proved.

More in specific, the ***Comparison of the Complete Analysis of the 4 Sequences of the Macaca Nemestrina organism***, considered here in this chapter, has blatantly demonstrated that:

1) the sequence XM_011721319.1 (Product = "insulin isoform X2") is graphically rappresented in a **slightly different manner** from the other three sequences (Product = "insulin isoform X1");

2) the sequences XM_011721316.1, XM_011721317.1 and XM_011721318.1 are graphically represented in a **very similar manner,** notwistanding the *quantity* and *"quality" of the single nitrogenous bases, different in part* in each of those sequences.

To be very precise, here, we have ascertained, without a doubt, that the *sequences* XM_011721316.1 and XM_011721317.1 are those that, amongst the others, have **much more similar** *general characteristics* in terms of *"Trends"*, *"quality"* of the bases (Tonality and % of Variation) and *identifying characteristics* than all the other sequences when compared to each other.

Finally, it has also emerged that the **Profile of the 64 Total Codes** seems to be capable, more than expected, of *SYNTHETISING* and sharply picking the *SPECIFIC SIMILAR CHARACTERISTICS* (in this case, the specific similar characteristic was represented by their *products* being "*insulin isoform*") of sequences that can have a different **quantity** e "**quality**" of bases.

In light of these results, we can then conclude that the **graphs** elaborated with the data obtained by the **T.T.E.S.** software and the Excel program are _more than adequate_ to **describe**, in a way that is innovative, clear and comprehensive, a Dna (or Rna) sequence of any lenght.

More precisely, the graphs here elaborated are capable of :

1) synthetising and highlighting the *general characteristics* of a *sequence*, above all in relation to the role carried out by the "Trend" (**Profile of the 8 Principal Codes**);

2) highlighting, in a clear way, the *characteristics* of the *"Trend"* of an entire sequence (**Distribution of the Variation Percentage of the 8 Principal Codes**);

3) highlighting the "*qualities*" (Tonality and % of Variation) of the *single bases* of a sequence (**64 Total Codes Tonalities**);

4) synthetising and highlighting faithfully the *specific and identifying characteristics* of a sequence (**Profile of the 64 Total Codes**);

5) discriminating and highlighting precisely the similar and different aspects of sequences that are partially different or very different from each other.

In conclusion, the **graphs** elaborated with the data obtained by the **T.T.E.S.** software and the Excel program can be considered as a new and important resource, that can surely be enhanced, but of immense value in the hands of researchers of different fields such as molecular biology, genetics, genomics, etc.

END OF CHAPTER II °

www.ttesystems.eu

E-mail: *nunzio.bonaventura@libero.it*

12 Agosto 2019