# 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		3
CHAPTER II°:	Comparison of the Complete Analyses of 4 Sequences of the Macaca Nemestrina organism obtained by the significant alignments of the Insulin A Chain	
Complete Analys PREDICTED: Ma 2.1 Characteristics	is of the Sequence XM_011721319.1 acaca nemestrina insulin (INS), transcript variant X4, mRNA of the Sequence XM_011721319.1	
PREDICTED: 2.2 Graphic results PREDICTED: 2.2 PREDICTED: 2.2 PREDI	Macaca nemestrina insulin (INS), transcript variant X4, mRNA	7 9
Complete Analys PREDICTED: Ma 2.3 Characteristics of PREDICTED: 2.4 Graphic results of PREDICTED: 1	<i>is of the Sequence XM_011721318.1</i> <i>acaca nemestrina insulin (INS), transcript variant X3, mRNA</i> f the Sequence XM_011721318.1 Macaca nemestrina insulin (INS), transcript variant X3, mRNA	14 16
Complete Analys PREDICTED: Ma 2.5 Characteristics on PREDICTED: 2 6 Graphic results of	<i>is of the Sequence XM_011721317.1</i> <i>acaca nemestrina insulin (INS), transcript variant X2, mRNA</i> of the Sequence XM_011721317.1 Macaca nemestrina insulin (INS), transcript variant X2, mRNA	21
PREDICTED:	Macaca nemestrina insulin (INS), transcript variant X2, mRNA	23
<b>Complete Analys</b> <b>PREDICTED:</b> Ma 2.7 Characteristics of	is of the Sequence XM_011721316.1 Acaca nemestrina insulin (INS), transcript variant X1, mRNA of the Sequence XM 011721316.1	
PREDICTED: 2.8 Graphic results of PREDICTED	Macaca nemestrina insulin (INS), transcript variant X1, mRNA	28 30
Comparison and	Comment on the Results	50
of the Complete A XM_011721319.1, X 2.9 Comparison of XM_01172131 2.10 Comparison an results of the c XM_01172131	Analyses of the Sequences   XM_011721316.1, XM_011721317.1 and XM_011721318.1   The bases of the Sequences   9.1, XM_011721316.1, XM_011721317.1 and XM_011721318.1   The bases of the Sequences   9.1, XM_011721316.1, XM_011721317.1 and XM_011721318.1   Part of the Sequences   9.1, XM_011721316.1, XM_011721317.1 and XM_011721318.1   Part of the Sequences   9.1, XM_011721316.1, XM_011721317.1 and XM_011721318.1	35 37

#### *General Discussion of the Results and Conclusion* 2.12 General discussion of the results and conclusion \_\_\_\_\_\_ pag. 49

#### **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 <u>www.ttesystems.eu</u>).

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° - 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 <mark>2/1 3/1 8/1 17/1</mark>	PREDICTED: <mark>Macaca</mark> nemestrina insulin (INS), transcript variant X4, mRNA	110	110	100%	6e- 21	98%	<u>XM_011721319.1</u>	
XM_011721318.1 <mark>2/1 3/1 8/1 17/1</mark>	PREDICTED: <mark>Macaca</mark> nemestrina insulin (INS), transcript variant X3, mRNA	110	110	100%	6e- 21	98%	<u>XM 011721318.1</u>	
XM_011721317.1 2/1 3/1 8/1 17/1	PREDICTED: <mark>Macaca</mark> nemestrina insulin (INS), transcript variant X2, mRNA	110	110	100%	6e- 21	98%	<u>XM_011721317.1</u>	
XM_011721316.1 <mark>2/1 3/1 8/1 17/1</mark>	PREDICTED: <mark>Macaca</mark> nemestrina insulin (INS), transcript variant X1, mRNA	110	110	100%	6e- 21	98%	<u>XM_011721316.1</u>	

**<u>ATTENTION</u>**: 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) <u>XM\_011721319.1</u> Product = "insulin isoform <u>X2</u>";
- 2) <u>XM\_011721318.1</u> Product = "insulin isoform X1";
- 3) <u>XM\_011721317.1</u> Product = "insulin isoform X1";
- 4)  $\underline{XM}_{011721316.1}$  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:

- 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  $T_TES$  (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	<u>GenBank</u>	<b>Graphics</b>	<u>FASTA</u>			
Score			Expe	ct Id	entities	Gaps	Strand	
110 bit	ts(121	1)	7e-2	1 62	2/63(98%)	0/63(0%)	Plus/Plus	
Query	1	GGCAT	CGTGGAGCAG	төстөсасси	ΑGCATCTGTTCCCTCTAC	AGCTGGAGAACTACTG	60	
Sbjct	232	ĠĠĊĂ	rcataaaacaa	toctockck	AdcAtctdctcctctacd	AGCTGGAGAACTACTG	291	
Query	61		63					
Sbjct	292	ÅÅĊ	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 mRNA.	a insulin (INS), transcript variant X4,								
ACCESSION	XM 011721319									
VERSION	XM_011721319.1									
DBLINK	BioProject: PRJNA279145									
KEYWORDS	RefSeq.									
SOURCE	Macaca nemestrina (pig-tailed macaque)									
ORGANISM	Macaca nemestrina									
	Eukarvota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;									
	Mammalia; Eutheria; Euarchor	ntoglires; Primates; Haplorrhini;								
	Catarrhini; Cercopithecidae; Cercopithecinae; Macaca.									
COMMENT	MODEL <u>REFSEQ</u> : This record is predicted by automated computational									
	analysis. This record is der	rived from a genomic sequence								
	( <u>NW_012013911.1</u> ) annotated u	sing gene prediction method: Gnomon,								
	supported by mRNA evidence.									
	Also see:									
	Documentation of NCBI's	Annotation Process								
	##Genome-Annotation-Data-STA	ART##								
	Annotation Provider	:: NCBI								
	Annotation Status	:: Full annotation								
	Annotation Name	:: <u>Macaca nemestrina Annotation Release</u> 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	D##								

FEATURES		Location/Qu	ualifiers					
sour	ce	1297						
		/organism='	'Macaca neme	estrina"				
		/mol type='	'mRNA"					
		/isolate="N	495218"					
		/db xref="t	axon:9545"					
		/chromosome	e="Unknown"					
		/sex="femal	Le"					
		/tissue typ	pe="blood"					
gene		1297						
		/gene="INS'	•					
		/note="Der	ived by auto	omated compu	utational an	nalysis using		
		gene predio	ction method	d: Gnomon. S	Supporting e	evidence		
		includes st	imilarity to	o: 1 mRNA, 1	l Protein, a	and 17%		
		coverage of	f the annota	ated genomic	c feature by	y RNAseq		
		alignments'	•					
		/db xref="0	GeneID:1054	59786 <b>"</b>				
CDS		1297						
		/gene="INS"						
		/codon_star	ct=1					
		/product="insulin isoform X2"						
		/protein_id=" <u>XP_011719621.1</u> "						
		/db_xref="0	GeneID:1054	59786 <b>"</b>				
		/translatio	on="MALWMRL]	LPLLALLALWG	PDPAPAFVNQHI	LCGSHLVEALYLVCG		
		ERGFFYTPKT	RREAEDPQGSL	QPLALEGSLQK	RGIVEQCCTSI	CSLYQLENYCN"		
STS		1296						
		/gene="INS"						
		/standard_r	name="PMC123	3023P3"				
		/db_xref="U	JniSTS: <u>27042</u>	24"				
STS		114295						
		/gene="INS'	•					
		/standard_r	name="Ins1"					
		/db_xref="U	JniSTS: <u>26700</u>	<u>)3</u> "				
STS		139290						
		/gene="INS'	•					
		/standard_r	name="PMC246	544P6"				
		/db_xref="U	JniSTS: <u>26549</u>	94"				
ORIGIN								
1	atggccctgt	ggatgcgcct	cttgcccctg	ctggcgctgc	tggccctctg	gggacctgac		
61	ccggccccgg	cctttgtgaa	ccagcacctg	tgcggctccc	acctggtgga	agctctctac		
121	ctggtgtgcg	gggagcgagg	cttcttctac	acacccaaga	cccgccggga	ggcagaggac		
181	cctcagggca	gcctgcagcc	cttggcgctg	gaggggtccc	tgcagaagcg	cggcatcgtg		
241	gagcagtgct	gcaccagcat	ctgctccctc	taccagctgg	agaactactg	caactag		
11								

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.

- 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. <u>https://www.ncbi.nlm.nih.gov/pmc/articles/PMC146917/</u>
- (2) National Center for Biotechnology Information (NCBI)[Internet]. Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information; [1988]. Available from <u>https://www.ncbi.nlm.nih.gov/</u>

#### 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 <u>lenght</u>; 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

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** is not limited by the sequence lenght; 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** is not limited by the lenght 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*.

Attention: 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** is not limited 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*.

Attention: 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<sup>o</sup>- Second Part, ["DNA or RNA Sequence Analysis and Modification through  $\underline{T} \underline{T} \underline{E} \underline{S}$  (Chapter I<sup>o</sup> - 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 Graphice	EVET	٨			
Score	10 457 <u>Gen</u>		t T	raji dentitic	<u>A</u>	Gans	Strand	
110 bits(12	21)	7e-2	1 6	2/63(9	8%)	0/63(0%)	Plus/Plus	
Query 1	GGCATCGT	GGAGCAG	тостосьс		TOTICCCTCTAC		ACTGC 60	
Query 1	ĬĬĬĨĬĬĬĬ	ĬĬĨĬĬĨĬ						
50]00 595	GGCATCGT	GGAGCAG	IGCIGCACO	AGCATO		LAGCIGGAGAACIA	40100 454	
Query 61	AAC 63							
Sbjct 455	AAC 457							
PREDICTE	D: Ma	icaca	nemest	trina	a insulin	(INS),	transcript	
variant	<mark>X3</mark> , mRN	IA	(NCBI	Refe	rence Sequ	ence: XM_0	11721318.1)	
		21.0		5 3 A J			4 3 3 3 0 0 1 0	
DEFINITION	REDICTED	318 : Macaca	nemestr	532 bp ina in	o mRNA L usulin (INS).	inear PRI 2 transcript va	4-APR-2018 riant X3.	
222 2112 2 2011	mRNA.	• 1140400			(110),,	oranoorrpo ta		
ACCESSION	XM_011721	318						
DBLINK	BioProject	318.1 t: prinz	279145					
KEYWORDS	RefSeq.	<u></u>	11.0110					
SOURCE	Macaca ner	mestrina	a (pig-ta	iled m	lacaque)			
ORGANISM	Eukarvota	, Metazo	a: Chord	lata; C	raniata; Vert	ebrata; Eutel	eostomi;	
	Mammalia;	Euther:	La; Euarc	hontog	lires; Primat	es; Haplorrhi	ni;	
CONGUENE	Catarrhin	i; Cerco	opithecid	lae; Ce	rcopithecinae	; Macaca.		
COMMENT	analysis.	<u>SEQ</u> : Tr This re	nis recor ecord is	derive	edicted by a deno	nutomated comp mic sequence	outational	
	( <u>NW_01201</u>	3911.1)	annotate	d usin	ng gene predic	tion method:	Gnomon,	
	supported	by mRNA	A evidenc	e.				
	Docum	entation	n of NCBI	's Ann	otation Proce	SS		
			-					
	##Genome-J	Annotat: n Provid	ion-Data- Jor	START#	H NCBT			
	Annotatio	n Status	3	::	Full annotati	on		
	Annotatio	n Name		::	Macaca nemest	rina Annotati	on Release	
	Annotatio	n Versi	מר		101			
	Annotatio	n Pipel:	Lne	::	NCBI eukaryot	ic genome ann	otation	
					pipeline			
	Annotatio:	n Softwa n Method	are Versi M	on ::	8.0 Best-placed R	efSea: Gnomon		
	Features 1	Annotate	- ed	::	Gene; mRNA; C	DS; ncRNA		
	##Genome-	Annotat	ion-Data-	END##				
FEATURES		Location	n/Oualifi	ers				
source		1532						
		/organis	sm="Macac	a neme	strina"			

			/mol_type="	'mRNA" 195218"							
			/db xref="t	axon • 9545"							
			/chromosome="Unknown"								
			/sex="female"								
			/tissue typ	be="blood"							
a	rene		1532								
<u>_</u>			/gene="INS'	,							
			/note="Deri	ved by auto	omated compu	utational an	nalysis using				
			gene predic	tion method	d: Gnomon.	Supporting e	evidence				
			includes si	milarity to	o: 2 mRNAs,	10 Proteins	s, and 9%				
			coverage of	the annota	ated genomia	c feature by	y RNAseq				
			alignments'	'							
			/db_xref="G	GeneID: <u>1054</u>	69786"						
S	TS		8254								
			/gene="INS'	1							
			/standard_r	name="GDB:18	31496"						
			/db_xref="U	JniSTS: <u>1552</u> 4	<u>48</u> "						
C	DS		128460	_							
			/gene="INS'								
			/codon_star	rt=1							
			/product="]	Insulin iso	torm XL"						
			/protein_ic	a = "XP UII/IS	<u>9620.1</u> "						
			/db_xrel="0	eneid: <u>1054</u>	09/00 	עוסואזים גם גםם					
			/ LIANSIALIC	DI MALWMRL	UPLLALLALWG		CGSHLVEALILVCG				
			LAGEFILENIE VOLENVON!!	(KEAEDFQVGQ)	VELGGGFGAG51	гдглятгезгді	VKGIVEQCCISICSL				
q	2TTS		128459								
	10		/gene="INS"	,							
			/standard r	ame="PMC123	3023P3"						
			/db xref="l	JniSTS:27042	24"						
S	TS		241.458								
_			/gene="INS'	,							
			/standard r	name="Ins1"							
			/db_xref="U	JniSTS:26700	)3 <b>"</b>						
S	TS		266453								
			/gene="INS'	1							
			/standard_r	name="PMC246	644P6"						
			/db_xref="U	JniSTS: <u>26549</u>	<u>94</u> "						
S	TS		315532	_							
			/gene="INS'								
			/standard_r	name="GDB:1	79433"						
ODICIN	,		/db_xrei="(	JniSTS: <u>15504</u>	46"						
ORIGIN	1	agaaaagat	aastasassa	200000000000000000000000000000000000000	aaaaatatat	+	ttagataga				
	⊥ 61	gygacagget	gcatcagaag	ayyccaycaa	geaggietge	tecaagggee	atcactatog				
	121	ttcccccato	ggetgeteta	tacacetett	accetacta	acactactaa	ccctctggg				
	181	acctgacccg	accccaacct	ttataacca	gcacctgtag	gagetgetgg	taataaaaac				
	241	tetetaceta	atatacaaaa	agcgaggdt+	cttctacaca	cccaagaccc	accaadaaaac				
	301	agaggaccct	caqqtaaaac	aggtggagct	dadcdagaaac	cctqqcqcaq	gcagcctgca				
	361	gcccttggca	ctqqaqqqqt	ccctgcagaa	gcgcggcatc	gtggagcagt	gctgcaccag				
	421	catctgctcc	ctctaccage	tggagaacta	ctgcaactag	atgcggcccg	caggeggeee				
	481	acaccctcca	cctcctgcac	caagagagat	cgaataaagc	ccttgaacca	gc				
//			_		2						

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</u> <u>synthesis</u> of the entire sequence.



**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 <u>single bases</u> 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 nemestri	na insulin (INS), t	transcript varian	t X2, mRNA	
Sequence ID:	XM_011721317.1	2/1 3/1 8/1 17/	1	,	
Product="insi	ulin isoform X1" Number of Matche	xe: 1			
Range 1: 345	to 407 GenBank	Graphics FAS	ТА		
Score	Ex	pect Ident	ities	Gaps	Strand
110 bits(12	21) 70	-21 62/6	3(98%)	0/63(0%)	Plus/Plus
Query 1	GGCATCGTGGAGC	АӨТӨСТӨСАССАӨС	ΑΤΑΤΑΤΑΤΑ	ACCAGCTGGAGAACTA	астос 60
Sbjct 345	GGCATCGTGGAGC	AGTGCTGCACCAGC	Atctdctccttt	ACCAGCTGGAGAACT	ACTGC 404
Query 61	AAC 63				
Sbjct 405	AAC 407				
PREDICTE	D: Macaca	a nemestrin	a insulin	(INS), tr	anscript
variant	X2, mRNA	(NCBI Refe	erence Sequ	ence: XM 011	721317.1)
		400.1		_	
DEFINITION	PREDICTED: Maca	482 i ca nemestrina :	p mrna i Insulin (INS),	inear PRI 24-A. transcript varia	unt X2,
	mRNA.		· · · ·	±	
ACCESSION	XM_011721317 XM_011721317_1				
DBLINK	BioProject: PRO	NA279145			
KEYWORDS	RefSeq.				
ORGANISM	Macaca nemestri	na (pig-tailed	macaque)		
01(011(1011	Eukaryota; Meta	zoa; Chordata;	Craniata; Vert	ebrata; Euteleos	tomi;
	Mammalia; Euthe	ria; Euarchonto	oglires; Primat	es; Haplorrhini;	
COMMENT	MODEL REFSEO:	This record is	predicted by a	; Macaca. nutomated computa	tional
001111111	analysis. This	record is deriv	ved from a gend	omic sequence	0101041
	( <u>NW 012013911.1</u>	) annotated us:	ing gene predic	tion method: Gno	omon,
	supported by ma	NA evidence.			
	Documentati	on of NCBI's An	notation Proce	ess	
	##Genome-Annota	tion-Data-STAR	r##		
	Annotation Prov	rider :	NCBI		
	Annotation Stat	us :	: Full annotati	.on	
	Annotation Name		: <u>Macaca nemest</u> 101	rina Annotation	Release
	Annotation Vers	ion :	101		
	Annotation Pipe	line :	NCBI eukaryot	ic genome annota	tion
	Annotation Soft	ware Version :	: 8.0		
	Annotation Meth	iod :	Best-placed F	RefSeq; Gnomon	
	Features Annota	ted :	Gene; mRNA; C	CDS; ncRNA	
FEATURES	##Genome-Annota Locati	on/Oualifiers	t		
source	1482	, gaarriet			
	/orgar	ism="Macaca ner	nestrina"		
	/mol_t	ype="mKNA"			

			/isolate="N	195218"						
			/db_xref="t	axon • 9545"						
			/chromosome	="Unknown"						
			/cnromosonic	o"						
			/sex- remai	e 						
			/LISSUE_LYP	pe- prood						
(	gene		1482							
			/gene="INS'	•						
			/note="Deri	ved by auto	omated compu	utational ar	nalysis using			
			gene predio	ction method	d: Gnomon. S	Supporting e	evidence			
			includes si	milarity to	o: 3 mRNAs,	10 Proteins	s, and 10%			
			coverage of	the annota	ated genomi	c feature by	y RNAseq			
			alignments'	1						
			/db xref="0	GeneID:10546	59786 <b>"</b>					
	STS		8204							
-			/gene="INS"	,						
			/standard r	ame="GDB · 18	81496"					
			/db_vrof="I	Iniere.1552/	18"					
,	ana		700_XIEI= 0	<u>111515.</u> <u>1552</u>	10					
-	CDS		/8410							
			/gene="INS'							
			/codon_stai	rt=1						
			/product="i	nsulin iso	Eorm X1"					
			/protein_id=" <u>XP_011719619.1</u> "							
			/db_xref="GeneID: <u>105469786</u> "							
			/translation="MALWMRLLPLLALLALWGPDPAPAFVNQHLCGSHLVEALYLVCG							
			ERGFFYTPKTF	REAEDPQVGQV	/ELGGGPGAGSI	LQPLALEGSLQF	KRGIVEQCCTSICSL			
			YQLENYCN"							
	STS		78409							
-			/gene="INS"	,						
			/standard r	ame="PMC121	302323"					
			/db_vrof="I	IniSTS • 27043	24"					
	CTC		101 /08							
-	010		(mama_UTNC)							
			/gene="INS"							
			/standard_r	name="insi"						
			/db_xref="l	JniSTS: <u>26/00</u>	<u>)3</u> "					
1	STS		216403							
			/gene="INS'	1						
			/standard_r	name="PMC246	544P6"					
			/db xref="l	JniSTS:26549	94"					
	STS		265482							
-			/gene="INS'	•						
			/standard r	ame="GDB:1"	79433"					
			/db_vref="I	InisTS • 15504	16"					
ORIGI	N		/ db_Aici (	<u>111010.</u>	10					
01(101)	1	aaaaaaaat	aastasaasa	20000000000	aggagtatat	+	ttabartaba			
	£ 1	gggacagget	ycaccayaay ttaacaata	aggecageaa	taaggicigi	racatata	reacter			
	101	gleactglee	LLCCGCCalg	geeelgigga	Lycycclcll	geeeetgetg	gegelgelgg			
		ccctctgggg	acctgacccg	yccccggcct	LIGTGAACCA	ycacctgtgc	yyctcccacc			
	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	qc		-			-			
11		-								

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 <u>very general</u> <u>synthesis</u> 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.



Pct. 10

The graph of the **64 Total Code Tonalities** (see the graph in Pct. 11) highlights the <u>single bases</u> 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.



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**).

PREDIC Sequence Product= Length= Range 1	PREDICTED: Macaca nemestrina insulin (INS), transcript variant X1, mRNA Sequence ID: <u>XM_011721316.1</u> 2/13/18/117/1 Product="insulin isoform X1" _ength= <mark>454</mark> Number of Matches: 1 Range 1: 319 to 381 <u>GenBank</u> Graphics FASTA								
Score 110 bi	ts(12	1)	E) 76	e-21	Identities 62/63(98%)	Gaps 0/63(0%)	Strand Plus/Plus		
Query	1	GGCA	TCGTGGAG				c 60		
Sbjct	319	ĠĠĊĂ	t c á t á á á á	AGTGCTGCA	ĊĊĂĠĊĂŦĊŦĠĊŦĊĊĊŦĊŦĂĊ	CAGCTGGAGAACTACTG	Ċ 378		
Query Sbjct	61 379	aac 111 aac	63 381						

PREDICTE	ED: Macaca	nemestr	rin	a insu	ulin	(INS)	), 1	transcript	
variant	X1, mRNA	(NCBI R	efe	rence	Sequ	ence:	XM_01	1721316.1)	
LOCUS DEFINITION	XM_011721316 PREDICTED: Maca mRNA.	4: ca nemestrin	54 b na i	p mRN nsulin (	NA 1 (INS),	inear transcr	PRI 24 ript var	-APR-2018 iant X1,	
ACCESSION	XM_011721316 XM_011721316_1								
DBLINK KEYWORDS	BioProject: <u>PRJ</u> RefSeq.	NA279145							
SOURCE	Macaca nemestri	na (pig-tai)	led	macaque)	)				
ORGANISM	Eukaryota; Meta Mammalia; Euthe Catarrhini; Cer	zoa; Chordat ria; Euarcho copithecidae	ta; onto e; C	Craniata glires; ercopith	a; Vert Primat necinae	cebrata; ces; Hap e; Macac	Eutele Dorrhin	ostomi; i;	
COMMENT	MODEL <u>REFSEQ</u> : analysis. This ( <u>NW 012013911.1</u> supported by mR Also see: Documentati	This record record is de ) annotated NA evidence	is eriv usi • s An	predicte ed from ng gene notatior	ed by a a genc predic	automate omic sec otion me	ed compu quence ethod: G	tational	
	##Genome-Annota	tion-Data-S	TART	##					
	Annotation Prov	ider	::	NCBI					
	Annotation Stat	us	::	Full ar	inotati	.on		. Delesse	
	Annotation Name		::	Macaca 101	nemest	rina An	INOTATIO	n Release	
	Annotation Vers	ion	::	101					
	Annotation Pipe	line	::	NCBI eu pipelir	ıkaryot ne	ic genc	ome anno	tation	
	Annotation Soft	ware Version	n ::	8.0					
	Annotation Meth	od	::	Best-pl	laced F	RefSeq;	Gnomon		
	Features Annota	ted	::	Gene; n	nRNA; C	CDS; ncF	RNA		
	##Genome-Annota	tion-Data-EN	ND##						

FEATURES		Location/Qua	lifiers						
sour	ce	1454							
		/organism="Macaca nemestrina"							
		/mol type="mRNA"							
		/isolate="M95218"							
		/db xref="taxon:9545"							
		/chromosome=	"Unknown"						
		/sex="female	"						
		/tissue_type="blood"							
gene		1454							
		/gene="INS"							
		/note="Derive	ed bv auto	mated compu	utational ar	nalvsis using			
		gene prediction method. Chomon Supporting evidence							
		includes sim	ilarity to	5 mRNAs,	10 Proteins	s. and 11%			
		coverage of	the annota	ited genomic	feature by	z RNAseq			
		alignments"	0110 01110 00	gonomize	2000010 2	i iuniooq			
		/db xref="Gei	neTD:10546	9786"					
STS		8178							
<u></u>		/gene="INS"							
		/standard na	me="GDB:18	1496"					
		/db xref="Un	iSTS:15524	8"					
CDS		52 384							
000		/gene="INS"							
		/codon start:	=1						
		/product="in	sulin isof	form X1"					
		/protein id=	"XP 011719	618.1"					
		/db xref="Ge	neTD:10546	9786"					
		/translation	="MALWMRLT	.PT.T.AT.T.AT.WGF	PDPAPAFVNOHI	CGSHLVEALYLVCG			
		FRGFFYTPKTRR	EAEDPOVGOV	TLGGGPGAGSI	OPLALEGSLOP	RGIVEOCCTSICSI			
		YOLENYCN"		1100001011001					
STS		52 383							
<u></u>		/gene="INS"							
		/standard na	me="PMC123	1023P3"					
		/db_xref="IIn	ists:27042	· · · · · · · · · · · · · · · · · · ·					
STS		165382	<u> </u>						
<u></u>		/gene="INS"							
		/standard na	me="Insl"						
		/db_xref="Un	ists:26700	13"					
STS		190377	<u> </u>						
<u></u>		/gene="INS"							
		/standard na	me="PMC246	44P6"					
		$/ \text{Standard}_\text{Haller} = \text{FPC24044F0}$ /db $\text{wrof}=\text{Hip}(\text{SPC})(265/0/1)$							
ORIGIN		,	<u> </u>						
1	addacaddct	gratcagaag ag	aaccaacaa	acadatcact	atecttecae	catggccctg			
61	tagatacacc	tettaceet a	ctaacacta	ctarcctct	addacctda	cccaacccca			
121	acctttataa	accagcacct g	tacaactcc	cacctonton	aageteteta	cctaatatac			
1.81	geeeegega	acttetteta e	acacccaad	acccaccaa	addcadadda	ccctcadata			
241	addcadatad	auctanacaa a	aaccetaac	acaaacaaaca	tacaaccc++	aacactaaaa			
301	agatectac	agaaggggggggg	atcatagaa	cantactaca	ccarcatctr	ctccctctac			
361	caucturana	actactocaa c	tagatgggug	cccaceaaaca	accacacco	tocacotoot			
421	acaccaadaa	agatogaata a	agecettaa	acca	goodacacee				
//	geaceaayay	agatoguata a	agoooccya						

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.

- 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. <u>https://www.ncbi.nlm.nih.gov/pmc/articles/PMC146917/</u>
- (2) National Center for Biotechnology Information (NCBI)[Internet]. Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information; [1988]. Available from <u>https://www.ncbi.nlm.nih.gov/</u>

#### 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</u> <u>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.



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 CTTCTTCTACACACCCCAAGACCCGCCGGGAGGCAGAGGACCCTCAGGGCAGCCTGCAGCCCTTGGCGCTG GAGGGGTCCCTGCAGAAGCGCGGCATCGTGGAGCAGTGCTGCACCAGCATCTGCTCCCTCTACCAGCTGG AGAACTACTGCAACTAG

#### XM\_011721316.1 PREDICTED: Macaca nemestrina insulin (INS), transcript variant X1, mRNA

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

### $\rm XM\_011721317.1$ PREDICTED: Macaca nemestrina insulin (INS), transcript variant $\rm X2$ , mRNA

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

GGGACAGGCTGCATCAGAAGAGGCCAGCAAGCAGGT<mark>CTGTTCCAAGGGCCTTCACGTCAGGT</mark>

CACTGTCCTTCCGCCAGGCCCTGTGGATGCGCCTCTTGCCCCTGCTGGCGCTGCTGGCCCTCTGGGGAC CTGACCCGGCCCCGGCCTTTGTGAACCAGCACCTGTGCGGCTCCCACCTGGTGGAAGCTCTCTACCTGGT GTGCGGGGAGCGAGGCTTCTTCTACACACCCAAGACCCGCCGGGAGGCAGAGGACCCTCAGGTGGGGGCAG GTGGAGCTGGGCGGGGGCCCTGGCGCAGGCAGCCTGCAGCCCTTGGCGCTGGAGGGGTCCCTGCAGAAGC GCGGCATCGTGGAGCAGTGCTGCACCAGCATCTGCTCCTCTACCAGCTGGAGAACTACTGCAACTAGAT GCGGCCCGCAGGCGGCCCCACACCCTCCACCTCCTGCACCAAGAGAGATCGAATAAAGCCCTTGAACCA

#### XM\_011721318.1 PREDICTED: Macaca nemestrina insulin (INS), transcript variant X3, mRNA

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

GGGACAGGCTGCATCAGAAGAGGCCAGCAAGCAGGTCTGTTCCAAGGGCCTTCACGTCAGGT GGGCTCAGGGCTGCCCCACTTGGGGGGTTCCAGGGTGGCTGGACCCAGGTCACTGTCCTCCGCC ATGGCCCTGTGGATGCGCCTCTTGCCCCTGCTGGCGCTGCTGGCCCTCTGGGGGACCTGACCCGGCCCCGG CCTTTGTGAACCAGCACCTGTGCGGCTCCCACCTGGTGGAAGCTCTCTACCTGGTGTGCGGGGGAGCGAGG CTTCTTCTACACACCCAAGACCCGCCGGGAGGCAGAGGGACCCTCAGGTGGGGCCAGGTGGAGCTGGGGCGGG GGCCCTGGCGCAGGCAGCCTGCAGCCCTTGGCGCTGGAGGGGCCCTGCAGAAGCGCGGCGGG AGTGCTGCACCAGCATCTGCTCCCTCTACCAGCTGGAGAACTACTGCAACTAGATGCGGCCCGCAGGCGG CCCACACCCTCCACCTCCTGCAAGAGAGATCGAATAAAGCCCTTGAACCA

#### 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 observing 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 highlighted in <u>YELLOW</u> 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 **RED** 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*  $XM_011721319.1$  (Pct. 26) *e*  $XM_011721318.1$  (Pct. 29), the comparisons of the **Profiles of the 64 Total Code Tonalities**, highlight that the *sequences*  $XM_011721319.1$  and  $XM_011721318.1$  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\_011721319.1** (already denominated 9) the **BLACK** colour; to the **Sequence XM\_011721316.1** (already denominated 6) the **YELLOW** colour; to the **Sequence XM\_011721317.1** (already denominated 7) the **RED** colour; and finally, to the **Sequence XM\_011721318.1** (already denominated 8) the **WHITE** colour.

Surprisingly, it can be ascertained by observing the Pct. 34, the differences among the four sequences <u>seem</u> 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\_011721316.1 (Pct. 31) and XM\_011721317.1 (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\_011721316.1 and XM\_011721317.1 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\_011721319.1 (Pct. 30) and XM\_011721318.1 (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. 19 (Sequence XM\_011721317.1)



Pct. 20 (Sequence XM\_011721318.1)











Pct. 34



Pct. 35



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 <u>more than</u> <u>adequate</u> 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**);
- 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 $^\circ$



### www.ttesystems.eu

E-mail: nunzio.bonaventura@libero.it

12 Agosto 2019