

# Chimeric mitochondrial peptides from contiguous regular and swinger RNA

# Hervé Seligmann

Unité de Recherche sur les Maladies Infectieuses et Tropicales Émergentes, Faculté de Médecine, URMITE CNRS-IRD 198 UMER 6236, Université de la Méditerranée, Marseille, France

#### ARTICLE INFO

Article history: Received 8 March 2016 Received in revised form 19 June 2016 Accepted 23 June 2016 Available online 29 June 2016

Keywords: RNA-DNA differences Bijective transformation Nucleotide substitution Proteome Systematic deletions delRNA

# ABSTRACT

Previous mass spectrometry analyses described human mitochondrial peptides entirely translated from swinger RNAs, RNAs where polymerization systematically exchanged nucleotides. Exchanges follow one among 23 bijective transformation rules, nine symmetric exchanges ( $X \leftrightarrow Y$ , e.g.  $A \leftrightarrow C$ ) and fourteen asymmetric exchanges  $(X \rightarrow Y \rightarrow Z \rightarrow X, e.g. A \rightarrow C \rightarrow G \rightarrow A)$ , multiplying by 24 DNA's protein coding potential. Abrupt switches from regular to swinger polymerization produce chimeric RNAs. Here, human mitochondrial proteomic analyses assuming abrupt switches between regular and swinger transcriptions, detect chimeric peptides, encoded by part regular, part swinger RNA. Contiguous regular- and swinger-encoded residues within single peptides are stronger evidence for translation of swinger RNA than previously detected, entirely swinger-encoded peptides: regular parts are positive controls matched with contiguous swinger parts, increasing confidence in results. Chimeric peptides are 200× rarer than swinger peptides (3/100,000 versus 6/1000). Among 186 peptides with >8 residues for each regular and swinger parts, regular parts of eleven chimeric peptides correspond to six among the thirteen recognized, mitochondrial protein-coding genes. Chimeric peptides matching partly regular proteins are rarer and less expressed than chimeric peptides matching non-coding sequences, suggesting targeted degradation of misfolded proteins. Present results strengthen hypotheses that the short mitogenome encodes far more proteins than hitherto assumed. Entirely swinger-encoded proteins could exist. © 2016 The Author. Published by Elsevier B.V. on behalf of Research Network of Computational and Structural

Biotechnology. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

# 1. Introduction

Mitochondrial genomes apparently compensate for their reduced size by cumulating multiple functions for single sequences [12,14]. For example, tDNA, DNA templating for tRNAs, probably functions also occasionally as light strand replication origin [79,80,83,106–108,110]. The complementary strand of tDNA has similar secondary structure formation capacities and might template for additional functional tRNAs with anticodons usually corresponding to the inverse complement of the tRNA's regular anticodon [7,81,82,84]. Mitochondrial tRNA sidearm loops might also function as anticodons, potentially increasing further mitochondrial anticodon repertoires [90,95]. Translation of stop codons also increases protein coding repertoires [81], reassigning stop codons to amino acids [26,84–87,91,7,11,98].

These various mechanisms expand protein-coding potentials of DNA/RNA sequences. Multifunctional sequences, as suggested for tRNA synthetase genes [53,69,70] are presumably relicts of ancient, short protogenomes, plausibly consisting of ancestors of ribosomal RNAs [72,73,111], where sequence multifunctionality was probably

essential. Presumably, alternative codings are relicts of mechanisms that increase sequence multifunctionality.

# 1.1. Swinger polymerization

A further little known phenomenon increases DNA's protein coding repertoire: nucleotide polymerization that systematically exchanges nucleotides. This alters gene and mRNA coding properties. Assuming this phenomenon enables to detect homology relationships of otherwise 'orphan' DNA and RNA sequences. The homology of these orphan sequences had not been determined because these apparently orphan sequences are so much transformed as compared to their 'parent' homologue that homology is undetectable without assuming a systematic exchange between nucleotides, but becomes obvious after taking the systematic exchange(s) into account. These transformations consist of systematic exchanges between nucleotides during DNA or RNA polymerization, producing so-called swinger sequences.

The first described swinger RNAs were from vertebrate mitogenomes, and correspond to a 3'-to -5' inversion, without complementing, of the homologous, template sequence [88,92], also called 'reversing' transformation [27,28]. When considering a specific sequence, this transformation follows the swinger rule  $A \leftrightarrow T + C \leftrightarrow G$  (bijective transformation rule  $\pi$ 9 according to the annotation system in [58]) of the negative

E-mail address: varanuseremius@gmail.com.

http://dx.doi.org/10.1016/j.csbj.2016.06.005

<sup>2001-0370/© 2016</sup> The Author. Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

strand of the specific sequence, one among nine systematic symmetric exchanges, of type X  $\leftrightarrow$  Y, e.g. A  $\leftrightarrow$  C [93]. Fourteen asymmetric exchanges exist, of type X  $\rightarrow$  Y  $\rightarrow$  Z  $\rightarrow$  X, e.g. A  $\rightarrow$  C  $\rightarrow$  G > A [94]. About hundred mitochondrial transcripts corresponding to one of these 23 swinger types have been detected within the human EST database of GenBank, with about twice as many from the nine symmetric exchanges than from the fourteen asymmetric exchanges.

Swinger RNAs matching eleven exchange types were detected within GenBank's EST database, six symmetric, and four asymmetric transformations. Most of these swinger RNAs (obtained by classical Sanger sequencing) are longer than 100 nucleotides and have >90% similarity with the mitogenome if the swinger transformation is assumed over their complete length [90,91]. All 23 swinger types exist in the human mitochondrial transcriptome among short reads produced by RNA seq (Illumina) ([103], data from [30]). Abundances of different swinger types as estimated from GenBank's ESTs (sequenced by classical methods) and next generation massive sequencing (RNA seq) are overall congruent (i.e. Figure 2 in [103]). This congruence between swinger RNA abundances is remarkable for two reasons: first because comparable results are obtained by two independent methods (Sanger versus next generation sequencing); and second because biological samples differ (not the same cells/mitochondrial lines were analyzed). This suggests that mitochondrial swinger transcription is general to mitochondria, not tissue- or line-specific.

Hence sequences potentially template for 23 swinger transformed versions, increasing considerably the potential coding density of any sequence. Swinger DNA was also detected for nuclear and mitochondrial genes [96,97], especially ribosomal RNAs [99], but for now only according to swinger rule  $A \leftrightarrow T + C \leftrightarrow G$ . Swinger sequences detected in Genbank originate from numerous independent research projects and laboratories, only this author describes them as swinger-transformed.

### 1.2. Swinger versus chimeric RNAs and peptides

Some detected sequences are not entirely swinger-transformed, sequences contiguous to the swinger sequence match the untransformed, contiguous DNA template, and hence are regular RNA [100]. These RNAs transcribed partly by regular, and partly by swinger transcriptions, are termed chimeric RNAs [100]. The transition from one to the other part is frequently abrupt, suggesting sudden switches in the polymerization mode of the same polymerase.

Analyzes here search for peptides matching translation of such chimeric RNAs, where contiguous parts of the peptide are translated from regular and swinger parts of the sequence. These peptides are also considered chimeric, and differ from previously described swinger peptides [103] because the latter are only translated from swingertransformed RNA, while chimeric peptides would be transcribed from RNA that is in part regular, and in part swinger-transformed.

The principle according to which chimeric sequences are produced is shown for a specific 120 nucleotides long sequence of the human mitogenome (Fig. 1). The mid-forty nucleotides are swinger transformed according to swinger rule A  $\leftrightarrow$  C + G  $\leftrightarrow$  T (second sequence in Fig. 1). Swinger RNAs consist solely of swinger-transformed regions such as the underlined transformed regions. Chimeric RNAs have at least one of the contiguous, untransformed (5' and/or 3') parts. Swinger peptides are solely translated from swinger-transformed sequences (such as the underlined sequence in Fig. 1). Chimeric peptides are translated from a sequence that stretches over a regular and a swingertransformed RNA region. Only a minority of detected RNA sequences bearing swinger transformations are chimeric, most follow in their entirety a given nucleotide exchange rule [90,91,100]. Detection of chimeric peptides would be evidence independent of previous descriptions of swinger peptides (entirely encoded by swinger-transformed RNA [103]) for translation of swinger-transformed RNA. Much fewer chimeric human mitochondrial RNAs [100] than entirely swingertransformed human mitochondrial RNAs [90,91] have been detected. Hence I expect to detect fewer chimeric peptides than for previous analyzes searching for entirely swinger-encoded peptides.

#### 1.3. Swinger polymerization by regular polymerases?

Swinger polymerizations could result from unusual polymerization modes by regular polymerases because the principle of swinger polymerization does not differ from that of point nucleotide misinsertions. The difference is in the systematic change in templating rules, from f(A,C,G,T) = (A,C,G,T/U) (regular DNA replication/transcription), to a different rule, e.g. A  $\leftrightarrow$  C, which can also be annotated as f(A,C,G,T/U) =(C,A,G,T/U), stressing its systematic, rather than punctual nature. It seems plausible that point nucleotide misinsertions are due to switches to unstable, unusual conformations of polymerases, lasting the time of a misinsertion. Hypothetically, these unstable, misinsertion-inducing conformations are occasionally stabilized, so that the nucleotide exchange corresponding to that misinsertion occurs systematically along the sequence stretch polymerized while the polymerase is in that unusual conformation, producing a swinger DNA/RNA. Swinger RNAs are for now the only evidence indicating the existence of such unusual, stabilized polymerase states.

This hypothesis on polymerase conformations yields two testable predictions. The first prediction is that biochemical parameters experimentally estimated for point misinsertions by polymerases predict properties of swinger sequences. In this respect, the affinity (Km) and Vmax of each of the twelve misinsertions, and the four regular

C	GAT	CAC	AGG	TC	TAT	CA	cc	СТА	TTA	AC	CAC	CTC	ACG	GG	AGC	TCI	CCZ	ATG	CAI	TT	GGT	AT	ΓTT	CGT	CTC	GGG	GGG	TA	TGC	ACC	GCG	ATA	GCA	TTC	GCG	AGA	CGC	стG	GAG	SCC	GGA	GCA	ccc	TAT	GTO	CGCAG
\$0	GAT	CAC	AGG	TC	TAT	CA	CC	CTA	TTA	AC	CAC	CTC	ACG	GG	AGC	TAG	AAC	CGT.	ACG	GG	TTG	CG	GGG	ATG	AGI	FTT	TTT	GC	GTA	CAI	AG	ATA	GCA	TTC	GCG	AGA	CGC	CTG	GAG	CCO	GGA	GCA	CCC	TAT	GT	CGCAG
Γ		Н	*	S	I	Г	•	L	L	т	Т	H	G		s .	*	N	V	R	v	A	(	G	М	S	F	L	R	Т	*	1	4	A	L	R	D	A	G	I	4	G	A	Ρ	Y	v	A
	I	т	G	L	5	5	Ρ	Y	*	Ρ	, I	5	т	G	A	*	Т	Y	G	3	L	R	G	W	V	F	С	2	V	H	*	*	Н	С	Е	Т	I	5	Е	Ρ	Е	Н	Ρ	М	S	
	S	Q	V		Y	Н	Ρ	I	N	1	H	S	R	Е	L	E	F	2	Т	G	С	G	D	E	E	F I	F	A	Y	М	_D	S	I	. 7	· ·	ł	R	W	S	R	S	Т	I	C	2	х
\$	AT	CAC	AGG	TC	TAT	CA	cc	CTA	TTA	AC	CAC	CTC	ACG	GG	AGC	TCG	AAC	CGT.	ACG	GG	TTG	CG	GGG	ATG	AG	TTT	TTT	GC	GTA	CAT	TAT	ATA	GCA	TTC	GCG	AGA	CGC	CTG	GAG	SCC	GGA	GCA	ccc	TAT	GTO	CG
\$	Т	CAC	AGG	TC	TAT	CA	CC	CTA	TTA	AC	CAC	CTC	ACG	GG.	AGC	TCI	AAC	CGT.	ACC	GG	TTG	CG	GGG	ATG	AGI	TTT	TTT	GC	GTA	CAT	TAT	CTA	GCA	TTC	GCG	AGA	CGC	CTG	GAG	CCO	GGA	GCA	ccc	TAT	GT	CGC
\$	(	CAC	AGG	TC	TAT	CA	CC	СТА	TTA	AC	CAC	CTC	ACG	GG	AGC	TCI	CAC	CGT	ACC	GG	TTG	CG	GGG	ATG	AG	rtt'	TTT	GC	GTA	CAI	TAT	CGA	GCA	TTC	GCG	AGA	CGC	CTG	GAG	SCC	GGA	GCA	ccc	TAT	GTO	CGCA
\$		AC	AGG	TC	TAT	CA	CC	CTA	TTA	AC	CAC	CTC	ACG	GG	AGC	TCI	cc	GT.	ACO	GGG	TTG	CG	GGG	ATG	AGI	TTT	TTT	GC	GTA	CAT	TAT	CGC.	AGC	AT	GC	GAG	ACG	GCT	GGF	GC	CGG	AGC	ACC	CTA	TG	ICGCAG

**Fig. 1.** Example of running windows reduced to 120 nucleotides for illustration purposes, for the human mitochondrial genome, and swinger transformation of the mid-third 40 nucleotides. First row: regular genomic sequence; rows starting by \$ are the first 5 running windows, with the mid third swinger transformed according to swinger rule  $A \leftrightarrow C + G \leftrightarrow T$  (as an example). Running windows used for actual analyses are 270 nucleotides long, and transformed according to each of the 23 swinger transformations, along the same principles as shown above for 120 nucleotides. The three peptides translated from the first running window sequence are also indicated, stops codons are translated here as "\*. Analyses searching for mass spectrometry data matching these predicted peptides consider the possibility that any amino acid is integrated at stops (19 possibilities, merging leucine and isoleucine, which are undistinguishable by the sequencing technique used here, because their molecular weights are identical). This means that for a single running window sequence after a single swinger transformation,  $3 \times 19 = 57$  hypothetical peptides are considered. This number is doubled to 114 considering the inverse complement sequence, and multiplied by 23 considering all 23 potential swinger transformations. Hence 2622 peptides are translated from the 23 swinger transformations of each running window sequence indicated by \$. This running window structure enables detection of chimeric peptides where the regular part is translated from the 5', as well as from the 3' side of the swinger part.

#### Table 1

Human mitochondrial peptides detected assuming abrupt switches between regular and swinger parts of RNA, for peptides where each regular and swinger parts have >8 amino acids (mass spectra from [33]). Columns are: 1. Peptide number; 2. swinger type; 3. amino acid inserted at stop(s) ('no' indicates lack of stops); 4. strand and frame; 5. peptide sequence; 6. PSMs; 7. Xcorr; 8. trypsin miscleavage; 9. PEP; 10–13. Positive strand positions of 5' and 3' extreme amino acids of regular and swinger parts of detected peptide; 14. Peptide extremity matching regular transcription. Underlined: peptide swinger part; \*, \$ marks swinger peptide parts covering previously described swinger reads, respectively previously described swinger peptides [103]. Peptides 8 and 9 differ in posttranslational amino acid modifications (not indicated). Highlighted peptide parts match both translations according to vertebrate mitochondrial, and nuclear (standard) genetic code. Peptide parts not highlighted match only translation according to the vertebrate mitochondrial genetic code. For example, peptide 3 could be translated in the cytosol on the base of RNA transcribed from mitochondrial inserts in the nuclear chromosome (numts), peptides 5 could not, as peptides 1 ad 2 because at least one part of the peptide is not compatible with translation according to the nuclear genetic code. Further analyses (see text) show that fewer detected peptides are compatible with the nuclear genetic code than expected by chance, and that more peptides than expected by chance are compatible only with translation according to the mitochondrial genetic code.

1         e         a         0         VXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX	1	2	3	4	5	6	7	8	9	10	11	12	13	14		
2         a.e.         v         1         SCLAPMONUNCSDPURAUM         30.1         0         0.00         7.20         2.20         1.20        <	1	ac	a	r0	YGVSEGLAAPVGAYNVGAFAALYMAANFSFLNQDAVQVADK	3	2.77	0	0.374	3246	3156	3153	3123	5′	29	12
3         acc         r         7         0         MACGGEPURWWCSDNYMMR         (43)         3.06         0         7.45         16188         16118         1711         16118         17118	2	ac	v	r1	SCLLAFLMGSLMLTLIVGLSK	20	3.04	0	0.169	3465	3432	3492	3468	3′	12	9
4         ag         k         2         Accurates Numery Numery Numery         185         3.06         I         0.082         2.289         2.282         2.292         2.291	3	ac	r	r0	AWGGGFDVDWWGSDDIVAMR	43	3.06	0	0.745	16188	16161	16158	16131	5′	10	10
5         ag         f         r         r)         r)         r)         ag         ag <td>4</td> <td>ag</td> <td>k</td> <td>r2</td> <td>AIGKVAFSTSVMLEVMFLVNK</td> <td>185</td> <td>3.06</td> <td>1</td> <td>0.082</td> <td>2289</td> <td>2262</td> <td>2319</td> <td>2292</td> <td>3′</td> <td>11</td> <td>10</td>	4	ag	k	r2	AIGKVAFSTSVMLEVMFLVNK	185	3.06	1	0.082	2289	2262	2319	2292	3′	11	10
6         ag         5         10         CSSIDSTINGQLISUMAK         149         3.8         0         0.755         10.125         10.125         10.101         1	5	ag	f	r1	SYTFPPGSSSVACWLGCSPSPTLTLIFGLSK	49	3.38	0	1.000	3456	3432	3522	3459	3′	9	22
7         as         10         10         10         10         10         13         1370         1377         13844         1370         13         10         32         10         32         10         33         10         33         10         35         10         37         10         38         327         1         0.05         1477         10         35         373         37         10         37         10         37         10         37         10         37         10         37         10         35         10	6	ag	S	r0	GGSPSDSTTSSQQLLSSILWSK	149	3.98	0	0.755	10125	10101	10164	10128	3′	10	12
8         90         10         11         100         11         100         11         100         11         1	7	ag	no	f0		1	2.37	1	0.931	13704	13767	13644	13701	3'	29	12
9         9         10 <td>8</td> <td>at</td> <td>no</td> <td>f1 61</td> <td></td> <td>399</td> <td>2.77</td> <td>1</td> <td>0.196</td> <td>882</td> <td>924</td> <td>855</td> <td>879</td> <td>3'</td> <td>11</td> <td>9</td>	8	at	no	f1 61		399	2.77	1	0.196	882	924	855	879	3'	11	9
10         10         12         10         12         10         13         12         11         12         11         12         12         11         12<	9	dL	no	11 f1		604	2.77	1	0.033	882 1467	924	800 1407	8/9	3'	11	12
12         11         12         13         13         13         13         13         13         13         13         13         13         13         13         14           15         6         7         7         10         TUTUITUTUTUTUTUTUTUTUTUTUTUTUTUTUTUTUTU	10	dL	Р	11 r2	MIPHSPINWDWGGINSSKLPWECLWEQGPEAK	25	4.24	1	0.005	5601	5574	5621	1404 5604	2/	17	12
15         16         2.60         9         0.004         11.41         1985         1164         15         16         10         5         16         10	12	at	y po	12 r2		10	4.54	0	0.830	6033	6006	6063	6036	2/	10	10
14         cp         12         1.4.SLP_OPTIVETSINGSPACE/LOWWERK         2         2.0.0         13.032         1105         11105         21         2         2         0.0.0         105         1105         11105 <th< td=""><td>12</td><td>cσ</td><td>no</td><td>f0</td><td>TLCOCVAHDIRTNPVDFVCDK</td><td>6</td><td>2.17</td><td>1</td><td>0.121</td><td>1344</td><td>1365</td><td>1368</td><td>1404</td><td>5/</td><td>9</td><td>12</td></th<>	12	cσ	no	f0	TLCOCVAHDIRTNPVDFVCDK	6	2.17	1	0.121	1344	1365	1368	1404	5/	9	12
15         c         10         DUNNILLICSAMINATARY A         540         4.89         0         0.633         1125         1101         1161         1128         2         1         1           17         Ct         m         1         DUNNILLICSAMINATARY         8         4.203         0         0.6337         12155         1215         1215         1	14	Cg Cg	k	r2	LIASLPOPTVVPSTMPTISVRSGVLAGCLIGWWKPK	2	2.00	1	0.592	11103	11058	11163	11106	31	16	20
16       c       no       rt       SQTDLITEPPTYPERUMASYNCK       8       2.03       0       0.532       12150       12163       121533       121533       1	15	cg	t	rO	TDNTNHHLTGSAIMTMTAPVK	540	4.89	0	0.653	11625	11601	11661	11628	3'	11	10
17       cf.       m. ft       MWEDLAWEAMSLSSATVCR       880       449       0       0.837       92019       2022       2022       2052       5'       9       16         19       ct       t       0       CACCMANTALALERNANNACES       324       1       0.862       4233       4201       4230       4261       4251       4287       3'       1       10         21       ct       w       10       AMMCGACACCTAMASMCR       13       203       0       0.42       1387       13860       5'       10       10         22       ct       v       N       MMCGACACCTAMASMCRACCMACCTMASMARK       13       231       0       0.423       1321       1380       1521       1520       1524       1523       12       10       10       132       3'       10       1321       1321       131       141       130       1484       132       14       132       14       132       14       132       14       142       424       424       424       424       424       424       424       424       424       424       424       424       424       424       424       424       424       424	16	cg	no	r1	TSQTDLLTDPPITYEFLWAFSVNK\$	8	2.03	0	0.592	12150	12126	12195	12153	3′	9	14
18       et       et <th< td=""><td>17</td><td>ct</td><td>m</td><td>f1</td><td>MWEDLMVEAMNSLSSATVGR</td><td>880</td><td>4.49</td><td>0</td><td>0.837</td><td>1995</td><td>2019</td><td>2022</td><td>2052</td><td>5′</td><td>9</td><td>11</td></th<>	17	ct	m	f1	MWEDLMVEAMNSLSSATVGR	880	4.49	0	0.837	1995	2019	2022	2052	5′	9	11
19         ct         t         7	18	ct	e	r0	GMGPMAYLASLALKENMVNNAEGFK	642	2.47	1	0.182	4233	4209	4206	4161	5′	9	16
20         ct         a         rz         CVNWAKAMIACYESSYNEQR         5         2.76         1         0.32         101         100         9           22         ct         v         r0         MAKOCXCETAMMAKCK         31         3.31         1         0.321         1521         1527         15240         57         1540         57         1541         1571	19	ct	t	fO	NPSLSISVPSTRHVSMPITISSIPPQTTEMCLMK	3	2.54	1	0.866	4290	4350	4251	4287	3′	21	13
21       ct       m       no       AMMCDCAVCCTREMMSMCIK       13       2.03       0       0.21       1314       13890       13887       13860       5       10       9         23       ct       y       fi       DVSCPSSPSSIMUTILISPDLIG/PENDTIANPLINTPY       2       2.27       0       0.44       1329       1327       1365       1332       3       10       13         25       acgt       s       fi       MVCSMMCSGCMPTPCDSRCEPRSACACPCOSTLQAR       38       2.43       1       0.49       1328       1368       1328       3       1       1.20         26       acgt       s       fi       EXEMPSENDAGENCEPREADERCEPCACHONINGR       8       2.43       1       0.49       1252       126       126       12       2       2       2       2       2       12       2       2       2       12       2       2       12       2       2       12       2       2       12       12       2       2       12       2       2       13       1       13       3       3       3       3       3       3       3       3       3       3       3       3       3       3       <	20	ct	a	r2	GVNWAKMNIAGYESSYNEQR	5	2.76	1	0.352	10512	10485	10542	10515	3′	10	10
22       ct       v       v0       NVWSWAVAAMMKGCVCVCMGCHMEMK       31       31       31       31       31       31       31       31       31       327       15276 <td< td=""><td>21</td><td>ct</td><td>m</td><td>r0</td><td>AMMGDCAVCGTEMMSMCIK</td><td>13</td><td>2.03</td><td>0</td><td>0.42</td><td>13914</td><td>13890</td><td>13887</td><td>13860</td><td>5′</td><td>10</td><td>9</td></td<>	21	ct	m	r0	AMMGDCAVCGTEMMSMCIK	13	2.03	0	0.42	13914	13890	13887	13860	5′	10	9
23       ct       y       f1       DVSCRSSPSSUMTLIFSPDLCOPNYTLANPLIMPY       2       22.7       0       0.724       157.11       15801       15684       157.11       3       30       10         25       acgt       5       0       0.464       1325       10.25       10       10.458       13356       1333.5       133.65       133.3       14       24         27       acgt       5       7       0       0.458       138.5       11       0.458       138.2       10       0.458       53.00       0.461       135.00       65.01       64.86       65.88       33       21       10         29       acgt       x       r1       CDREVEWSENSULLISUPCIANTIMETTER       9       2.77       1       0.425       7.67       7452       74.06       74.0       57.09       74.9       3       9       11         31       acgt       n       10       VMVTDLQUSWSPHSYNNYTNR       12       25.31       10.001       150.57       74.05       74.05       74.05       74.15       21.1       13.33       32.08       10       10.268       153.85       18.33       31.0       10       10.33       36.26       13.0       10.268	22	ct	v	r0	NVVWSVAVAAMMKGGVGVGMGGHMEMK	31	3.31	1	0.393	15321	15279	15276	15240	5′	14	13
24       gt       q       r.2       NNLFSLYCYLPQLWMMDPEHMSMALK       1       2.55       00       MCSTMCSGNETPESACHOCSTLQSAR       38       2.44       1       0.494       1329       1287       13365       1332       3'       16       11         26       acgt       b       r.2       SWEMSGSDKTPETCHOSTEGERSEACHOCSTLQSAR       38       2.43       1       0.495       199       2049       220       202       2.00       2.18       4.24       8.383       4.1000       6558       6588       58.58       32       2       1         28       acgt       y       10       EATASSACINASYUCAGGMUNT       8       3.85       1       1.000       6579       7479       747       746       746       745       759       7479       3'       9       11         31       acgt       n       0       WATUPDICUSWORHSPHNITINR       6       2.23       1.001       1329       3338       13266       3'       1       1       10         32       acgt       n       1       SUMWOPSHNITINE       12       2.11       1       10       10       10       10       10       10       10       10       10       10	23	ct	У	f1	DVSGPSSPSSSLMTLTLFSPDLLGDPDNYTLANPLNTPPY	2	2.27	0	0.724	15714	15801	15684	15711	3′	30	10
25       ar. erg f.       10       MVGSFMGSGDR/TER/DSFCDPREACH/CFICSTLQSAR       38       2/43       1       0.459       1995       2049       2052       2109       5       19       19         27       ar. gr f.       5       2       SCUVFEWSRISSULCITSTRISULCIGATIVITER       8       3.85       1       0.0459       6648       6558       6588       3       2       1       0.0459       6644       6558       6588       3       2       1       0.01       7503       7476       7462       745       74	24	gt	q	r2	NNLFSLYCYLFQLWMMDPEHMNSMALK	1	2.55	0	0.464	13329	13287	13365	13332	3′	16	11
2b       acgf       n       f.2       WISSELARPARIVUCAISSMELTUVCCI HYPH-HYMWONK       1       2.7.1       1       0.218       442.8       4389       4500       44.31       3       1       1.2         28       acgf       y       10       EATASSACINASYDCQSCKDSQATPYTKIPTK       9       2.77       1       0.425       7221       7254       7161       7248       3'       1       1.0         30       acgf       x       r1       CREUPSCALWSSVLTYLR       129       5.31       0       0.425       7221       7246       7463       3'       3'       10         31       acgf       n       10       WINTDLIGNSWPRISTNNTTRR       6       2.39       1       0.016       7476       74/5       3'       1'       1         33       acgf       n       10       SLRQNNDESPRSTNWNTRR       1       2.12       17       150       150       15264       1264       1261       126       1'       1'       1'       1.000       1329       1338       1366       1'       1'       1'       1'       1'       1'       1'       1'       1'       1'       1'       1'       1'       1'       1'       1' </td <td>25</td> <td>ac gt</td> <td>S</td> <td>t0</td> <td>MVGSFMGSGDKPTEPGDSFGEPRSEAGPGPGSTLQSAR</td> <td>38</td> <td>2.43</td> <td>1</td> <td>0.459</td> <td>1995</td> <td>2049</td> <td>2052</td> <td>2109</td> <td>5'</td> <td>19</td> <td>19</td>	25	ac gt	S	t0	MVGSFMGSGDKPTEPGDSFGEPRSEAGPGPGSTLQSAR	38	2.43	1	0.459	1995	2049	2052	2109	5'	19	19
27       acgt       5       SUMVEWSHIPSVILLSLPVLACHTMETPK       9       3.85       1.000       b648       b528       <	26	ac gt	h	r2	WSSSLAAPSAFVLVGMSSRHSLLVCGTHVYFFGHNWNK	1	2.71	1	0.218	4428	4389	4500	4431	3'	14	24
28       acg f       y       10       EARLSSACUADSYDUCSGEDQUALTY IRVITYK       39       2.7       0.023       7.21       7.254       7.161       7.218       37       2       9       11         30       acg f       x       r1       RUEKTPXCLUSCAGEMAR       129       5.31       0       0.516       7.476       7.425       7.509       7.476       37       31       10         31       acg f       n       10       VMTDULQSNSPHYSTNYTTR       12       2.39       10       0.516       7.476       7.452       7.509       7.476       31       10         32       acg f       n       12       SNALINNAGEMAECHYSSSPNINK       1       2.275       1       1.000       8559       8583       8520       856       32       10       10         33       acg f       n       11       EMEZAMEZCHYERAMARYESCHYEVULCEPEDPSLITIK       1       2.010       10.338       13266       13266       13266       13266       13266       13266       13266       13266       13266       13266       13266       13266       1480       10       10       10       10       10       10       10       10       10       11       10	27	ac gt	S	f2	SGWVEWSRHSVLLLLSLPVLAAGITMLLTDR	8	3.85	1	1.000	6591	6648	6558	6588	3'	21	10
29       acg f       x       r1       CDAPTICALLWARAMANE       435       430       1       0.501       74/9       74/9       74/0	28	ac gt	У	f0	EATASSAGNDASYDGQSGKDSQATPYTKPTPK	425	2.77	1	0.425	7221	7254	/161	7218	3'	12	20
30         a eg b         x         11         MEDRALINGSVEATILA         129         53.1         0         0.316         7470         7470         7479         3         1         31         1           31         a egt         n         10         WINDLUGSWSVEATILA         12         123         1         1000         8559         8193         8127         8174         3'         13         10           32         a egt         n         12         SIALUNACKNAECHYSSENSINIT         12         12         1         1000         8559         8833         8520         8556         3'         9         13           34         a egt         n         16         MAID         1         1000         1329         1333         1326         12         21         1         1000         1329         1333         13'         11         19           35         agt         t         r         ICDMUEDMCKNSNNOL         13         20.8         0         0.875         7485         7515         7518         5'         10         13           4         agt         n         r         r         r         r         r         r	29	ac gt	X	Г1 1		435	4.00	1	0.901	7503	7479	7470	7440	2' 2'	9	11
1         acg         n         12         SNALDNARGRAGHYSSPNNC         12         2.5.7         1         0.5.7         0.15.7         <	30	ac gi	X n	fO	VMVTDU OKSWSDHSVNNVITNP	129	2.21	1	0.510	7470 8157	7452 8103	7309 8127	7479 8157	2/	13	10
32       acgi n       c       11       11       2.275       0       0.205	32	ac gi	n	f2	SNALNNACKNAFCHVSSSPNNK	1	2.55	1	1 000	8559	8583	8520	8556	3/	0	13
a ag t         no         fill         SikQNWDFSPNSSTMVVAGIFILIR         1         2.10         1         1.000         1329         1.333         1326         1329         3         1326         3         14         11          35         arg t         x         fi         EMHICSSEDSRAHNTWCALK         13         2.02         1         0.899         16728         16752         1698         16723         3'         11         19           36         ag ct         d         1         LIDDWLEDMCSNNQNULK         3         2.02         0         0.899         3426         339         3396         3375         5'         9         9           38         ag ct         n         fl         MULTSCNSSVSLAPLAATGSWGCLNQULR         7         2.96         0         0.437         7515         7518         7518         7518         7518         7518         7518         7518         7518         7518         7518         7518         7518         7518         7518         718         719         776         7719         5'         10         20           42         ag ct         v         r1         RMPEGAVCITVETISTINTTISTICTURGENKK         2.14         1         0.0421	33	ac gt	C	f1	TPC///VPFPAPAANV/HSSCPPCPW/IPCFPPSI PPSI TI TK	15	2.52	0	0.296	12564	12624	12510	12561	31	22	17
35       ac gt       x       ft       EMHLCSXEDSRAHNTWCXLK       13       2.02       1       0.989       16728       16752       16698       16725       3'       10       10         36       ag ct       k       r       ISLAPSGWSLINLTNPLFSMNLPTILLHKR       16       4.31       1       0.208       1728       1731       3'       11       19         37       ag ct       no       12       VALPISCTDSSVSLAPLAATCSWGCLNQTQLR       7       2.96       0       0.433       5076       4980       5040       3'       12       21         40       ag ct       q       r       VCVEDESAVTNTSTNLTLPTICSQPSNCKK       2       2.14       1       0.443       5043       575       575       5754       5'       10       12         41       ag ct       q       r       VCVEDESAVTNTSTNLTLPTICSQPSNCKK       2       2.14       1       0.4021       14874       16404       11644       11644       11647       5'       10       13         42       ag ct       v       r       RSPLPCDQVDYVVHCKGDSQFUVAFSVNK       34       3.08       1       1.000       1335       3129       11647       13'       1       94 <t< td=""><td>34</td><td>ac gt</td><td>no</td><td>fO</td><td>SIKONWDESENSSTMVVAGIELUR</td><td>1</td><td>2.13</td><td>1</td><td>1 000</td><td>13299</td><td>13338</td><td>13266</td><td>13296</td><td>31</td><td>14</td><td>11</td></t<>	34	ac gt	no	fO	SIKONWDESENSSTMVVAGIELUR	1	2.13	1	1 000	13299	13338	13266	13296	31	14	11
36       ag ct       k       r.2       SLAPSCWSLIMTNPESSMNLPTILLHKR       16       4.31       1       0.208       1728       1704       1788       17.31       3'       11       19         37       ag ct       d       n       12       WALPSCMSSUMPLAATGSWCCLNQTQLR       7       2.96       0       0.43       5043       5076       4980       5040       3'       12       21         39       ag ct       v       ft       MPPYTWSDVMSINCTVCLGCLR       7       2.96       0       0.435       5043       5076       4980       5040       3'       12       21         40       ag ct       v       r1       MVPVEDSAMTDISTINLTIPTICPSNCKK       2       2.14       1       0.875       7485       7515       7786       7716       7516       7517       756       7517       776       7716       7517       7518       7517       7751       7517       7756       7517       7518       754       7       10       12         42       ag ct       v       r1       AUXOVVVHCLGSCNOPGUAGSVNK       373       3       1       0.001       1280       12161       12183       12163       318       318       318	35	ac gt	x	f1	EMHLCSXEDSRAHNTWGXLK	13	2.02	1	0.989	16728	16752	16698	16725	3′	10	10
37       ag ct       d       r1       ICDDWLEDMCNNONQLK       3       2.08       0       0.899       3426       3399       3396       3375       5'       9       9         38       ag ct       no       12       WALPLSCDSSSVLAPLAATCSWGCLNQTQLR       7       2956       0       0.875       7485       7515       7518       754       5'       10       12         40       ag ct       r       1'       VVGVEDESAVTINTSITULPTICQPSNCKK       2       2.14       1       0.472       7809       7777       7776       7719       5'       10       12         41       ag ct       r       r1       CACWCPVPSQLGGGQAGVVKGLQCLHQCGPQNGCR       1       2.24       1       0.4024       14843       11646       11670       5'       14       9         43       ag ct       r       r1       RSPLPCQUVVVVLGCMSQUPLAYSVNK       34       3.08       1       0.000       11604       11643       11645       11670       5'       14       9         43       ag ct       r       r1       RSPLPCQUVDVVVLGCMCGMSQUPLAYSVNK       351       0       0.243       18474       14871       14371       11       9       16       14 <td>36</td> <td>ag ct</td> <td>k</td> <td>r2</td> <td>SLAPSGWSLLNLTNPLFSSMNLPTILLHKR</td> <td>16</td> <td>4.31</td> <td>1</td> <td>0.208</td> <td>1728</td> <td>1704</td> <td>1788</td> <td>1731</td> <td>3′</td> <td>11</td> <td>19</td>	36	ag ct	k	r2	SLAPSGWSLLNLTNPLFSSMNLPTILLHKR	16	4.31	1	0.208	1728	1704	1788	1731	3′	11	19
38       agct       90       72       940       0.143       5043       5074       980       30       3       12       21         39       agct       9       10       NPPTIWSDYMSDYMSDRGPVCLGCLR       1       0.437       5043       50715       7518       7515       5'       10       13         41       agct       9       72       CDACWCPVPSQLGCQCQCQACWKCLQCLHQQGCPQNGCR       1       2.24       1       0.804       9456       9387       9384       9342       5'       10       13         43       agct       no       f1       ANEAPLACQCQACWKCLQCHAQCCNWCK       34       3.08       1       1.000       12180       12126       1213       12183       3'       19       11         44       agct       r       f1       INPECOUPTKPTINENK       34       3.08       1       1.000       1315       3129       3126       306       5'       13       18         46       atcg       no       f2       SSERPTKCVVELASEEVK       4       2.41       1       1.000       3135       3129       3126       306       5'       9       10         47       atcg       atcg       r	37	ag ct	d	r1	LGDDWLEDMGNSNQNQLK	3	2.08	0	0.899	3426	3399	3396	3375	5′	9	9
99       ag ct       y       f0       NPPTWSDYMSIFCPVCLGCLR       15       2.08       0       0.875       7485       7515       7518       7554       5'       10       12         40       ag ct       e       r1       VVGVEDEAATINISTINITIPTIGOPSNCK       2       21.4       1       0.472       7809       7770       7776       7716       5'       10       13         41       ag ct       n       f1       AHVEAPIACSMVLAVISPCSNNR       377       374       0       1.000       11643       11643       11646       11670       5'       14       9         43       ag ct       v       r1       RNPERCDVDVVVHGGMSVQFLWAFSNK       914       3.51       0       0.423       14874       1490       14847       14871       3'       11       9         45       ag ct       v       r1       RSPIPCDVDVVVHGGMSVQFLWAFSONK       14       3.23       0       0.224       3858       3822       3912       3061       3'       13       18         46       at cg       e       r1       SLS       1       1.000       313       312       366       3'       13       21         47       at cg </td <td>38</td> <td>ag ct</td> <td>no</td> <td>f2</td> <td>WALFLSGTDSSSVSLAPLAATGSWGGLNQTQLR</td> <td>7</td> <td>2.96</td> <td>0</td> <td>0.143</td> <td>5043</td> <td>5076</td> <td>4980</td> <td>5040</td> <td>3′</td> <td>12</td> <td>21</td>	38	ag ct	no	f2	WALFLSGTDSSSVSLAPLAATGSWGGLNQTQLR	7	2.96	0	0.143	5043	5076	4980	5040	3′	12	21
40       ag ct       e       r1       VVCVEDESAVINTSTILTIPTIQPSNCKK       2       2.14       1       0.472       7809       7779       7779       7719       5'       10       20         41       ag ct       q       r2       GDACWGPVFSQLGGQQACWKGLQCLHQGGPQNGCR       37       1       0.804       9456       9384       9384       9342       5'       10       20         43       ag ct       r       RSPLFQQVDVVVHGGMSVGLWASSNNR       37       3.74       0       0.804       9456       9384       9342       5'       14       9         44       ag ct       r       RSPLFQQVDVVVHGGMSVGLWASSVNK       34       3.08       1       1.000       1218       14847       14847       14847       3'       11       9         45       ag ct       v       r1       PMPFEGVCPTIKUTVELASEEVX       1       3.23       10       0.423       1884       14847       14847       14847       13'       13	39	ag ct	У	fO	NPPYTWSDYMSIFCFVVCLGGLR	15	2.08	0	0.875	7485	7515	7518	7554	5′	10	13
41       ag ct       q       r2       CDACWCPVPSQLCGCQCQACVVKSLQCILHQQGGPQNGGR       1       2.42       1       0.804       9456       9387       9384       9342       5'       23       15         42       ag ct       no       f1       AHVEAPIACSMVLAVSPCSNNR       37       3.74       0       1.000       11643       11643       11670       5'       14       9         43       ag ct       v       r1       RSPLPCDQVDYVVHGCMSVQFWAPSOLCOVSVMLVR       914       3.51       0       0.423       14874       14807       14871       3'       11       9         45       ag ct       v       r2       FEPSDEVAAYGANSIRCVGVGVVMLVR       1       3.23       10       0.721       15015       14979       1508       3'       13       18         46       at cg       no       72       SERPTRCVFLASESSASTVVPEK       215       3.23       0       0.224       3858       3822       3921       3861       3'       13       12         48       at cg       e       r2       GISWPRDEGCGPFAGEAPACIK       59       3.82       1       0.169       5847       574       59       94       12         51       at cg	40	ag ct	e	r1	YVGVEDESAVTNTSTNLTLPTIGQPSNGKK	2	2.14	1	0.472	7809	7779	7776	7719	5′	10	20
42       ag ct       no       f1       AHVEAPIACSMVLAVTSPGSNNR       37       3.74       0       1.000       11643       11646       11670       5'       14       9         43       ag ct       v       r1       RSPLPGQVDYVVHGGMSVQFLWAFSVNK       34       3.08       1       1.000       12180       12126       12213       12138       3'       19       11         44       ag ct       r       r1       FNPFFGVQPTKVFLASEVK       91       3.51       0       0.423       14874       14904       14847       14817       3'       13       18         45       ag ct       v       r2       HYHPEPSDEVAAVGANSIRCVGCVVVMLVR       1       3.23       0       0.243       14874       14904       14847       14871       3'       13       18         46       at cg       no       f2       SSLRPYTKCVFLASEEVK       4       2.41       1       10.00       3135       3126       306       5'       9       10       18         48       at cg       a       r0       FEDNKMDSEDFYQTYFLGLAGNAGDCNGYGDMSYK       1       2.82       1       10.00       4074       4002       4107       3'       12       1	41	ag ct	q	r2	GDACWGPVPSQLGGQGQAGVVKGLQGLHQQGGPQNGGR	1	2.24	1	0.804	9456	9387	9384	9342	5′	23	15
43       ag ct       v       r1       REPLPCDQUDYVWHGGMSVORLWAFSVNK       34       3.08       1       1.000       12180       12120       12213       12183       3'       19       11         44       ag ct       v       r1       FNPFFGFVGMTKPTLNFNK       914       3.51       0       0.423       14874       14904       14871       14871       3'       11       9         45       ag ct       v       r2       HVHPEPSDEVAAYGANSIRCVCVGVVMLVR       1       3.23       1       0.721       15015       14979       15069       15018       3'       13       18         46       at cg       no       f2       SSLRPYTKCVVFLASEEVK       4       2.41       1       1.000       3135       3129       3126       3096       5'       9       10         47       at cg       e       r2       GISWPKLDEEGGOPFEAGEASYVVPEK       215       3.23       0       0.224       3858       3822       3010       4002       4107       407       3'       14       12         48       at cg       e       r1       SIG       GISWPRIDEGCOPFEAGEAPAGLK       59       3.82       1       0.0407       4002       4107	42	ag ct	no	f1	AHVEAPIAGSMVLAVTSPGSNNR	37	3.74	0	1.000	11604	11643	11646	11670	5′	14	9
44       ag ct       t       t       t       t       fill       FIPH-FCPVQPTIKPTINFNK       914       3.51       0       0.423       148/4       148/4       148/4       148/4       148/1       3'       11       9         45       ag ct       v       r.2       HVHEPEDEVAAYCANSIRC/GVGVVVMLVR       1       3.23       1       0.721       15015       14979       15069       15018       3'       13       18         46       at cg       no       f2       SSLRPYTKC/VFLASEEVK       4       2.41       1       1.000       3135       3129       3126       3096       5'       9       10         47       at cg       e       r       Q       QAVFTSLQSSSQNHCFMQHISSCESASYVPEK       215       3.23       0       0.224       3851       50       3126       3096       5'       9       16         50       at cg       e       r1       SIAGVDVAMAVSCTKTLYLLHSNTHNR       1       3.31       1       0.886       6201       6231       6147       6198       3'       10       18         51       at cg       c       r       ACELLEVWKASEPNSAVAK       8       3.17       1       0.171       9072	43	ag ct	V	r1	RSPLPGDQVDYVVVHGGMSVQFLWAFSVNK	34	3.08	1	1.000	12180	12126	12213	12183	3'	19	11
45       ag ct       v       r.2       HHPE/SDEVAAYGANSIKUCUCUVVMLVK       1       3.23       1       0.721       15015       149/9       15069       15018       3'       13       18         46       at cg       no       f2       SSLRPYTK/UVELASEEVK       4       2.41       1       1.000       3135       3129       3126       3096       5'       9       10         48       at cg       a       r0       FEDNKVDSFIDFYQTYFLGLAGNAGDCNGYGDMSYK       1       2.82       1       1.000       4074       4002       4107       4077       3'       24       12         49       at cg       a       r0       FEDNKVDSFIDFYQTYFLGLAGNAGDCNGYGDMSYK       1       2.82       1       1.000       4074       4002       4107       4077       3'       24       12         50       at cg       k       f1       SLACDVAMANSCTKULLEGCGCFFEACEAPAGLK       59       3.82       1       0.160       5847       5874       579       5844       3'       10       18         51       at cg       r       f2       WLWLGXCSCCWCRITSTPTYFHYSR       1       3.41       0.491       8070       8027       8020       8196       5'	44	ag ct	f	f1	FNPFFGFVGPITKPTLNFNK	914	3.51	0	0.423	14874	14904	14847	14871	3'	11	9
46       at cg       no       fz       SSLRFY IRCVPLASEEVK       4       2.41       1       1.000       3135       3129       3126       3096       5       9       10         47       at cg       e       r2       QAEVFLSUQSSQNHCFMQHISSCESASYVVPEK       215       3.23       0       0.224       3858       3822       3921       3861       3'       13       21         48       at cg       e       r2       GISWPKIDEEGGPFAGEAPACLK       59       3.82       1       0.169       5847       5874       5799       5844       3'       9       16         50       at cg       r       fz       MAGVDAMAVSCTKTLYLLHSNTHNR       1       3.31       1       0.886       6201       6231       6147       6198       3'       10       18         51       at cg       r       fz       MAGVCARECSCORCHRISTPTYFPHYSR       1       2.49       1       0.901       8007       8007       8022       8067<'/th>       3'12'       10''       10''         52       at cg       c       r1       AGC       0.171       9072       9042       9039       9012       5'       10       10         54       at	45	ag ct	v	r2	HVHPEPSDEVAAYGANSIRCVGVGVVVMLVR	1	3.23	1	0.721	15015	14979	15069	15018	3'	13	18
47       at cg       e       12       QAEVFLALQSSQNMERTMONTISUEARSTVYFEK       213       3.23       0       0.224       3336       3622       3921       3601       3       13       21         48       at cg       a       r0       FEDRKWDSFIDFVQTFLGLAGNAGDCNGYGDMSYK       1       2.82       1       0.169       5847       5874       5799       5844       3'       24       12         50       at cg       k       f1       SIAGVDVAMAVSGTKTLYLLHSNTHHNR       1       3.31       1       0.886       6201       6231       6147       6198       3'       10       18         51       at cg       c       r1       ZIGVDVAMAVSGTKTLYLLHSNTHHNR       1       3.31       1       0.886       6201       6231       6147       6198       3'       10       18         52       at cg       c       r1       ACCGULVWKASEPENSAVAK       8       3.17       1       0.171       9072       9042       9039       9012       5'       10       10       10       10       11       3.57       0       0.249       9648       9669       9672       9705       5'       9       11       155       at cg       n       <	40	at cg	110	12	SSERPTIKUV VPLASEEVK	215	2.41	1	0.224	3133	3129	3120	3090	2' 2'	12	10
46       at cg       a       10       ILDIARWDSHDFRQTTREGENGREDMERK       1       2.62       1       1.000       4074       4002       4107       3       24       12         49       at cg       e       f2       GISWPKIDEEGGPFEACEAPAGLK       59       3.82       1       0.169       5847       5874       5799       5844       3'       9       16         50       at cg       k       f1       SIAGVDVAMAVSGTKTLYLLHSNTHNR       1       3.31       1       0.886       6201       6231       6147       6198       3'       10       18         51       at cg       c       r2       WLPWLGCSCGWCRLITSTPTYFPHYSR       1       2.49       1       0.941       8070       8097       8022       8067       3'       11       16         52       at cg       c       r2       CYUGAFHCNLHNQENCK       27       3.92       0       0.690       8247       8223       8220       8196       5'       9       9         53       at cg       e       f1       3.57       0       0.249       9648       9669       9672       9705       5'       10       10         54       at cg <td< td=""><td>47</td><td>at cg</td><td>e</td><td>12 r0</td><td></td><td>215</td><td>5.25 2.92</td><td>1</td><td>1.000</td><td>2020 4074</td><td>3622</td><td>3921 4107</td><td>2001 4077</td><td>2/</td><td>15</td><td>21 12</td></td<>	47	at cg	e	12 r0		215	5.25 2.92	1	1.000	2020 4074	3622	3921 4107	2001 4077	2/	15	21 12
10       allog       10       0.1011       10011       0.1011       0	40 49	at cg	d P	f2	CISWPKI DEFCCCPEFACEAPACI K	59	2.82	1	0.169	4074 5847	4002 5874	5799	5844	ר זי	24 0	16
551       at cg       r       12       MUPWLGCSCGWCRLITSTPTYFHYSR       1       0.941       8071       8071       8077       3'       11       16         52       at cg       r       12       WUPWLGCSCGWCRLITSTPTYFHYSR       1       0.941       8070       8097       8022       8067       3'       11       16         52       at cg       c       r2       CYLVGAFHCNLHNQENCK       27       3.92       0       0.690       8247       8223       8200       8167       3'       11       16         54       at cg       e       f1       EIFLSLLPQVCSGMCGESSR       8       3.17       1       0.171       9072       9042       9039       9012       5'       10       10         55       at cg       h       f2       GFLCIKLSCVGCCPHLLASLYYFLK\$       8       2.08       1       0.935       11037       11070       10992       11034       3'       13       17         56       at cg       r       f0       RPRLTSLPSLNDIMINIEWSGSAGSVNMGSVCEPVCR       1       2.94       1       0.124       15687       15735       15738       15738       15798       5'       18       20         58 <td< td=""><td>50</td><td>at cg</td><td>k</td><td>f1</td><td>SIAGVDVAMAVSGTKTI YI I HSNTHHNR</td><td>1</td><td>3 31</td><td>1</td><td>0.105</td><td>6201</td><td>6231</td><td>6147</td><td>6198</td><td>31</td><td>10</td><td>18</td></td<>	50	at cg	k	f1	SIAGVDVAMAVSGTKTI YI I HSNTHHNR	1	3 31	1	0.105	6201	6231	6147	6198	31	10	18
52       at cg       c       r.2       CYLVGAFHCNLHNQENCK       27       3.92       0       0.690       8247       8223       8036       5'       10       10         53       at cg       e       r.1       AGEGLLEVWKASEPNSAVAK       8       3.17       1       0.171       9072       9042       9039       9012       5'       10       10         54       at cg       e       f.1       EIFLSLLPQVCSGMGGESSR       11       3.57       0       0.249       9648       9669       9672       9705       5'       9       11         55       at cg       r       f.2       GFLCIKLSCVGGCPHILASSLYYFLTK\$       8       2.08       1       0.935       11037       11070       10992       11034       3'       13       17         56       at cg       r       f.0       RPRLTSLPSLINDINTILWSGCSAGSVMGSVCEPVCR       1       2.94       1       0.124       15687       15735       15738       15798       5'       18       200         58       acg       n       r2       nd       1.000       3315       3255       3345       3318       3'       21       10       10       10       10       10	51	at cg	r	f2	WIPWIGCSCGWCRUITSTPTYFPHYSR	1	2.49	1	0.941	8070	8097	8022	8067	31	11	16
53       at cg       e       f1       AGEGLLEVWKASEPNSAVAK       8       3.17       1       0.171       9072       9042       9039       9012       5'       10       10         54       at cg       e       f1       EIFLSLLPQVGSGMGGESSR       11       3.57       0       0.249       9648       9669       9672       9705       5'       9       11         55       at cg       h       f2       GFLCIKLSCVGGCPHILASSLYYFLTK\$       8       2.08       1       0.935       11037       11070       10992       11034       3'       13       17         56       at cg       x       r0       DGGNXGSQGXGAMSSHVPMMKMNLNVVLK\$       38       2.25       1       0.766       12321       12291       12288       12237       5'       12       17         57       at cg       r       f0       RPRLTSLPSLLNDINTIWSGCSAGSVNMGSVCEPVGR       1       2.94       1       0.124       15687       15735       15738       15798       5'       18       20         58       acg       n       r2       NTTLSRTLNVGAVMNNVMVDVAGFNGSLVK       60       2.68       1       1.000       3315       3255       3345       3318       3'<	52	at cg	c	r2	CYLVGAFHCNLHNOENCK	27	3.92	0	0.690	8247	8223	8220	8196	5′	9	9
54       at cg       e       f1       EIFLSLLPQVGSGMGGESSR       11       3.57       0       0.249       9648       9669       9672       9705       5'       9       11         55       at cg       h       f2       GFLCIKLSCVGGCPHILASSLYYFLTK\$       8       2.08       1       0.935       11037       11070       10992       11034       3'       13       17         56       at cg       x       r0       DGGNXGSQGXGAMSSHVPMMKMNLNVVLK\$       38       2.25       1       0.766       12321       12291       12288       12237       5'       12       17         57       at cg       r       f0       RPRLTSLPSLLNDINTILWSGCSAGSVNMGSVCEPVGR       1       2.94       1       0.124       15687       15735       15738       15738       5'       18       20         58       acg       n       r2       NTTLSRTLNVGAVMNNVMVDVAGFNGSLVK       60       2.68       1       1.000       3315       3255       3345       3318       3'       21       10         59       acg       e       f0       SEHTPQLPTETSSALSDRR       159       3.52       1       0.411       5895       5109       5112       5142       5 </td <td>53</td> <td>at cg</td> <td>e</td> <td>r1</td> <td>AGEGLLEVWKASEPNSAVAK</td> <td>8</td> <td>3.17</td> <td>1</td> <td>0.171</td> <td>9072</td> <td>9042</td> <td>9039</td> <td>9012</td> <td>5′</td> <td>10</td> <td>10</td>	53	at cg	e	r1	AGEGLLEVWKASEPNSAVAK	8	3.17	1	0.171	9072	9042	9039	9012	5′	10	10
55       at cg       h       f2       GFLCIKLSCVGCCPHILASSLYYFLTK\$       8       2.08       1       0.935       11037       11070       10992       11034       3'       13       17         56       at cg       x       r0       DGGNXGSQGXGAMSSHVPMMKMNLNVVLK\$       38       2.25       1       0.766       12321       12291       12288       12237       5'       12       17         57       at cg       r       f0       RPRLTSLPSLLNDINTILWSGGSAGSVNMGSVCEFVGR       1       2.94       1       0.124       15687       15735       15738       15798       5'       18       20         58       acg       n       r2       NTTLSPLNUQAVMNNVMVDVAGFNGSLVK       60       2.68       1       1.000       3315       3255       3345       3318       3'       21       10         59       acg       e       f0       SEHTPQLPTETTSSALSDRR       159       3.52       1       0.141       5085       5109       5112       5142       5'       10       10         60       acg       g       r2       QLVEQRKDTVEWQDMVEVGYNVVR       3       5.28       1       0.401       6891       6450       6531       6484       3	54	at cg	e	f1	EIFLSLLPQVGSGMGGESSR	11	3.57	0	0.249	9648	9669	9672	9705	5′	9	11
56       at cg       x       r0       DGGNXGSQGXGAMSSHVPMMKMNLNVVLK\$       38       2.25       1       0.766       12321       12291       12288       12237       5'       12       17         57       at cg       r       f0       RPRLTSLPSLLNDINTILWSGGSAGSVNMGSVGEFVGR       1       2.94       1       0.124       15687       15735       15738       15798       5'       18       20         58       acg       n       r2       NTTTLSRTLNVGAVMNVVVQAGFNGSLVK       60       2.68       1       1.000       3315       3255       3345       3318       3'       21       10         59       acg       e       f0       SEHTPQLPTETTSSALSDRR       159       3.52       1       0.141       5085       5109       5112       5142       5'       10       10         60       acg       g       r0       WGGSTINGGEPTGGSTLVGGEYKLQGDR       173       2.52       1       0.401       6891       6450       6531       6484       3'       12       16         61       acg       q       r2       DVPCKDTVEWQDMVEVGYNVR       3       528       1       0.401       6891       6852       6924       6884       3'	55	at cg	h	f2	GFLCIKLSCVGGCPHLLASSLYYFLTK\$	8	2.08	1	0.935	11037	11070	10992	11034	3′	13	17
57       at cg       r       f0       RPRLTSLPSLLNDINTILWSGCSAGSVNMGSVCEPVGR       1       2.94       1       0.124       15687       15735       15738       15738       5'       18       20         58       acg       n       r2       NTTILSRTLNVGAVMNVMVDVAGFNGSLVK       60       2.68       1       1.000       3315       3255       3345       3318       3'       21       10         59       acg       e       f0       SEHTPQLPTETTSSALSDRR       159       3.52       1       0.141       5085       5109       5112       5142       5'       10       10         60       acg       g       r0       WGGSTINGGEPTGGSTLVGGEYKLQGDR       173       2.52       1       0.411       5085       6450       6531       6484       3'       12       16         61       acg       q       r2       QLVEQKDTVEWQDWVEVGYNVVR       3       5.28       1       0.401       6891       6852       6924       6894       3'       12       14         62       acg       t       f2       TSKPHPTTTPPSSSTPLQCLQCGCVRCVQAHQGAGHVQGR       73       3.03       1       0.866       13167       13230       1323       13290 <t< td=""><td>56</td><td>at cg</td><td>х</td><td>r0</td><td>DGGNXGSQGXGAMSSHVPMMKMNLNVVLK\$</td><td>38</td><td>2.25</td><td>1</td><td>0.766</td><td>12321</td><td>12291</td><td>12288</td><td>12237</td><td>5′</td><td>12</td><td>17</td></t<>	56	at cg	х	r0	DGGNXGSQGXGAMSSHVPMMKMNLNVVLK\$	38	2.25	1	0.766	12321	12291	12288	12237	5′	12	17
58       acg       n       r2       NTTLSRTLNVGAVMNVMVDVAGFNGSLVK       60       2.68       1       1.000       3315       3255       3345       3318       3'       21       10         59       acg       e       f0       SEHTPQUPTETTSSALSDRR       159       3.52       1       0.141       5085       5109       5112       5142       5'       10       10         60       acg       g       r0       WGGSTINGGEPTGGSTLVGGEYKLQGDR       173       2.52       1       0.141       5085       6109       6131       6486       3'       12       16         61       acg       q       r2       OLVEQPKDTVEWQDMVEVGYNVR       3       5.28       1       0.401       6891       6852       6924       6884       3'       12       16         62       acg       t       f2       TSKPHPTTTPPSSSTPLGCLQCCGVRGVQAHQGAGHVQGR       73       3.03       1       0.876       13167       13215       13218       13237       5'       17       24         63       acg       t       f2       TSKPHPTTTPPSSSTQISPTCGGVRQVQAHQGAGHVQGR       73       3.03       1       0.866       13167       13230       1323       13290 <td< td=""><td>57</td><td>at cg</td><td>r</td><td>fO</td><td>RPRLTSLPSLLNDINTILWSGGSAGSVNMGSVGEFVGR</td><td>1</td><td>2.94</td><td>1</td><td>0.124</td><td>15687</td><td>15735</td><td>15738</td><td>15798</td><td>5′</td><td>18</td><td>20</td></td<>	57	at cg	r	fO	RPRLTSLPSLLNDINTILWSGGSAGSVNMGSVGEFVGR	1	2.94	1	0.124	15687	15735	15738	15798	5′	18	20
59       acg       e       f0       SEHTPQLPTETTSSALSDRR       159       3.52       1       0.141       5085       5109       5112       5142       5'       10       10         60       acg       g       r0       WGGSTINGGEPTGGSTLVGGEYKLQGDR       173       2.52       1       0.351       6483       6450       6531       6486       3'       12       16         61       acg       q       r2       QLVEQFKDTVEWQDMVEVGYNVVR       3       5.28       1       0.401       6891       6852       6924       6894       3'       13       11         62       acg       t       f2       TSKPHPTTTPPSSSTPLQGLQCGCVRGVQAHQGAGHVQGR       73       3.03       1       0.876       13167       13215       1328       13287       5'       17       24         63       acg       t       f2       TSKPHPTTTPPSSSTQISPTCGGVRGVQAHQGAGHVQGR       73       3.03       1       0.866       13167       13230       1323       13290       5'       22       19         64       acg       no       f2       GKQEPGLEQLCASSAALGEIPLNNNPTLPK       17       2.36       1       0.236       13989       14022       13986       3'	58	acg	n	r2	NTTTLSRTLNVGAVMNNVMVDVAGFNGSLVK	60	2.68	1	1.000	3315	3255	3345	3318	3′	21	10
60       acg       g       r0       WGGSTTNGGEPTGGSTLVGGEYKLQGDR       173       2.52       1       0.351       6483       6450       6531       6486       3'       12       16         61       acg       q       r2       QUVEQPKDTVEWQDMVEVGYNVVR       3       5.28       1       0.401       6891       6852       6924       6894       3'       13       11         62       acg       t       f2       TSKPHPTTTPPSSSTPLQGLQCGCVRGVQAHQGAGHVQGR       73       3.03       1       0.876       13167       13215       1328       13287       5'       17       24         63       acg       t       f2       TSKPHPTTTPPSSSTQISPTCGGVRGVQAHQGAGHVQGR       73       3.03       1       0.866       13167       13230       1323       13290       5'       22       19         64       acg       no       f2       GKQEPGLEQLCASSAALGEIPLNNNPPLPK       17       2.36       1       0.236       13989       14022       13932       13986       3'       12       19         65       acg       e       f0       GLPPHQVHHKPHKPHYETHTQQK       2       2.05       0       1.000       14855       14859       3'       14	59	acg	e	fO	SEHTPQLPTETTSSALSDRR	159	3.52	1	0.141	5085	5109	5112	5142	5′	10	10
61       acg       q       r2       QLVEQPKDTVEWQDMVEVGYNVVR       3       5.28       1       0.401       6891       6852       6924       6894       3'       13       11         62       acg       t       f2       TSKPHPTTTPPSSSTPLQCLQCCGVRGVQAHQGAGHVQGR       73       3.03       1       0.876       13167       13215       13218       13287       5'       17       24         63       acg       t       f2       TSKPHPTTTPPSSSTQISPTCGCVRGVQAHQGAGHVQGR       29       2.99       1       0.866       13167       13230       13233       13290       5'       22       19         64       acg       no       f2       GKQEPGLEQLCASSAALGEIPLNNNPPLPK       17       2.36       1       0.236       13989       14022       13980       3'       12       19         65       acg       e       f0       GLPQHQVHHKPHKPHYETHTQQK       2       2.05       0       1.000       14852       14859       3'       14       9	60	acg	g	r0	WGGSTTNGGEPTGGSTLVGGEYKLQGDR	173	2.52	1	0.351	6483	6450	6531	6486	3′	12	16
62       acg       t       f2       TSKPHPTTTPPPSSSTPLQCLQCGCVRGVQAHQGAGHVQGR       73       3.03       1       0.876       13167       13215       13218       13287       5'       17       24         63       acg       t       f2       TSKPHPTTTPPSSSTQISPTCGGVRGVQAHQGAGHVQGR       29       2.99       1       0.866       13167       13215       13218       13287       5'       17       24         64       acg       no       f2       GKQEPGLEQLCASSAALGEIPLNNNPPLPK       29       2.99       1       0.866       13167       13230       13233       13290       5'       22       19         64       acg       no       f2       GKQEPGLEQLCASSAALGEIPLNNNPPLPK       17       2.36       1       0.236       13989       14022       13986       3'       12       19         65       acg       e       f0       GLPQHQVHHKPHKPHYETHTQQK       2       2.05       0       1.000       14862       14901       14855       3'/       14       9	61	acg	q	r2	QLVEQPKDTVEWQDMVEVGYNVVR	3	5.28	1	0.401	6891	6852	6924	6894	3′	13	11
b3       acg       t       t2       15KPHPTTTPPSSS1QISPTICGGVRGVQAHQGAGHVQGR       29       2.99       1       0.866       13167       13230       13230       5'       22       19         64       acg       no       f2       GKQEPGLEQLCASSAALGEIPLPNNNPPLPK       17       2.36       1       0.236       13989       14022       13932       13986       3'       12       19         65       acg       e       f0       GLPQHQVHHKPHKPHYETHTQQK       2       2.05       0       1.000       14862       14901       14855       3'       14       9	62	acg	t	f2	TSKPHPTTTPPPSSSTPLQGLQCGGVRGVQAHQGAGHVQGR	73	3.03	1	0.876	13167	13215	13218	13287	5′	17	24
b4         acg         no         I2         GKQEPGLEQULASSAALGEIPLENNNPPLPK         17         2.36         1         0.236         13989         14022         13932         13986         3'         12         19           65         acg         e         f0         GLPQHQVHHKPHKPHYETHTQQK         2         2.05         0         1.000         14862         14901         14835         14859         3'         14         9	63	acg	t	t2	ISKPHP1TTPPPSSSTQISPTTCGGVRGVQAHQGAGHVQGR	29	2.99	1	0.866	13167	13230	13233	13290	5'	22	19
י ארא <u>הויעתעיחת</u> גיאגיאזידוחועעג 2 2.05 ט 1.000 14862 14901 14835 14859 3' 14 9	64 65	acg	no	t2 f0	GRUEPGLEQLCASSAALGEIPLPNNNPPLPK	17	2.36	1	0.236	13989	14022	13932	13986	3′ 2'	12	19
	00	acg	е	10		Z	2.05	U	1.000	14602	14901	14633	14609	2	14	9

(continued on next page)

# Table 1 (continued)

1	2	3	4	5	6	7	8	9	10	11	12	13	14		
66	acg	e	f0	GLPQHQVHHQPHKPHYETHTQQK	2	2.05	0	1.000	14865	14901	14835	14862	3′	13	10
67	acg	а	f1	GGFSGPSQILIATILCSFMGK	4	2.24	0	0.319	16266	16305	16242	16263	3′	12	9
68 60	act	e t	r2 r1	GENAGELEGWIAILSCSSPEGPIILGPLK	6	2.59	0	0.156	3228 5508	3186 5511	3273 5508	3231 5472	3' 5/	14 30	15 12
70	act	k	f2	NLPILLLTNVEPOPFPPTPK	27	4.28	0	0.948	11097	11118	11064	11094	3'	9	12
71	act	k	f2	NLPILLLTNVDPQPFPPTPK	197	4.97	0	1.000	11097	11118	11064	11094	3′	9	11
72	act	n	f2	WPLQCRQMSTTSSNFCHFNPNNNPPLPK	3	2.93	1	0.970	13998	14022	13941	13995	3′	9	19
73	agc	r L	f2	DQPNPLRPCPAPLTDAMAIR	45	2.19	0	0.272	3945	3969	3972	4002	5′ 2′	10	10
74 75	age	K V	r2	FOLOSMPVLGSVGGVGGAMGVMR	125	5.80 4.64	0	0.582	8157	8115	8184	8160	3/	15 14	9
76	agc	no	r2	TSYLLLQQPLLWWLCWFKLCVFWK	5	2.01	1	0.163	10749	10716	10785	10752	3′	12	12
77	agc	q	r1	MQQAAFALQWMLLWGQDWQQQWMR	8	2.16	0	0.264	12432	12390	12459	12435	3′	15	9
78	agc	w	r2	MSSTCGNDISMSRISGLFSAWGWK	4	2.04	1	0.739	13098	13068	13137	13101	3'	11	13
/9	agc	q	f0 f0		13	3.//	0	0.353	13314	13338	13281	13311	3' 5/	12	11
81	age	d	r1	FLVDLGLGGVGAGFGSDEGLDDGLCGVWCDFMTVSLLMNK	200	2.80	0	0.029	13458	134887	13451	13854	5′	30	10
82	agt	a	fO	TDAQSGGASAYKAHENILLR	1	2.25	1	1.000	2343	2370	2313	2340	3′	10	10
83	agt	n	f2	LIYSTSITLLPMTGGNGEGMR	194	4.00	0	0.586	5442	5475	5478	5502	5′	12	9
84	agt	d	f2	LIYSTSITLLPMTGGDGEGMR	61	3.96	0	0.460	5442	5475	5478	5502	5′	11	10
85 86	agt	Γ σ	r2 r2	AASCCPPSCI PEVCINGCONCMISTAACCPCIVCAMNEANC*	1	2.41	1	0.180	6423 8493	6369 8409	6366 8526	6312 8496	5' 3'	20 30	18 11
87	agt	a	r2	AAEGNSYAEEFYGEADAGGGYAVEAATAWGGPPLAEAVPR	1	2.43	0	0.518	10143	10068	10065	10026	5′	25	15
88	agt	d	r0	NYSSAMGACQGGSDESNDDGSGVCVWFARGPVVAAPGAVDR	6	2.55	1	0.953	13566	13491	13488	13446	5′	26	15
89	agt	У	f2	RIRPVEVVGAIPHYFLLQYPHHR*	1	2.81	1	0.173	13713	13749	13680	13710	3′	12	11
90 01	atc	h	f1 £1		12	2.55	1	0.124	1413	1452	1455	1515	5'	15	20
91	atc	q v	11 r2	GILIVQQQHIMIPPIPLGNAR GIVIVVVKALGSVGGVGGAMGVMR	429	3.19 4.41	1	0.059	2010 8157	2043 8115	2040 8187	2073	יב גי	10 14	10
93	atc	v	r2	GLVIVVVKAUGSVGGVGGAMGVMR	1057	4.48	1	0.140	8160	8115	8187	8163	3'	14	10
94	atc	х	r1	MXLMLMIALVAXEXIQMVVLFSVMAGMLGVVGWCR\$	1	2.16	0	0.818	13263	13233	13335	13266	3′	11	24
95	atc	k	r0	FNYAFLGWGDWLLLWNYHMGMKVK	17	3.81	1	1.000	14049	14022	14019	13980	5′	10	14
96	atc	q	r2	LWLCSKGGQWLQLGFVMNQFLMNDPK	35	2.60	1	0.630	15408	15381	15378	15333	5′	11	15
97 98	atg	n d	10 r1	FPMMHOVSMCKKPVSCCDPPDEDDVDCIK	1 14	2.13	1	0.076	1344 5061	1389 5007	1392 5088	1419 5064	י בי	17	10
99	atg	no	f0	TMASSSPPSVPPAPAGSASASVTVASPPLALVAPVR\$	13	2.42	0	1.000	5376	5409	5412	5481	5′	9	27
100	atg	no	f2	GASFLFIWNSLYLLFGAWAGVLGTALSLLIRAELGQPGNL	1	2.24	1	1.000	6054	6141	6024	6051	3′	30	10
101	atg	m	r0	IMRMGAFGIGNMSGENTSTK	65	2.86	1	1.000	6291	6264	6261	6234	5′	10	10
102	atg	p	fO	APPTALVGTDSPHSTEAMWNDLLQCSEPPDSSFFSPPVA	10	2.64	0	0.564	6987	7074	6960	6984	3'	30	9
103	atg	K n	rı r1	KPY ILPMESMNPFCSHSKAK	51	2.17	1	0.924	10284	10263	10314	10287	3' 3'	10 30	10 11
101	atg	x	r1	EWAEVSSCGEEGXGGADAASEEPTKTTGER	11	2.51	1	0.991	14826	14784	14781	14739	5′	10	19
106	cgt	q	r1	GLYWWDQQYGSGQGGWSLASPLDLGAQWTQGVGFR	1	2.15	0	1.000	228	189	186	126	5′	14	21
107	cgt	х	fO	HPKPKPWEMXLXIPLXIXLLVQXGTALWTLGK	1	2.72	0	0.632	2103	2166	2073	2100	3′	22	10
108	cgt	m	r0	LLQSDHTAFGSAPMSPQPK	2	2.13	0	0.217	2604	2577	2631	2607	3'	10	9
110	cgt	d t	12 r2	AGASEGINTS YMMICLITHYNLQHSPSNLANMISDK RPPSCWPSCTSTI VSCATTTTCRCACVTVFTTFCCSSTDNM	6	2.06	1	0.034	4278 4989	4347 4902	4251 5028	4275	3' 3'	23 30	10
111	cgt	no	rO	LFLGMCLKQENPVMMSGLK	393	4.09	1	0.162	12066	12033	12093	12069	3′	10	9
112	cgt	t	fO	VFLLTMTFNQNITLWIWQHIASTGHPGMNATMSCK	2	2.14	0	0.703	12330	12384	12387	12435	5′	19	17
113	cgt	f	f2	MNGTEGHVVHVPDGVASMIYPTLQLMFPTTNSPSK	8	2.57	0	0.942	13107	13161	13059	13104	3'	19	16
114	ctg	v	r0 f1		1	2.98	0	1.000	3126	3078	3192	3129	3'	17	22
116	ctg	k	r2	DOVRPLVLCMVMLYFTIHLLHAYK	20	2.55	0	0.480	8601	8571	8568	8532	5 5'	12	10
117	ctg	k	fO	MNNKVFLVFVQTTIPLYLK	791	3.61	1	0.130	14001	14025	13971	13998	3′	9	10
118	ctg	no	f2	SPSSMYPNNKLEEGLYELK	70	4.56	1	0.048	15915	15945	15948	15972	5′	10	9
119	acgt	t	f1	GGAGGTPAATGTRTPSAGSDSVSTDFTQPTTSTTTPTNLK*	5	2.24	1	0.825	2520	2466	2586	2523	3'	18	22
120	acgt	d d	r1	AVEDMTGWADGUSTGDVDPTFSGVPKUSGGSAK	6	2.88	1	0.342	2957 4305	4236	4233	4206	5'	23	9 11
122	acgt	f	r2	EEMLDGSFCGTFVFGGPVLALFSVMR	1	2.20	0	0.163	4452	4416	4413	4371	5′	14	12
123	acgt	no	r2	VPRQALVPFEVNEASYDGK	141	3.13	1	0.592	6930	6906	6960	6933	3′	9	10
124	acgt	no	r2	VPRQALVPFDVNEASYDGK	388	3.16	1	0.496	6930	6906	6960	6933	3′	9	10
125	acgt	no	fl r1		29	2.97	1	1.000	/28/	/350	/353	/404	5' 2'	21	18
120	acgt	ч no	r2	VVGGVGWVPLAWLSLDMLOR	167	3.64	0	0.555	14448	14424	14481	14451	3'	9	10
128	acgt	a	r2	WMSGALILLGGAFCVLGSFRGGVGFVLLPGDVMADAGVER*	4	2.37	1	1.000	14772	14718	14715	14652	5′	18	22
129	actg	d	r1	SSYKDPFAEDADLDNDIALLSLGDLVPLVK	149	3.81	1	0.469	2733	2682	2769	2736	3′	17	13
130	actg	q	r2	GMGQGVHSQQAMVQAKVGAVMQQVMVDVAGGQNVGQPQGR	2	2.36	1	0.409	3363	3276	3273	3246	5′	30	10
131	actg	s	f0	NSVCSDGSAKAVSPLAPGLSHK	25	2.75	1	0.560	3273	3303	3300	3330	5'	10	12
133	actg	s	rO	MIMSAWSWKVMSSSMMETSMVEHLLDMIEIRPR	6	2.41	1	0.948	7527	7482	7479	7431	5′	17	16
134	actg	no	f2	SLSPFMITPSSVGVGVMVAVER	59	3.28	0	0.884	7767	7794	7797	7830	5′	11	11
135	actg	t	r0	TPKEQEIGEATVGTGIFNLTAK	2	2.42	1	1.000	8778	8748	8811	8781	3′	11	11
136	actg	no £	f2	NQMIQALLITILLGLYFTLLSIVTAGTVFGLR	1	2.29	0	0.690	9834	9894	9897	9930	5′ 2′	20	12
137	actg	ı r	r1 r2	WGSSNHEHGGAGCLMGMVOGR	1 571	2.42 4.35	1	0.403	11400	11307	11451 11517	11403	3'	12 9	17
139	actg	k	r2	WGSSNHEHGGAGGLMGMVQGKGK	67	4.67	1	0.973	11478	11454	11517	11481	3′	10	13
140	actg	d	r0	AMLLDMGAWVSKVETWVDAR	9	4.42	1	1.000	12186	12153	12150	12126	5′	11	9
141	actg	d	r0	AMLLDMGAWVSQVETWVDAR	9	4.42	0	1.000	12186	12153	12150	12123	5′	10	10

 Table 1 (continued)

1	2	3	4	5	6	7	8	9	10	11	12	13	14		
142	actg	no	f2	VSAASSGFYRPLPNNNPPLPK	81	3.67	0	1.000	13992	14022	13962	13989	3′	11	10
143	agct	k	r2	SKTLLLMWTLILIQSTSGK	110	3.76	1	0.377	2757	2730	2727	2703	5′	10	9
144	agct	S	fO	ASLSVALDPSGSTNTSLTAKHPNQLASIYFSR	2	2.24	1	0.987	5775	5832	5730	5772	3′	20	12
145	agct	no	fO	KAPNPCLAICALDACMLEK	1222	3.24	1	0.448	5931	5955	5958	5985	5′	9	10
146	agct	S	fO	GGPAGSVFWATIQANPMASMTFSKSYSK	4	2.67	1	0.429	7608	7653	7572	7605	3′	16	12
147	agct	h	r0	TLLNKTSPTFTYLSGEAHLTCHGEK*	4	2.19	1	0.342	13599	13569	13641	13602	3′	11	14
148	agtc	h	fO	SLLPSLSTQHHRYSGWVPGGSGLDIHAPK	1	2.81	1	0.420	2445	2472	2475	2529	5′	12	17
149	agtc	r	f2	AATDDERPTTTGLNSSTTTLLLSR	5	2.66	0	0.303	5214	5247	5163	5211	3′	11	13
150	agtc	e	f1	LTEPLTNGESSWEKASGSMVLAAVLLK\$	155	2.13	1	0.932	11628	11658	11580	11625	3′	11	16
151	agtc	m	fO	IVAFSTSSQLGLMASEALAGMK	168	3.53	0	0.790	13458	13494	13497	13521	5′	13	9
152	agtc	W	fO	IVAFSTSSQLGLMASEALAGWK	434	4.72	0	0.715	13458	13494	13497	13521	5′	13	9
153	agtc	d	fO	IVAFSTSSQLGLMASEALAGDK	448	4.33	0	0.817	13458	13494	13497	13521	5′	12	10
154	agtc	r	r1	SLLPFVLFTFTPWSTLGLSMFLWVLGRGGLLFGR*	1	3.40	1	1.000	13758	13725	13821	13761	3′	13	21
155	agtc	no	r2	VAPVMSLIWFVLLPVSGPILEWCGRLMK	16	3.06	1	0.437	14934	14907	14991	14937	3′	9	19
156	atcg	k	r2	ATKTVGGVFGQTNQSPDPK*	1	2.68	1	0.251	321	294	291	267	5′	10	9
157	atcg	У	f1	SVGGGSAGYCVCGAAWGLGEPTKPQYHPPQFMYLTSSK	8	2.37	0	0.517	555	609	498	552	3′	19	19
158	atcg	e	r0	VISSEFIMQSQSPKHELEK	42	3.08	1	0.103	1629	1605	1602	1575	5′	10	9
159	atcg	а	f2	LYSQAFNSSSAQHTHGVGGCGVHWVRFEFK	1	2.25	1	0.669	3324	3369	3372	3411	5′	15	25
160	atcg	no	f2	LMPPLCKIHHESVALLVR	20	2.55	1	1	3936	3912	3909	3885	5′	9	9
161	atcg	g	r1	EKNQAVPEGPSMFISGPTQVK	3	2.47	1	0.573	4383	4353	4413	4386	3′	11	10
162	atcg	g	r1	EKNQAVPEGLAMFISGPTQVK	15	2.81	1	0.57	4380	4353	4413	4383	3′	11	10
163	atcg	g	r1	EKNQAVPEGLSMFISGPTQVK	20	2.80	1	0.57	4383	4353	4413	4386	3′	11	10
164	atcg	no	fO	AHTPKMLVMGPGLLPSGQGLGR	27	2.13	1	0.472	4503	4527	4530	4566	5′	10	12
165	atcg	Х	r1	SEASASGSAKAAHDHLDDHPMMXLLFFVNSSMMAHLGK	1	2.79	1	0.34	5115	5064	5175	5118	3′	18	20
166	atcg	q	r2	QDCCDQDGSDEDPSNQNPQPAPKQER	1	2.42	1	0.309	6315	6282	6279	6240	5′	12	14
167	atcg	v	r2	GPVTVQAKVVGSVGGVGGAMGVMR	203	4.83	1	0.074	8160	8115	8187	8163	3′	15	9
168	atcg	v	r2	<u>GPVTVQAKVLGSVGGVGGAMGVMR</u>	181	4.59	1	0.082	8157	8115	8187	8160	3′	14	10
169	atcg	a	f2	AASHPVPVPMTLLMLGLLTNTLTMYQWWR	24	2.66	0	0.053	9477	9537	9450	9474	3′	19	10
170	atcg	a	r0	AMTLHAHAGAMFSEPAVLWVAISAMSAGAEPTAVANAK	2	2.44	0	0.693	11196	11115	11226	11199	3′	28	10
171	atcg	q	r2	GDAGEMLLVNAGLLGAQFLLASK	28	3.47	0	0.068	13719	13695	13692	13653	5'	10	13
172	atcg	e	r2	GDAGEMLLVIAGLLGAEFLLASK	29	3.54	0	0.147	13719	13695	13692	13653	5'	10	13
173	atcg	t	rU	SAEHSLGAGYHSGLMWGGVFKGLATVILSGSPITSGENT	1	2.93	1	0.885	15546	15456	155/3	15549	3'	30	9
174	atgc	С	IZ		[ 	2.36	1	0.239	18/2	1911	1815	1869	3'	13	19
175	atgc	e	IU	SVESWILLGEENNFAEEAKAK	641	3.52	1	0.148	1902	1929	1872	1899	3'	11	12
176	atgc	W	IU 61	WDISQGKTFAVILNLVLYPHPPK	23	3.62	1	0.851	3240	3270	3204	3237	3'	11	12
170	atgc	K	11 61		154	3.19	1	0.419	4272	4293	4296	4323	5'	9	10
178	atgc	n	11	LMHHHYKSSAHHVHSPMIVHHNNYQYK	35	2.40	1	0.591	6501	0522	6444	6498	3'	10	18
1/9	atgc	e	r1		32	3.62	1	0.040	/155	7131	/128	7101	5'	10	9
180	alge	e	6 11		42	4.38	1	0.150	7100	7131	7128	7101	יכ גע	10	14
101	alge	110 £	12 f1		12	2.48	1	1 000	7524	7500	7482	/521	5'	15	14
182	alge	1	11	AQVVLFAFALFLKIVIPPPFEVIVIFHIVISIVIK	2 14	2.10	1	0.204	9000	9051	9054	9081	5'	10	10
104	alge	11	r1 r0		14	3.54	1	0.304	1298	11208	11205	1229	יכ גע	12	12
104	alge	ĸ	FU F1	SKINKED HNASSINEVININSGLK	232	3.49	1	0.793	12054	12033	1208/	1205/	3' 2/	9	11
100	atge	C	11 f1		1 12	2.13	1	0.023	13428	13452	13333	13423	3' 2/	9 21	24
100	atge	у	11		12	2.47	U	0.404	15941	14001	15914	12920	Э.	21	9

insertions, as determined by Lee and Johnson [46]) (therein Table 1), were used to predict abundances of swinger RNAs. Indeed, these experimental kinetic parameters predict several properties of swinger RNAs [58,93,94], strengthening the hypothesis that regular polymerases are responsible for swinger polymerizations, by switching to unusual, stabilized 'swinger' conformations, similar (or even identical) to conformations causing point misinsertions, but lasting longer.

The second testable prediction of the hypothesis is that the same polymerase produces regular and swinger-transformed sequences. Hence occasionally, polymerases switch in the midst of replication/ transcription, so that part of the sequence follows regular templating rules, and the other, contiguous part, is swinger transformed according to one among the 23 swinger rules. The fact that there are far more RNAs that are entirely swinger transformed than chimeric RNAs [100] suggests that such switches during polymerization are rare, and usually occur before or at the onset of polymerization.

The 16S rRNA gene in the complete mitogenome of *Kamimuria* wangi is a swinger transformed  $A \leftrightarrow T + C \leftrightarrow G$  DNA sequence, embedded within an otherwise regular insect mitogenome [99]. The reasons why until now only DNA matching this  $A \leftrightarrow T + C \leftrightarrow G$  swinger transformation has been detected, remain unknown. This  $A \leftrightarrow T + C \leftrightarrow G$  exchange rule is also common among chimeric (part regular, part swinger) RNAs [100].

Peptides encoded by chimeric transcripts are detected for the first time here. An example of three peptides and the corresponding DNA, chimeric RNA is described in Fig. 1. Underlined parts are translated after swinger-transformation of the transcribed RNA sequence. When a detected peptide corresponds only to all or part of the underlined amino acid sequence, this peptide is considered a swinger peptide, as the peptides described in an earlier publication [103]. When the detected peptide encompasses part of the underlined, and part of the contiguous amino acid sequence(s) that are not underlined in Fig. 1, the peptide is considered chimeric, because translated from untransformed RNA (the part that is not underlined in Fig. 1), and from RNA that is swinger translated (underlined part in Fig. 1).

### 1.4. Previously detected swinger RNA

An anonymous reviewer suggested to add, for reader convenience, explanations on how swinger sequences described in previous publications had been detected in GenBank. These methods are not described in the Materials and methods section, as this would be inadequate and confusing: neither results nor analyses beyond those described in earlier publications on RNA were done in the context of the presently described proteomic analyses. The aim here is (chimeric) peptide detection. Following descriptions are only for the convenience of potential readers.

In a first step, the 23 swinger versions of the human mitogenome were transformed in silico. This means that for swinger transformation  $A \leftrightarrow C$  (as an example), all As in the human mitogenome are replaced by the 'replace' function in the software Word by 'X'. Then all Cs are replaced by A. The last step is to replace all Xs by C, producing a hypothetical, A ↔ C swinger transformed human mitogenome. Similar procedures produce all 23 possible swinger versions of the mitogenome. Each of these is then analyzed by BLASTn [2]. Two types of analyses have been done. The first analyzes (publications by Seligmann on swinger sequences prior to 2016, from 2012 on) search for alignments between the swinger transformed mitogenome and various sequence databases in GenBank, using standard default megablast parameters. This resulted in detecting long, highly similar sequences, as described in [88,92-94], for example. Such searches do not yield alignments with nuclear chromosome sequences, but detect about 100 ESTs (expressed sequence tags). The length of the alignments (>100 nucleotides) and the similarity with the hypothetical swinger-transformed mitogenome versions (>90%), as these were previously presented (Table 1 in [88,92], and Table 2 in [93,94]), are not compatible with randomly obtained results (as tested by simulations based on randomly shuffled swinger mitogenomes in [96] (therein Section 2.2.3.)).

These long EST swinger sequences were then confirmed by sequences detected within sequence read archives (SRA) of the human transcriptome published by Garzon et al. [30]), GenBank SRA entries SRX768406–SRX768440. For these analyses pertaining to short RNA reads (50 nucleotides, RNA seq, Illumina), BLASTn searched for 'somewhat similar sequences', using default search parameters, and detected swinger reads as these are described in [103] (therein supplementary data). These results on short swinger reads converge with those obtained by the first, EST-focused search ([103], therein Figure 2). Using the same search tool and criteria as used for the RNA seq reads, the 23 swingertransformed mitogenome versions also align with nuclear chromosome sequences (they did not align with human nuclear chromosomes when using megablast as for the EST search).

# 1.5. Nuclear origins of swinger sequences (numts)

Alignments detected with relaxed criteria (BLASTn, see previous section) between swinger mitogenome sequences and each short reads and nuclear chromosome sequences suggest the possibility that swinger reads (but not swinger ESTs) could originate from the cytosol. To some extent, this is not relevant to the main issue at stake, the very existence of swinger polymerizations, but could be relevant because the very large human nuclear genome could by chance be the origin of these alignments, due to its size.

In addition, mitogenome copies (called numts, [51]) are inserted within nuclear chromosomes [36]. Because nuclear copies of regular mitogenomes exist, their occurrence for swinger-transformed mitogenomes is plausible, and would consist in itself a possible independent confirmation of the existence of swinger sequences, as previously discussed [103]. In addition, the possibility of swinger transcription of regular numts in the nucleus can't be ruled out.

Previous analyses [103] showed that the majority of detected swinger reads have mitochondrial origins. On average, alignments between swinger-transformed mitogenomes and RNA seq reads have higher identity percentages than between the same swinger mitogenome sequences and nuclear chromosome sequences. This is the case for a statistically significant majority of comparisons of identity percentages obtained between RNA reads and the swinger-transformed mitogenome, versus that between the same swinger-transformed mitogenome region and nuclear chromosome sequences. This suggests that most potential swinger numts diverge from their ancestral mitogenomic sequence, and that most RNA reads aligning with swinger-transformed mitogenomes have mitochondrial origins, because the swinger-transformed mitogenomes resemble on average more RNA reads than putative swinger numt(s) [103]. The 'Discussion' below develops these points in relation to potential nuclear origins of chimeric peptides.

#### 1.6. Chimeric RNAs due to fusion between different RNAs

The term 'chimeric' transcripts has been used in the literature for a different type of RNA than the contiguous regular- and swingertranscribed RNAs [101]. These other types of chimeric RNAs refer to two or more different transcripts produced each by regular polymerization, on the template of disjunct DNA regions. These RNAs are then fused by natural [60,132] or artificial reverse-transcription-associated phenomena [131]. These chimeric RNAs differ from the regularswinger RNAs in the sense that for the latter the transcription process is chimeric (part regular-, part swinger transcriptions), but not in terms of their templating DNA regions, which are contiguous, not disjunct. It is possible that some unknown sequencing artifacts produce some of the detected swinger reads, but the non-random mapping of detected swinger peptides on detected swinger RNA reads, as previously described [103] shows that most swinger reads exist while translation occurs in the cell, and hence are not artifacts.

#### 1.7. Swinger polymerization creates new genomic sequences

Another type of analyses detected swinger repeats within the regular mitogenome. Swinger repeats are usually short repeats that can only be detected when taking into account swinger transformations. These short sequences are inserted within the regular mitogenome, suggesting that natural retrotransposition of swinger RNAs produces novel DNA sequences [101]. They are more frequent and longer than expected by chance, and their length is proportional to the probability that the specific swinger transformation conserves circular code signals that presumably maintain the ribosomal translation frame in the gene. The natural circular code is a punctuation code within the genetic code consisting of 20 codons that as a group, have properties that enable protein coding frame retrieval [4,22,55–57]. This indicates that insertion of swinger sequences in the human mitogenome depends on their capacity to integrate protein coding genes without disrupting punctuation that presumably enables ribosomal detection of the coding frame.

### 1.8. Chimeric peptides

Recent analyses show convergent frequencies between swinger RNAs sequenced by classical and next generation (RNAseg) sequencing methods [103]. Hence swinger RNA occurrence is relatively well confirmed by data from independent methods and research teams. Here analyses complement at peptide levels results on chimeric transcripts. The existence of transcripts that are part regular (untransformed), part swinger RNAs, with an abrupt switch between these parts, predicts the existence of 'chimeric' peptides matching translation of such chimeric transcripts. Hence MS/MS mass spectra of peptide data (from [33]) previously used to detect swinger peptides [103] are reanalyzed here, using the same methods as by Seligmann [103]), searching for peptides matching in part the translation of the untransformed human mitogenome, and in part the translation of the swinger-transformed, contiguous mitogenome sequence. Chimeric peptides are peptides where swinger-encoded parts of a peptide are contiguous with parts translated from regular RNA. These would be stronger evidence for translation of swinger RNA than previous detections of entirely swingerencoded peptides because the regular encoded parts function as matched positive controls, directly associated with swinger-encoded parts. In addition, chimeric peptides could suggest that swinger peptides are integrated within otherwise regular proteins, a further small step to understand functions associated with swinger phenomena.

# 2. Materials and methods

The revised Cambridge reference sequence for the human mitochondrial genome (NC\_012920, [3]) was cut according to a running window of 270 nucleotides. Analyzes do not account for known mitochondrial polymorphisms, as this would expand analyzes beyond computing powers. The six frames of each of these nucleotide sequences of 270 bases were translated into the corresponding hypothetical peptides according to the vertebrate mitochondrial genetic code, after the 90 nucleotide-long mid-third of that sequence was swinger transformed according to each of the 23 swinger transformations. Hence each running window (around 16,300 in total) is represented by the six peptides translated from each of its 23 partly swinger-transformed versions (6 frames  $\times$  23 swinger versions = 138 hypothetical chimeric peptides for each of the 16,300 running windows). The window length of 90 codons/amino acids is designed to match the length of the longest (non-chimeric) peptides (up to 40 amino acids, [103]) previously detected in this dataset [33]. All translated hypothetical peptides are used by Thermo Protein Discoverer to predict a theoretical mass spectrometry distribution, which is matched with observed MS/MS mass spectrometry data from Gueugneau et al. [33]).

Stops are translated as 'X', which Thermo Proteome Discoverer considers by default as leucine/isoleucine (these have equal masses, and are indistinguishable by mass spectrometry). Peptides including stops are duplicated 18 times, replacing 'X' by one of the 18 remaining amino acid species, excluding leucine and isoleucine. Hence predicted peptides include the possibility that any amino acid could be inserted at stops. Analyses assume that all stops in a single predicted peptide are translated by the same amino acid. Hence the 138 peptides for a single window of 270 nucleotides, if it includes at least one stop (the majority of cases), are represented 19 times, inserting X and each of the remaining amino acids at stops (19 x 138 = 2622 chimeric peptides). In total, approximately 42.7 million hypothetical chimeric peptides were tested.

Consensus searches were handled with the Sequest (Thermo Fisher Scientific, Illkirch) algorithm with the following mass tolerances: Parent = 1 Da and Fragment = 0.5 Da (monoisotopic masses). Fixed carbamidomethyl (C) and variable Oxydation (M) modifications were activated, as well as the lysine  $\rightarrow$  pyrrolysine modification, and only one missed trypsin cleavage was allowed. False discovery rate was estimated against a reverse decoy database using the Percolator algorithm. No protein grouping was allowed since the database only contained non redundant entries. Peptides with false discovery rate q < 0.05 and score Xcorr > 1.99 were considered identified. The score Xcorr is a likelihood of match between expected and observed MS/MS data that is unaffected by peptide length. Further explanations on peptide detection and characterization by the software are given in the Discussion. Observed mass spectra were compared separately to predicted peptides 19 times, each time inserting a different amino acid at stops. Here analyzes test the existence of a specific group of peptides, namely chimeric peptides. The false discovery rate q is adapted to such populations of detected items [41]. Results also indicate the posterior error probability PEP, an estimate of detection error specific to each individual peptide, which might be useful in the future, when analyzes focus on specific peptides, rather than on a population of peptides. Results are not analyzed according to this criterion more adapted to studies focusing on specific individual peptides.

## 3. Results

# 3.1. Chimeric peptides

Analyses detect according to the filtering criteria 1301 chimeric peptides, among approximately 42.7 million chimeric peptides produced by combinations of stop codon-amino acid insertions, swinger transformations and frames for the running window of 270 nucleotides (illustration in Fig. 1). Hence chimeric peptides are detected for approximately 3 among 100,000 hypothetical chimeric peptides. This is 200 times less than the rate of detection for 'regular' non-chimeric swinger peptides, using the same criteria and the same data, approximately 6 per 1000 predicted swinger peptides [103].

Previously detected chimeric human mitochondrial RNAs are about 3% of all RNAs detected with at least some swinger part. Part of the discrepancy between chimeric RNA versus peptide detections probably results from the fact that proteomic analyses only considered abrupt switches between regular and swinger parts of peptides. Blast analyses detecting RNAs are not limited by this consideration, and can detect RNAs where the switch is not abrupt: in a transition sequence between regular and swinger transformed sequences, nucleotides seem random. Hence for practical reasons, detection of chimeric RNAs encompasses more possibilities than chimeric peptide detection, explaining lower rates of chimeric peptide detection (relative to rates of swinger peptide detection); these are lower than rates of chimeric RNA detection (relative to swinger RNA detection rates).

Here we focus specifically on chimeric peptides for which each regular and swinger parts have more than 8 amino acids. This is because considering 19 different amino acid species (merging leucine and isoleucine), the e value for 42.7 million potential chimeric peptides is about 0.0001 for amino acid sequences of 9 residues  $(42,000,000 \times 1/19^{-9})$ . Hence the match of each regular and swinger part of the detected peptide with the predicted chimeric peptide is unlikely to be due to chance, as estimated by this approximate e value.

This restricts the sample of 1301 detected chimeric peptides to 186 chimeric peptides of at least 18 residues, from various swinger transformations and stop-amino acid insertions (Table 1). Among these 186 chimeric peptides, the regular-encoded part of the peptide corresponds to the 5' part of the peptide for 41% of the 186 chimeric peptides. This means that a statistically significant majority of chimeric peptides (two tailed sign test, P = 0.0061) correspond to the 5' translation of swinger RNA and 3' translation of regular RNA. Note that this statistically significant bias could not occur if detected chimeric peptides were due to random detection artifacts, strengthening the suspicion that results reflect a biological reality. Hence 41% of chimeric peptides reflect translation of regular transcripts that switch at a given point to swinger transcription. Frequencies and mean lengths of chimeric peptides, for each swinger type (Table 2)

#### Table 2

Frequencies and lengths of chimeric swinger peptides detected in Table 1: Columns indicate: swinger type; peptide number; mean PSMs; mean amino acid number in peptide non-swinger part; mean amino acid number in peptide swinger part.

Swinger type	Ν	PSMs	Reg	Swinger
$A \leftrightarrow C$	3	22.0	17.0	10.3
$A \leftrightarrow G$	4	96.0	14.8	14.0
$A \leftrightarrow T$	5	209.6	11.6	10.0
$C \leftrightarrow G$	4	139.0	11.3	14.0
$C \leftrightarrow T$	7	225.1	14.7	11.7
$G \leftrightarrow T$	1	1.0	16.0	11.0
$A \leftrightarrow C + G \leftrightarrow T$	11	59.6	13.8	14.2
$A \leftrightarrow G + C \leftrightarrow T$	10	103.0	13.2	14.4
$A \leftrightarrow T + C \leftrightarrow G$	12	31.2	12.3	14.8
$A \to C \to G \to A$	10	52.2	14.6	13.7
$A \to C \to T \to A$	5	47.0	14.2	13.6
$A \to G \to C \to A$	9	91.9	14.1	10.2
$A \to G \to T \to A$	8	33.3	18.3	12.4
$A \to T \to C \to A$	7	230.3	12.1	14.7
$A \to T \to G \to A$	9	18.8	18.3	12.8
$C \to G \to T \to C$	8	52.4	18.4	12.9
$C \to T \to G \to C$	5	179.2	11.4	12.6
$A \to C \to G \to T \to A$	10	82.1	14.0	14.3
$A \to C \to T \to G \to A$	14	71.8	13.6	11.9
$A \to G \to C \to T \to A$	5	268.4	13.2	11.4
$A \to G \to T \to C \to A$	7	173.1	12.1	13.6
$A \to T \to C \to G \to A$	18	33.7	14.6	12.3
$A \to T \to G \to C \to A$	13	92.6	11.8	12.7

show that the regular (non-swinger) part of the chimeric peptides is on average slightly longer than the swinger part, though this difference is not statistically significant.

#### 3.2. Swinger peptides and chimeric peptides

Note that chimeric peptides, due to their part that matches translation of regular transcripts, differ in mass spectrometry properties from peptides entirely translated from swinger RNA, even if these have the exact same swinger sequence. Hence detection of chimeric peptides with swinger parts overlapping previously detected 'regular' swinger peptides (as the swinger peptides described by [103]) would be strong, independent methodological confirmation that positive results are not artifacts. Indeed, the swinger parts of eight chimeric peptides in Table 1 overlap with one of the 263 previously described swinger peptides [103]. These previously described swinger peptides cover on average 1.1% of the swinger-transformed mitogenome, expecting approximately 2 overlaps with chimeric peptides in Table 1 if no association exists between the two independent analyses. This means that chimeric peptides map on previously described swinger peptides 4 times more frequently than expected. This association between two independent searches confirms that results are not false positive matches between the mass spectrometry data and some predicted hypothetical chimeric peptides among a very large number of predicted hypothetical chimeric peptides. In addition, note that even if detected swinger and chimeric peptides correspond to the same swinger region, the corresponding MS/MS mass spectra differ, because for chimeric peptides mass spectra include also the adjacent residues translated from regular, untransformed RNA, while for swinger peptides, mass spectra do not include the latter residues. This non-random correspondence between swinger peptides and swinger parts of chimeric peptides suggests that translation of swinger RNAs is not random, and probably specific to some mitogenome regions.

# 3.3. Swinger RNA and chimeric peptides

Previously detected swinger peptides preferentially map on human mitogenome regions covered by independently detected swinger RNAs [103]. Their numbers increase with numbers of detected swinger transcripts. These positive associations between swinger RNA and swinger peptides can also be expected for chimeric peptides described in Tables 1–2. Such associations would confirm that the detected chimeric peptides actually exist, because they would match two independent material evidences, peptides, and RNA fragments.

The mean number of PSMs (peptide spectrum matches) for chimeric peptides increases as a function of the number of human mitogenome regions covered by swinger RNA (also called contigs), for the swinger type corresponding to the swinger part of the chimeric peptides (Fig. 2). Swinger transcriptomic data are from Seligmann [103]). Chimeric peptides presumably reflect translation of chimeric RNAs, along part regular, and part swinger transcription rules. Hence amounts of chimeric peptides should reflect numbers of possible transitions between regular and swinger RNAs, estimated by the number of swinger contigs previously described by Seligmann [103]). Indeed, a positive association between PSMs of chimeric peptides and swinger RNA contigs exists (r = 0.64, one tailed P = 0.0006), strengthening confidence in the validity of results, and corresponding with previous results for swinger peptides [103]. Note that similar correlation analyses for numbers (not PSMs) of detected chimeric peptides do not yield statistically significant associations with contig numbers.

The swinger part of 8 chimeric peptides (marked by \* in Table 1) maps on human mitogenome regions also covered by the adequate type of swinger RNA (six swinger types, two matches for  $A \rightarrow G \rightarrow T \rightarrow A$  and  $A \rightarrow C \rightarrow G \rightarrow T \rightarrow A$ , and one match for each  $A \rightarrow G \rightarrow C \rightarrow T \rightarrow A$ ,  $A \rightarrow G \rightarrow T \rightarrow C \rightarrow A$ ,  $A \rightarrow T \rightarrow C \rightarrow G \rightarrow A$ , and  $A \rightarrow T \rightarrow G \rightarrow C \rightarrow A$  swinger transformations). Considering the overall



**Fig. 2.** Mean number of PSMs detected for chimeric peptides, as a function of the number of disjunct human mitogenome regions covered by swinger RNA (RNA data from [102]). The positive association indicates the expected causal link between swinger RNA and chimeric peptides.

mitogenome coverage by swinger RNAs (on average 2.6% of the genome), lack of association between swinger RNAs and the swinger part of chimeric peptides would expect 4.76 matches across all 23 swinger transformations, with 0.21 peptides for the average swinger transformation. This predicted number for specific swinger transformations was always <0.5 peptides. Detecting at least one match for six among 23 swinger types, when less than 0.5 are expected for all 23 swinger transformations has P = 0.022 according to a two-tailed Fisher exact test. This indicates that chimeric peptides associate with detected swinger RNA, though this association is weaker than the previously described association between swinger RNA and swinger peptides [103].

### 3.4. Chimeric peptides: strong validation of swinger sequences

Chimeric peptides are in terms of confirmation of swinger polymerization only secondary evidence, because peptides are translated from RNA, as compared to previous descriptions of swinger RNAs and chimeric swinger RNAs [100], which directly result from swinger polymerization. This point is also valid for swinger peptides. However, detection of (numerous) peptides matching translation of contiguous parts of the mitogenome, where one part reflects regular transcription, and the other swinger transcription, is a strong methodological confirmation for swinger phenomena and associated translation into peptides, which is not implied by the detection of 'pure' swinger peptides. This is because the non-swinger part of the peptide is a positive control paired to its contiguous swinger part. Hence in addition of describing a further aspect of the biological phenomenon of swinger polymerizations, chimeric peptides are also a further validation of the phenomenon's existence.

## 3.5. Chimeric peptides integrated in regular proteins?

An important question associated to swinger sequences is their function: among others, do they code for functional proteins, and are swinger peptides integrated into regular, perhaps functional proteins? A reanalysis of Table 1 yields a first insight into these important questions. The regular (non-swinger) part of eleven peptides matches the sequence of six among thirteen known, regular, mitogenomeencoded proteins. Their swinger parts correspond to the translation

#### Table 3

Chimeric peptides from Table 1 with regular part matching proteins translated from known mitogenome-encoded genes. Swinger parts are underlined, gene identity is followed by the position of the 'normal' part of the peptide matching the regular translation of the gene in the regular protein. The swinger transformation and the amino acid inserted at stop(s) are also indicated. Peptide parts matching translation according to both nuclear and mitochondrial genetic codes are highlighted: peptide 100 could be translated in the cytosol on the base of RNA transcribed from mitochondrial inserts in the nuclear chromosome (numts), all remaining peptides could not, as at least on part of the peptide is incompatible with translation according to both there are fewer detected peptides compatible with the nuclear genetic code than expected, and more than expected peptides compatible only with the mitochondrial genetic code.

Table 1 #	Peptide	Gene	Position	Swinger rule	Stop
100	GASFLFIWNSLYLLFGAWAGVLGTALSLLIRAELGQPGNL	COX1	18-47	A > T > G	r
27	SGWVEWSRHSVLLLLSLPVLAAGITMLLTDR	COX1	205-213	$A\leftrightarrow C+G\leftrightarrow T$	S
181	LLKECLSLASVPATPPYHTFEEPVYMK	COX1	500-512	A > T > G > C	х
136	NQMIQALLITILLGLYFTLLSIVTAGTVFGLR	COX3	157-168	A > C > T > G	e
169	AASHPVPVPMTLLMLGLLTNTLTMYQWWR	COX3	41-59	A > T > C > G	a
23	DVSGPSSPSSSLMTLTLFSPDLLGDPDNYTLANPLNTPPY	Cyt B	238-267	$C \leftrightarrow T$	У
19	NPSLSISVPSTRHVSMPITISSIPPQTTEMCLMK	ND1	305-318	$C \leftrightarrow T$	t
38	WALFLSGTDSSSVSLAPLAATGSWGGLNQTQLR	ND2	165-176	$A \leftrightarrow G + C \leftrightarrow T$	n
34	SLKQNWDFSFNSSTMVVAGIFLLIR	ND5	249-262	$A\leftrightarrow C+G\leftrightarrow T$	f
80	IVAFSTSSQLGLMVLEVPVGVK	ND5	301-313	A > G > T > C	d
7	LLGAVPLASASLTIGSLALAGMPFLTGFYSKDHIIETANMS	ND5	374-402	$A \leftrightarrow G$	

of the contiguous swinger transformation of these genes, along nine (four symmetric, and five asymmetric) systematic nucleotide exchange rules (Table 3). Note that up to three chimeric peptides are detected for two large mitochondrial proteins (cytochrome c oxidase I and NADH:ubiquinone oxidoreductase subunit 5). It is plausible that such peptides are integrated within complete proteins. These sequence alterations could modulate (or not) the regular function of the protein, and not necessarily impair function.

These 11 chimeric peptides integrated in regular proteins represent 5.9% of all 186 detected chimeric peptides. Considering that regular mitochondrion-encoded proteins have a total length of 3789 amino acids, the regular proteins represent 11.43% of the total number of amino acids that could be translated from the positive and negative strands of the human mitogenome. This means that chimeric peptides embedded within regular coding sequences are half as frequent as expected (5.9 versus 11.43%). This principle is further strengthened when examining the number of PSMs (number of identified peptide spectra matching a hypothetical peptide) for these 11 regular-proteinintegrated chimeric peptides, as compared to the mean number of PSMs for all chimeric peptides detected for that swinger transformation: their PSMs is in all but one case (peptide 80 in Table 1) lower than the mean PSMs of other chimeric peptides for that swinger transformation. Hence chimeric peptides within regular proteins are rarer, and less expressed (as far as PSMs numbers can be trusted to reflect peptide abundances), than chimeric peptides translated from noncoding sequences, and non-coding frames of regular protein coding genes.

3.6. The natural circular code and swinger RNA, peptides and chimeric peptides

An anonymous reviewer suggested examining whether properties of chimeric peptides can be predicted from frameshift error-correcting properties of the natural circular code. Indeed, abundances of detected swinger RNAs in GenBank's EST database are proportional to reading frame retrieval (RFR) after swinger transformation of the natural circular code [58]. In this context, RFR, which estimates the capacity of the natural circular code to retrieve the protein coding frame, is calculated for the 20 codons that form the natural circular code, after each of the 23 swinger transformations: some codons belonging to the natural circular code are transformed into another codon included in the natural circular code, meaning that this property is invariant in relation to that codon and swinger transformation. RFR estimates this across all 20 codons of the natural circular code, for each swinger transformation. The length of swinger repeats in the human mitogenome is proportional to the RFR of the swinger transformation [101], which suggests that RFR affects insertion rates of swinger repeats in protein coding regions, and hence could also affect chimeric peptide production.

The association between RFR and swinger RNA abundances for EST sequences occurs also for mitogenome coverage by swinger RNA reads sequenced by RNAseg in the transcriptome by Garzon et al. [30]) (Pearson correlation coefficient r = 0.528, one tailed P = 0.005). For swinger peptides as described by Seligmann [103]), the mean number of PSMs also increases with RFR (r = 0.364, one tailed P = 0.044). This positive association between mean PSMs numbers and RFR is also detected for chimeric peptides from Table 1 (r = 0.367, one tailed P = 0.043). These two results are independent, also because mean PSMs of swinger and chimeric peptides are only weakly correlated (r = 0.24, P > 0.05). Hence detections of chimeric and swinger peptides are proportional to extents by which swinger transformations conserve natural circular code 'frame' punctuations. Note that RFR, as mitogenome contig numbers in a previous section, associate with mean PSMs, rather than numbers of detected peptides, suggesting that in the context of these specific data, PSMs are better quantitative estimates than other variables.

# 4. Discussion

### 4.1. Statistical validity of peptide detections by mass spectrometry

An anonymous reviewer of a previous version indicated that detection of peptides with masses approximately matching the numerous possibilities produced by translation of all potential chimeric RNAs could be due to chance, due mainly to the large number of hypothetical chimeric peptides. Indeed, considering all 19 possible amino acids inserted at stops introduces a 'fudge' factor that enables adapting many hypothetical peptides to an actual fragment with a similar mass. Note that 28 among 186 (15%) detected peptides lack stops, invalidating this argument for several detected chimeric peptides. Independently of this, there are three reasons why this important point does not invalid the remaining results on chimeric peptides presented here. This is first because mean chimeric peptide PSMs converge with corresponding swinger RNA contig numbers, an independent type of data unrelated to the problems of proteomic analyses, already discussed above.

The other two points relate to the nature of the MS/MS mass spectrometry analyses themselves. The factor 'detection by chance' is integrated into the detection software used by Thermo Proteome Discoverer. The software compares the match between the mass spectrum of the actual fragment and the predicted mass spectrum of the hypothetical peptide, and its match with a dataset of decoy (false, negative controls) predicted peptides. The q value estimates the false detection rate (FDR, see explanations by [41]) of a peptide based on comparing matches by the actual predicted peptides and the decoy peptide database. This q is a probability of detection corrected for the false detection rates within the population of positive results (classical P values consider the whole population of statistical decisions, not only the subpopulation of positives). Hence the reported detections account for matches due to chance, considering the various parameters of the samples analyzed/compared, among them in particular sample sizes.

The third point relates to the nature of the statistic whose distribution is used to evaluate the above mentioned q (FDR). It is Xcorr, the cross correlation of the goodness of fit between the experimental peptide fragments and theoretical mass spectra. This integrates fits with each b and y ions, which correspond to asymmetry in the physical fragmentation of peptide bonds within the detected peptide, resulting into shorter peptide subparts: b ions occur when the residue's N-terminal is charged, y ions when the C-terminal is charged. Hence the match between the observed and the predicted peptide is not based solely on the similarity between their total masses, but also on fit between distributions of masses of sub-fragments of the (expected and observed) peptides, and this separately for b and y ions. The Xcorr statistic accounts, in addition to peptide size, for the number of matching masses of such sub-fragments. This allows inferring more precisely the residue sequences in the peptide, and means that peptide detection is not based only on a single measure, its total mass, but also on the mass of several subfragments.

In this context, the peptide ACD can function as a simplified example. Its mass corresponds to six possible peptides, ACD, ADC, CAD, CDA, DAC and DCA. Hence if ACD results from translation of swinger RNA, one can't assert that the observed mass is due to this peptide rather than any of the other five possibilities. However, Xcorr also considers the masses of subfragments of this peptide. Detection of a subfragment matching the mass of AC excludes four among the six possible peptides. A fragment matching the mass of CD matches only two peptides. If both subfragments AC and CD are detected, the characterization of the peptide ACD can be considered as assessed.

In addition, this process is done separately for b and y ions, because mass spectrometry analyses are in principle sufficiently precise to distinguish between these ions (remember that the precision of 0.5 Da of the analyzed data means a precision of half the mass of a hydrogen atom, which is also far less than the difference between amino acids with similar molecular masses). Hence Xcorr integrates information from both b and y ions, evaluating whether that information is congruent with the observed data. This procedure, coupled to q values based on comparisons of the Xcorr distribution obtained for negative controls (decoy peptides), renders detections relatively robust, despite fuzzy factors. In fact, large numbers of predicted peptides are necessary to estimate properly the distribution of random Xcorrs. The last point stresses that q (as P) values account for numbers of predicted peptides.

#### 4.2. Confirmation of chimeric peptides by Waters technology

An anonymous reviewer suggested to confirm the existence of chimeric peptides by additional, independent mitochondrial proteomic data. In this context, I focused on another analysis of trypsinized human mitochondrial peptides [1], extracted by a more up to date MS/MS technology (Waters, Milford, MA, http://www.waters.com). This technique yields more accurate mass estimates than the method used by [33] (0.5 Da for the latter versus 5 ppm for the Waters method, hence about 10× more accurate estimates).

Analyses of the twelve samples from Alberio et al. [1]) by the software PLGS yield relatively few hits matching chimeric peptides considering only peptides where each regular and swinger-encoded parts are each at least nine amino acids long. One peptide matches significantly according to PLGS a chimeric peptide whose swinger part (underlined) matches swinger transformation  $A \rightarrow T \rightarrow G \rightarrow A$ , LVSASVEMNQQQVPGSAGR (the regular part are residues 4228–4237 translated from the third frame of the negative strand of the

human mitogenome). The other peptide detected in these data has a swinger part that matches transformation  $C \rightarrow G \rightarrow T \rightarrow C$ , <u>SAAAARAGSACCLTSTAVTDRNLNTTF</u>, the regular-encoded part corresponds to COX1, residues 211–219 in that regular mitochondrion-encoded protein.

Hence a different technology detects within independent mitoproteomic data peptides matching translations of chimeric RNAs, with one part regular, the other swinger transformed RNA. Hence, at least qualitatively, these independent data and technology confirm the existence of chimeric peptides and their integration in regular mitochondrial proteins. A more detailed description of 'regular' swinger peptides (meaning peptides entirely coded by swinger transformations of the mitogenome (unlike chimeric peptides that are in part regular-, in part swinger-encoded)) detected in the data from Alberio et al. [1]) will be presented elsewhere.

These results from data by Alberio et al. [1]) are too scarce to indicate whether chimeric peptides are produced according to a non-random profile. However, the non-random convergence between chimeric and entirely swinger peptides (detected in the same dataset from [33]) noted in a previous section in Results is in itself an indication that swinger-encoded peptides or parts of peptides are non-randomly produced.

### 4.3. Nuclear mitogenome copies

Previous transcriptomic analyses that detected non-canonical RNAs transformed according to systematic rules, such as deletions of monoand dinucleotides after each transcribed trinucleotide (producing delRNAs, [102]), and swinger transformations [103], included controls that account whether the transformed mitogenome versions match nuclear chromosome sequences: mitogenome analyses are frequently contaminated by such chromosomic pseudogenes [9,10,48–51,62,66, 67,123–125,133,134].

These previous analyses blasted the swinger-transformed mitogenome versions versus the (regular) human nuclear chromosomes. For transformed mitogenome regions aligning with both transcriptomic reads and chromosomes, similarities between the transformed mitogenome and the RNA contigs were compared with the corresponding similarities between the same transformed mitogenome region and the chromosomes. For each del- and swinger RNAs (non-canonical RNAs), similarities with RNA contigs were greater than those with chromosome sequences in significant majorities of cases [102,103], as already discussed above for swinger RNA reads.

These results indicate two major issues. First, overall, RNA contigs result from non-canonical transcriptions of the mitogenome, the point that was being tested. Second, the observation that chromosome sequences match transformed versions of the mitogenome suggests that chromosomes include inserts of mitogenomic origins that were transformed according to systematic rules. The observation that these are on average less similar to the transformed mitogenome than RNA contigs suggests that these transformed mitochondrial sequences inserted in nuclear chromosomes mutated apart from the original sequence, as expected for inserts lacking function in the cell's nucleus [16,24,29,34–36,38,40,52,54,63,65,71,77,113,116–119,126,128,129].

# 4.4. Peptides translated according to nuclear or vertebrate mitochondrial genetic codes

Similar-minded analyses at the peptide level can test whether chimeric peptides in Table 1 were translated according to the human mitochondrial or the nuclear genetic codes. For that purpose, the regular and swinger transformed versions of the human mitogenome were translated according to the standard genetic code, which differs from the vertebrate mitochondrial genetic code by the reassignment of codon ATA from Met to Ile, of TGA from Trp to stop, and AGR from stop to Arg [23]. These four codons are 6.25% of all 64 codons.

Each swinger- and regular-encoded part of detected chimeric peptides has at least 9 amino acids. Hence the probability of detecting chimeric peptides that would have identical sequence according to both genetic codes is  $(1 - 0.0625)^{-k}$ , where k is the total length of the peptide. This principle is applied to the chimeric peptides in Table 1 so as to calculate the predicted number of peptides, for each size category, that is expected to match translation according to both nuclear and mitochondrial genetic codes. Lengths of chimeric peptides in Table 1 range from 18 to 42 residues, a total of 24 length categories. The observed number of chimeric peptides compatible with translation according to both genetic codes (in total 30 among the 186 chimeric peptides) is lower than expected in 16 among 24 size categories. Obtaining this result has P = 0.038 according to a one-tailed sign test. This means that, considering the length of chimeric peptides, there are statistically significantly fewer than expected peptides with sequences compatible with translation according to the nuclear genetic code.

The same principle can be applied to chimeric peptides in Table 1 whose sequences are only compatible with translation according to the mitochondrial genetic code, separately for each the regular- and the swinger-encoded parts. Here, the observed number (54) should be larger than the predicted number, if the sample is biased towards mitochondrion-encoded/translated peptides. Considering that 6.25% of codons differ in codon-amino acid assignments between the two genetic codes, the total expected number of chimeric peptides, considering their size, containing at least one of the 4 codons with coding assignment differing between nuclear and mitochondrial genetic codes is 35.97. This number is far lower than the observed 54 according to a chi-square test (P = 0.0027). Hence chimeric peptides with sequences compatible only with translation according to the mitochondrial genetic code are significantly more frequent than expected. This bias confirms the mitochondrial origin of chimeric peptides in Table 1. The number of peptide length categories where more observed peptides than expected are only compatible with mitochondrial translation is again 16 among 24 length categories, which has P = 0.038 according to a one tailed sign test.

These analyses show that detected peptides are more likely translated according to the mitochondrial genetic code than according to the nuclear genetic code. Note that translation, within the mitochondrion, according to the nuclear code is possible: it potentially depends for some codons upon the presence of cytosolic tRNAs, which could be occasionally imported in mitochondria [21,32,39,44,75,76,78,112,114]. However, this rationale is not symmetric: cytosolic translation according to the mitochondrial genetic code is much less probable than the opposite, so that nuclear origins are not compatible with the results obtained.

In fact, whether peptides have cytosolic or mitochondrial origins does not actually affect the main point that is addressed here, which is that these peptides were translated in part from swinger RNA. The same point applies to the potential nuclear (numt) origin of swingertransformed mitochondrial DNA: independently of the location of the process, detection of chimeric peptides implies that swinger transformations occurred, whether during transcription of the regular mitogenome or nuclear inserts, or during numt insertion, possibly by natural swinger retrotranscription. This does not exclude the possibility that some detected chimeric peptides originate from the cytosol, but stresses the fact that most are mitochondrial, and that this issue is not directly relevant to the fact that swinger RNAs, chimeric RNAs, and corresponding peptides, exist, independently of the question of which cellular compartments produce them.

# 4.5. Few chimeric peptides in regular proteins translated from mitochondrion-encoded genes

Chimeric peptides in Table 3 have regular parts that match sequences of regular mitochondrial proteins encoded by mitochondrial genes. These are about 5% of all 186 detected chimeric peptides.

Peptides translated from regular mitochondrial genes represent about 11% of the total length potentially translated from the complete mitogenome, considering all six frames. Hence these 11 chimeric peptides potentially integrated in regular mitochondrial proteins are half as frequent as one could expect. Their PSMs is lower than for other chimeric peptides. These are hence rarer and less expressed than one could expect. Possibly, chimeric peptides integrated in regular proteins perturb proper protein folding. Incorrect folding induces various degradation mechanisms associated with mitophagy [5,42,122], which could explain that only few chimeric peptides are detected within regular proteins. These findings are not incompatible with the possibility that at least some swinger transcripts and peptides are functional.

# 4.6. Secondary structure formation by swinger transformed RNA and swinger RNA detection

Secondary structure formation by self-hybridization of DNA/RNA groups bijective transformations into three classes of each eight transformations. These share self-hybridization properties within each class [27,28]. This means that seven bijective transformations (including  $A \leftrightarrow T + C \leftrightarrow G$ ) conserve self-hybridization properties of the original, untransformed sequence. Secondary structure formation by swinger RNA associates with swinger RNA detection [104], but these groupings/properties do not correlate with differences in chimeric or swinger peptide abundances/PSMs (not shown). The issue of regulation of alternative mitochondrial transcriptions, respectively post-transcriptional splicing, in relation to secondary structure formation by transformed RNA [61] remains unclear: a positive association exists between RNA occurrence and secondary structure formation for regular and swinger RNAs, but for transcripts resulting from systematic deletions (delRNAs), a negative association exists between secondary structure formation after deletions and delRNAs [105].

# 4.7. Swinger transformations, RNA–DNA differences (RDDs) and heteroplasmy

Specific non-random point differences occur between DNA and RNA sequences, either due to nucleotide substitutions [47] or inserts/ deletions [15], including for human mitochondrial transcripts [8,37, 59]. These RDDs appear shortly after transcripts exit polymerases [130], suggesting RDDs are due to post-transcriptional edition. The systematic repetition of transformations over long sequence stretches that characterize swinger RNA seem less likely produced by posttranscriptional edition than some unusual stabilized polymerase state, however, at this point, no possibility can be excluded, and potential connections of del- and swinger RNAs with RDDs should be kept in mind.

For the same reason, and by definition, punctual mitochondrial heteroplasmies [45,74,115,121] could not account for swinger parts of chimeric peptides, because these have to be translated from sequences differing from standard mitogenome sequences by far more than punctual nucleotide substitutions. Mitochondrial length heteroplasmies are common (49% of individuals, [64]), and in principle could, by chance correspond to swinger-like inserts in the mitogenome. Considering the seven regions containing length heteroplasmies described by Ramos et al. [64]) (therein table 3), only three among 186 chimeric peptides in Table 1 (peptide numbers 3, 156 and 157) potentially overlap (and this only in part) with these length heteroplasmies. Hence length heteroplasmies map non-randomly on chimeric peptides (3 among seven). Hence some presumed chimeric peptides might be translated from regions presenting length heteroplasmies, but this explanation is compatible with, at most, a small minority of chimeric peptides. Hence heteroplasmy could not explain chimeric peptides.

# 4.8. Translation increasing codon size or transcription systematically deleting nucleotides

It is important to note in the broader context of the discussion of results that further little known mechanisms increase the coding potential of sequences. A different, sometimes tRNA-based mechanism produces an alternative decoding of sequences, that of systematic frameshifting, which expands the codon from three to four (or five) nucleotides [6], called here tetracodons or pentacodons. This could result from systematic ribosomal slippages, a phenomenon that would correspond to programmed frameshifts (e.g. [25,43]), but occurring systematically, and serially; and/or from translational activities of tRNAs with expanded anticodons [68,120,127]. These cases relate to previously described isolated frameshift mutations, interpreted as isolated tetra-, pentacodons.

The hypothesis of an early genetic code based on quadruplets was suggested by Baranov et al. [6]) to solve the problem that the weak triplet codon-anticodon interactions could not occur from a thermodynamic point of view in the absence of ribosomes, especially if these occurred at high temperatures [17–20]. Molecules as complex as ribosomes probably were absent at proto-life stages. Codon-anticodon interactions between four (or more) base pairs are more stable than those between three base pairs. Symmetry considerations also enable the deduction that the primeval genetic code was based on a subset of 64 quadruplets, called the tesserae, specifically for the vertebrate mitochondrial genetic code [31].

The expanded codon hypothesis is that modern genes include overlapping coding regions that consist of series of tetra- or pentacodons. This hypothesis is compatible with bioinformatic analyses where all eight frames of mitochondrial genes were translated assuming tetracodons. Blast analyses detected alignments between parts of these hypothetical tetracoded peptides and regular proteins in GenBank. Several other analyses based on codon usages in these tetracoding sequences confirm their special coding status, including higher GC contents than in non-tetracoding neighboring mitochondrial sequences. This corresponds to the prediction that tetracoding is an adaptation to translation at high temperature [89]. This point was further confirmed by a positive correlation between predicted tetracoding in lizard mitogenomes and mean body temperature in these lizard species [109]. Accordingly, overlap coding by tetracodons increases with temperature.

At this point, and besides the proven existence of decoding mechanisms for isolated tetracodons, the strongest further evidence for the existence of protein coding regions based on tetracodons is the coevolution between predicted tetracoding regions and the predicted antisense mitochondrial tRNAs with expanded anticodons, which is observed in mammal and Drosophila mitochondria [88,90,95]. In addition, mitochondrial peptides matching translation of regular and swinger RNAs according to tetra- and pentacodons have been detected [103], as well as translation of delRNAs (or dRNAs), RNAs transcribed while systematically deleting every fourth, or every fourth and fifth nucleotide. Peptide translation of such transcripts uses regular tRNAs but produces peptides identical to those resulting from decoding by tRNAs with expanded anticodons of regular transcripts [102]. These delRNAs are produced by systematic deletions, every third nucleotide, and correspond at deletion level, to systematic nucleotide substitutions/ exchanges. This predicts that chimeric peptides consisting in part of regular-translated, and in part tetra- or pentacoded peptides, might exist.

The strongest evidence for swinger-encoding is the association between detected swinger RNAs and detected swinger peptides. Analyses detecting mass spectra matching predictions according to translations of tetra- and pentacodons suffer the caveat that evidence is based solely on mass spectra, with the above discussed difficulties in asserting the robustness of results based only on proteomics. However, further analyses detected peptides matching translations, according to expanded codons, of swinger-transformed sequences, and showed their association with detected swinger RNA [103]. Hence from a methodological point of view, translation according to expanded codons of swinger RNAs is stronger evidence for tetra- and pentacoding than such translation of regular RNA because it is confirmed by the independent detections of two 'unusual' types of molecules, swinger RNA and corresponding peptides matching expanded codons.

# 4.9. Robustness of experimental design

An anonymous reviewer indicated that analyzes comparing transcriptome and proteome make sense only if data originate from individuals with the same phenotypes, and if possible the same tissues and even the same individual(s), however analyzes compare tumor transcriptome [30] with normal proteome [33]. This setup is indeed suboptimal. However, considering this point, RNA and peptide data converge (also in previous analyzes, [102,103]) despite that RNA and peptide data originate from different tissues/individuals/phenotypes. This indicates that the phenomenon is general, and robust. This should not be surprising, because analyzes consider only RNA and peptides corresponding to the mitogenome. Most tissue-specific differences in mitochondrial RNA and protein profiles relate to molecules imported from the cytosol [13].

Methods used to detect the various types of unusual peptides take into account the large numbers of possibilities in matching observed and hypothetical mass spectra, so that positive detections are robust, and could not be due to chance.

Beyond methodological issues, occurrence of peptides coded by combinations of presumably unusual coding systems (translation of stops, together with translation according to expanded codons, and this for swinger RNAs), suggests that these basically ignored mechanisms expand more frequently than presumed the coding potential of genes, at least of the short mitogenomes. Detections of chimeric peptides, consisting of peptide parts corresponding to regular translation, adjacent to peptide parts matching translation of contiguous swinger RNA, strengthen confidence in the validity of results as positive controls, and expand our understanding of the phenomenon: swinger peptides are occasionally integrated in regular mitochondrion-encoded proteins, but their occurrence is downregulated.

# 5. Conclusions

- 1. Analyses of MS/MS mass spectrometry data detect peptides matching the translation of chimeric transcripts, RNA following in part regular, and in part swinger-transformed transcription, assuming abrupt switches between regular and swinger transformed parts of the RNA.
- 2. The 186 detected chimeric peptides (peptides consisting of a part encoded by regular RNA and a contiguous part encoded by swinger RNA) represent 3/100,000 among potential chimeric peptides, about 200 times fewer (6/1000) than detected swinger peptides (peptides entirely encoded by swinger RNA) in the same data. Eleven among these 186 chimeric peptides have a regular-encoded part that corresponds to proteins translated from classical mitochondrionencoded genes.
- Chimeric peptides map on previously detected swinger RNA. This association is weaker than a previously described association between 'regular' swinger peptides and swinger RNAs [103].
- 4. The vertebrate mitochondrial genetic code differs from the nuclear genetic code for four codons. Numbers of detected chimeric peptides that could be translated from human mitogenome sequences according to the nuclear genetic code are significantly fewer than expected considering the differences between the two genetic codes. This means that the majority of detected chimeric peptides are not cytosolic contaminations and were translated in the mitochondrion.

5. Previous detections of swinger peptides (predicted products of translation of swinger RNA) suggested that swinger transformed RNA is translation-competent [103]. Chimeric peptides where the regular part corresponds to known mitochondrion-encoded proteins might be incorporated into the respiratory chain complexes. Chimeric and swinger peptides might affect known mitochondrial functions despite low abundances if they have regulatory functions. Results are compatible with the possibility that some proteins are encoded by swinger transformations, with yet unknown functions.

#### Acknowledgments

This work has been carried out thanks to the support of the A\*MIDEX project (no ANR-11-IDEX-0001-02) funded by the « Investissements d'Avenir » French Government program, managed by the French National Research Agency (ANR). I thank Nicolas Armstrong for coaching the mass spectrometry analyses, and anonymous reviewers for constructive comments.

### References

- Alberio T, Bondi H, Colombo F, Alloggio I, Pieroni L, Urbani A, et al. Mitochondrial proteomics investigation of a cellular model of impaired dopamine homeostasis, an early step in Parkinson's disease pathogenesis. Mol BioSyst 2014;10:1332–44.
- [2] Altschul SF, Madden TL, Schaeffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 1997;25:3389–402.
- [3] Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N. Reanalaysis and revision of the Cambridge reference sequence for human mitochondrial DNA. Nat Genet 1999;23:147.
- [4] Arquès DG, Michel CJ. A complementary circular code in the protein coding genes. J Theor Biol 1996;182:45–58.
- [5] Ashrafi G, Schwarz TL. The pathways of mitophagy for quality control and clearance of mitochondria. Cell Death Differ 2013;20:31–42.
- [6] Baranov PV, Venin M, Provan G. Codon size reduction as the origin of the triplet genetic code. PLoS One 2009;4:e5708.
- [7] Barthélémy R-X, Seligmann H. Cryptic tRNAs in chaetognath mitochondrial genomes. Comput Biol Chem 2016;62:119–32.
- [8] Bar-Yaacov D, Levin AG, Richards AL, Hachen N, Rebolledo Jaramillo B, Nekrutenko A, et al. RNA–DNA differences in human mitochondria restore ancestral form of 16S ribosomal RNA. Genome Res 2013;23:1789–96.
- [9] Bensasson D, Zhang D, Hartl DL, Hewitt GM. Mitochondrial pseudogenes: evolution's misplaces witnesses. Trends Ecol Evol 2001;16:314–21.
- [10] Bensasson D, Feldman MW, Petrov DA. Rates of DNA duplication and mitochondrial DNA insertion in the human genome. J Mol Evol 2003;57:343–54.
- [11] Beznosková P, Gunišová S, Valášek LS. Rules of UGA-N decoding by nearcognate tRNAs and analysis of readthrough on short uORFs in yeast. RNA 2016; 22:456–66.
- [12] Breton S, Milani L, Ghiselli F, Guerra D, Stewart D, Passamonti M. A resourceful genome: updating the functional repertoire and evolutionary role of animal mitochondrial DNAs. Trends Genet 2014;30:555–64.
- [13] Calvo SE, Clauser KR, Mootha VK. MitoCarta2.0 : an updated inventory of mammalian mitochondrial proteins. Nucleic Acids Res 2016;44:D1251–7.
- [14] Capt C, Passamonti M, Breton S. The human mitochondrial genome may encode for more than 13 proteins. Mitochondrial DNA 2015. <u>http://dx.doi.org/10.3109/</u> 19401736.2014.1003924.
- [15] Chen C, Bundschuh R. Systematic investigation of insertional and deletional RNA– DNA differences in the human transcriptome. BMC Genomics 2012;13:616.
- [16] Dayama G, Emery SB, Kidd JM, Mills RE. The genomic landscape of polymorphic human nuclear mitochondrial insertions. Nucleic Acids Res 2014;42:12640–9.
- [17] Di Giulio M. The universal ancestor lived in a thermophilic or hyperthermophilic environment. J Theor Biol 2000;203:203–13.
- [18] Di Giulio M. The universal ancestor was a thermophile or a hyperthermophile. Gene 2001;221:425–36.
- [19] Di Giulio M. The universal ancestor and the ancestor of bacteria were hyperthermophiles. J Mol Evol 2003;57:721–30.
- [20] DI Giulio M. The universal ancestor was a thermophile or a hyperthermophile: tests and further evidence. J Theor Biol 2003;221:425–36.
- [21] Duchêne AM, Pujol C, Maréchal-Drouard L. Import of tRNAs and aminoacyl-tRNA synthetases into mitochondria. Curr Genet 2009;55:1–18.
- [22] El Soufi K, Michel CJ. Circular code motifs near the ribosome decoding center. Comput Biol Chem 2015;59A:158–76.
- [23] Elzanowski A, Ostell J. The genetic codes. NCBI. 2013 accessed 10V2016 http://www.ncbi.nlm.nih.gov/Taxonomy/Utils/wprintgc.cgi?mode=t.
- [24] Ermakov OA, Simonov E, Surin VL, Titov SV, Brandler OV, Ivanova NV, et al. Implications of hybridization, NUMTs, and overlooked diversity for DNA barcoding of Eurasian ground squirrels. PLoS One 2015;10:e0117201.
- [25] Farabaugh PJ. Programmed translational frameshifting. Microbiol Rev 1996;60: 103–34.

- [26] Faure E, Delaye L, Tribolo S, Levasseur A, Seligmann H, Barthélémy RM. Probable presence of an ubiquitous cryptic mitochondrial gene on the antisense strand of the cytochrome oxidase I gene. Biol Direct 2011;6:56.
- [27] Fimmel E, Danielli A, Struengmann L. On dichotomic classes and bijections of the genetic code. J Theor Biol 2013;336:221–30.
- [28] Fimmel E, Giannerini S, Gonzalez DL, Struengmann L. Circular codes, symmetries and transformations. J Math Biol 2015;70:1623-164.
- [29] Françoso E, Gomes F, Arias MC. A protocol for isolating insect mitochondrial genomes: a case study of NUMT in *Melipona flavolineata* (Hymenoptera: Apidae). Mitochondrial DNA A DNA MappSeq Anal 2016;27:2401–4.
- [30] Garzon R, Volinia S, Papaioannou D, Nicolet D, Kohlschmidt J, Yan PS, et al. Expression and prognostic impact of lncRNAs in acute myeloid leukemia. Proc Natl Acad Sci U S A 2014;111:18679-1868.
- [31] Gonzalez DL, Giannerini S, Rosa R. On the origin of the mitochondrial genetic code: towards a unified mathematical framework for the management of genetic information. Nat Precedings 2012. http://dx.doi.org/10.1038/npre.2012.7136.1.
- [32] Gowher A, Smirnov A, Tarassov I, Entelis N. Induced tRNA import into human mitochondria: implication of a host aminoacyl-tRNA-synthetase. PLoS One 2013; 8:e66228.
- [33] Gueugneau M, Coudy-Gandilhon C, Gourbeyre O, Chambon C, Combaret L, Polge C, et al. Proteomics of muscle chronological ageing in post-menopausal women. BMC Genomics 2014;15:1165.
- [34] Hazkani-Covo E, Graur D. A comparative analysis of numt evolution in human and chimpanzee. Mol Biol Evol 2007;24:13–8.
- [35] Hazkani-Covo E. Mitochondrial insertions into primate nuclear genomes suggest the use of numts as a tool for phylogeny. Mol Biol Evol 2009;26:2175–9.
- [36] Hazkani-Covo E, Zeller RM, Martin W. Molecular poltergeists: mitochondrial DNA copies (numts) in sequenced nuclear genomes. PLoS Genet 2010;6:e1000834.
- [37] Hodgkinson A, Idaghdour Y, Gbeha E, Grenier JC, Hip-Ki E, Bruat V, et al. High resolution genomic analysis of human mitochondrial RNA sequence variation. Science 2014;344:413–5.
- [38] Hu QX, Fan Y, Xu L, Pang W, Wang S, Zheng YT, et al. Analysis of the complete mitochondrial genome and characterization of diverse NUMTs of *Macaca leonina*. Gene 2015;571:279–85.
- [39] Igloi GL, Aldinger CA. Where have all the inosines gone? Conflicting evidence for A-to-I editing of the anticodon of higher eukaryotic tRNAACGArg questions the dogma of a universal wobble-mediated decoding of CGN codons. IUBMB Life 2016;68:419–22.
- [40] Jensen-Seaman MI, Wildschutte JH, Soto-Calderón ID, Anthony NM. A comparative approach shows differences in patterns of numt insertion during hominoid evolution. J Mol Evol 2009;68:688–99.
- [41] Käll L, Storey JD, MacCoss MJ, Noble WS. Posterior error probabilities and false discovery rates: two sides of the same coin. J Proteome Res 2008;7:40–4.
- [42] Kim S-H, Park Y-Y, Yoo Y-S, Cho H. Self-clearance mechanism of mitochondrial E3 ligase MARCH5 contributes to mitochondria quality control. FEBS J 2015. <u>http://dx. doi.org/10.1111/febs.13568</u>.
- [43] Klobutcher LA, Farabaugh PJ. Shifty ciliates: frequent programmed translational frameshifting in euplotids. Cell 2002;111:763–6.
- [44] Koley S, Adhya S. A voltage-gated pore for translocation of tRNA. Biochem Biophys Res Commun 2013;439:23–9.
- [45] Korpelainen H. The evolutionary processes of mitochondrial and cholorplast genomes differ from those of nuclear genomes. Naturwissenschaften 2004;91: 505–11.
- [46] Lee HR, Johnson KA. Fidelity of the human mitochondrial DNA polymerase. J Biol Chem 2006;281:36236–40.
- [47] Li M, Wang IX, Li Y, Bruzel A, Richards AL, Toung JM, et al. Widespread RNA and DNA sequence differences in the human transcriptome. Science 2011;333: 53–8.
- [48] Lopez JV, Cevario S, O'Brien SJ. Complete nucleotide sequences of the domestic cat (*Felis catus*) mitochondrial genome and a transposed mtDNA tandem repeat (Numt) in the nuclear genome. Genomics 1996;33:229–46.
- [49] Lopez JV, Cumver M, Stephens JC, Johnoson WE, O'Brien SJ. Rates of nuclear and cytoplasmic mitochondrial DNA sequence divergence in mammals. Mol Biol Evol 1997;14:277–86.
- [50] Lopez JV, Stephens JC, O'Brien SJ. The long and short nuclear mitochondrial DNA (Numt) lineages. Trends Ecol Evol 1997;12:114.
- [51] Lopez JV, Yukhi N, Masuda R, Modi W, O'Brien SJ. Numt, a recent transfer and tandem amplification of mitochondrial DNA to the nuclear genome of the domestic cat. J Mol Evol 1994;39:174–90.
- [52] Mao X, Dong J, Hua P, He G, Zhang S, Rossiter SJ. Heteroplasmy and ancient translocation of mitochondrial DNA to the nucleus in the Chinese Horseshoe Bat (*Rhinolophus sinicus*) complex. PLoS One 2014;9:e98035.
- [53] Martinez-Rodriguez L, Erdogan O, Jimenez-Rodriguez M, Gonzalez-Rivera K, Williams T, Li L, et al. Functional class I and II amino acid-activating enzymes can be coded by opposite strands of the same gene. J Biol Chem 2015;290: 19710–25.
- [54] Michalovova M, Vyskot B, Kejnovsky E. Analysis of plastid and mitochondrial DNA insertions in the nucleus (NUPTs and NUMTs) of six plant species: size, relative age and chromosomal localization. Heredity 2013;111:314–20.
- [55] Michel CJ. Circular code motifs in transfer and 16S ribosomal RNAs: a possible translation code in genes. Comput Biol Chem 2012;45:17–29.
- [56] Michel CJ. An extended genetic scale of reading frame coding. J Theor Biol 2015; 365:164–74.
- [57] Michel CJ. The maximal C(3) self-complementary trinucleotide circular code X in genes of bacteria, eukaryotes, plasmids and viruses. J Theor Biol 2015;380: 156–77.

- [58] Michel CJ, Seligmann H. Bijective transformation circular codes and nucleotide exchanging RNA transcription. Biosystems 2014;118:39–50.
- [59] Moreira S, Valach M, Aoulad-Aissa M, Otto C, Burger G. Novel modes of RNA editing in mitochondria. Nucleic Acids Res 2016;44:4907–19.
- [60] Frenkel-Morgenstern M, Gorohovski A, Lacroix V, Rogers M, Ibanez K, Boullosa C, et al. ChiTaRS: a database of human, mouse and fruit fly chimeric transcripts and RNA-sequencing data. Nucleic Acids Res 2013;41:D142–51.
- [61] Ojala D, Montoya J, Attardi G. tRNA punctuation model of RNA processiong in human mitochondria. Nature 1981;290:470-4.
- [62] Olson LE, Yoder AD. Using secondary structure to identify ribosomal numts: cautionary examples from the human genome. Mol Biol Evol 2002;19:93–100.
- [63] Ramos A, Barbena E, Mateiu L, del Mar González M, Mairal Q, Lima M, et al. Nuclear insertions of mitochondrial origin: database updating and usefulness in cancer studies. Mitochondrion 2011;11:946–53.
- [64] Ramos A, Santos C, Mateiu L, del Mar Gonzalez M, Alvarez L, Azevedo L, et al. Frequency and pattern of heteroplasmy in the complete human mitochondrial genome. PLoS One 2013;8:e74636.
- [65] Ren T, Liang S, Zhao A, He K. Analysis of the complete mitochondrial genome of the Zhedong White goose and characterization of NUMTs: reveal domestication history of goose in China and Euro. Gene 2016;577:75–81.
- [66] Ricchetti M, Tekala F, Dujon B. Continued colonization of the human genome by mitochondrial DNA. PLoS Biol 2004;2:E273.
- [67] Richly E, Leister D. NUMTs in sequenced eukaryotic genomes. Mol Biol Evol 2004; 21:1081–4.
- [68] Riddle DL, Carbon J. Frameshift suppression: a nucleotide addition in the anticodon of a glycine transfer RNA. Nat New Biol 1973;242:230–4.
- [69] Rodin SN, Ohno S. 2 Types of aminoacyl-transfer-RNA synthetases could be originally encoded by complementary strands of the same nucleic-acid. Orig Life Evol Biosph 1995;23:393–418.
- [70] Rodin S, Rodin A, Ohno S. The presence of codon-anticodon pairs in the acceptor stem of tRNAs. Proc Natl Acad Sci U S A 1996;93:4537–42.
- [71] Rogers HH, Griffiths-Jones S. Mitochondrial pseudogenes in the nuclear genomes of Drosophila. PLoS One 2012;7:e32593.
- [72] Root-Bernstein M, Root-Bernstein R. The ribosome as a missing link in the evolution of life. J Theor Biol 2015;367:130–58.
- [73] Root-Bernstein M, Root-Bernstein R. The ribosome as a missing link in prebiotic evolution II: ribosomes encode ribosomal proteins that bind to common regions of their own mRNAs and rRNAs. J Theor Biol 2016;397:115–27.
- [74] Rose G, Passarino G, Scornaienchi V, Romeo G, Dato S, Bellizzi D, et al. The mitochondrial DNA control region shows genetically correlated levels of heteroplasmy in leukocytes of centenarians and their offspring. BMC Genomics 2007;8:293.
- [75] Rubio MA, Hopper AK. Transfer RNA travels from the cytoplasm to organelles. Wiley Interdiscip Rev RNA 2011;2:802–17.
- [76] Salinas T, Duby F, Larosa V, Coosemans N, Bonnefoy N, Motte P, et al. Co-evolution of mitochondrial tRNA import and codon usage determines translational efficiency in the green alga *Chlamydomonas*. PLoS Genet 2012;8:e1002946.
- [77] Schmitz J, Piskurek O, Zischler H. Forty million years of independent evolution: a mitochondrial gene and its corresponding nuclear pseudogene in primates. J Mol Evol 2005;61:1–11.
- [78] Schneider A. Mitochondrial tRNA import and its consequences for mitochondrial translation. Annu Rev Biochem 2011;80:1033–53.
- [79] Seligmann H. Hybridization between mitochondrial heavy strand tDNA and expressed light strand tRNA modulates the function of heavy strand tDNA as light strand replication origin. J Mol Biol 2008;379:188–99.
- [80] Seligmann H. Mitochondrial tRNAs as light strand replication origins: similarity between anticodon loops and the loop of the light strand replication origin predicts initiation of DNA replication. Biosystems 2010;99:85–93.
- [81] Seligmann H. Avoidance of antisense, antiterminator tRNA anticodons in vertebrate mitochondria. Biosystems 2010;101:42–50.
- [82] Seligmann H. Undetected antisense tRNAs in mitochondrial genomes? Biol Direct 2010;5:39.
- [83] Seligmann H. Pathogenic mutations in antisense mitochondrial tRNAs. J Theor Biol 2011;269:287–96.
- [84] Seligmann H. Two genetic codes, one genome: frameshifted primate mitochondrial genes code for additional proteins in presence of antisense tRNAs. Biosystems 2011;105:271–85.
- [85] Seligmann H. An overlapping genetic code for frameshifted overlapping genes in *Drosophila* mitochondria: antisense antitermination tRNAs UAR insert serine. J Theor Biol 2012;298:51–76.
- [86] Seligmann H. Overlapping genetic codes for overlapping frameshifted genes in Testudines, and *Lepidochelys olivacea* as special case. Comput Biol Chem 2012;41: 18–34.
- [87] Seligmann H. Coding constraints modulate chemically spontaneous mutational replication gradients in mitochondrial genomes. Curr Genomics 2012;13:37–54.
- [88] Seligmann H. Overlapping genes coded in the 3'-to-5'-direction in mitochondrial genes and 3'-to-5' polymerization of non-complementary RNA by an 'invertase'. J Theor Biol 2012;315:38-52.
- [89] Seligmann H. Putative mitochondrial polypeptides coded by expanded quadruplet codons decoded by antisense tRNAs with unusual anticodons. Biosystems 2012; 110:84–106.
- [90] Seligmann H. Pocketknife tRNA hypothesis: anticodons in mammal mitochondrial tRNA side-arm loops translate proteins? Biosystems 2013;113:165–76.
- [91] Seligmann H. Putative protein-encoding genes within mitochondrial rDNA and the D-Loop region. In: Lin Z, Liu W, editors. Ribosomes: molecular structure, role in biological functions and implications for genetic diseases; 2013. p. 67–86 [chapter 4].

- [92] Seligmann H. Triplex DNA:RNA, 3'-to-5' inverted RNA and protein coding in mitochondrial genomes. J Comput Biol 2013;20:660–71.
- [93] Seligmann H. Polymerization of non-complementary RNA: systematic symmetric nucleotide exchanges mainly involving uracil produce mitochondrial transcripts coding for cryptic overlapping genes. Biosystems 2013;111:156–74.
- [94] Seligmann H. Systematic asymmetric nucleotide exchanges produce human mitochondrial RNAs cryptically encoding for overlapping protein coding genes. J Theor Biol 2013;324:1–20.
- [95] Seligmann H. Putative anticodons in mitochondrial sidearm loops: Pocketknife tRNAs? J Theor Biol 2014;340:155–63.
- [96] Seligmann H. Mitochondrial swinger replication: DNA replication systematically exchanging nucleotides and short 16S ribosomal DNA swinger inserts. Biosystems 2014;25:22–31.
- [97] Seligmann H. Species radiation by DNA replication that systematically exchanges nucleotides? J Theor Biol 2014;363:216–22.
- [98] Seligmann H. Phylogeny of genetic codes and punctuation codes within genetic codes. Biosystems 2015;129:36–43.
- [99] Seligmann H. Sharp switches between regular and swinger mitochondrial replication: 16S rDNA systematically exchanging nucleotides  $A \leftrightarrow T + C \leftrightarrow G$  in the mitogenome of *Kamimuria wangi*. Mitochondrial DNA 2016;27:2440–6.
- [100] Seligmann H. Swinger RNAs with sharp switches between regular transcription and transcription systematically exchanging ribonucleotides: case studies. Biosystems 2015;135:1–8.
- [101] Seligmann H. Systematic exchanges between nucleotides: genomic swinger repeats and swinger transcription in human mitochondria. J Theor Biol 2015; 384:70–7.
- [102] Seligmann H. Codon expansion and systematic transcriptional deletions produce tetra-, pentacoded mitochondrial peptides. J Theor Biol 2015;387:154–65.
- [103] Seligmann H. Translation of mitochondrial swinger RNAs according to tri-, tetraand pentacodons. Biosystems 2016;140:38–48.
- [104] Seligmann H. Swinger RNA self-hybridization and mitochondrial non-canonical swinger transcription, transcription systematically exchanging nucleotides. J Theor Biol 2016;399:84–91.
- [105] Seligmann H. Systematically frameshifting by deletion of every 4th or 4th and 5th nucleotides during mitochondrial transcription: RNA self-hybridization regulates delRNA expression. Biosystems 2016;142:43–51.
- [106] Seligmann H, Krishnan NM. Mitochondrial replication origin stability and propensity of adjacent tRNA genes to form putative replication origins increase developmental stability in lizards. J Exp Zool B Mol Dev Evol 2006;306:433–49.
- [107] Seligmann H, Krishnan NM, Rao BJ. Possible multiple origins of replication in primate mitochondria: alternative role of tRNA sequences. J Theor Biol 2006;241: 321–32.
- [108] Seligmann H, Krishnan NM, Rao BJ. Mitochondrial tRNA sequences as unusual replication origins: pathogenic implications for *Homo sapiens*. J Theor Biol 2006; 243:375–85.
- [109] Seligmann H, Labra A. Tetracoding increases with body temperature in Lepidosauria. Biosystems 2013;447:155–63.
- [110] Seligmann H, Labra A. The relation between hairpin formation by mitochondrial WANCY tRNAs and the occurrence of the light strand replication origin in Lepidosauria. Gene 2014;542:248–57.
- [111] Se<sup>l</sup>igmann H, Raoult D. Unifying view of stem-loop hairpin RNA as origin of current and ancient parasitic and non-parasitic RNAs, including in giant viruses. Curr Opin Microbiol 2016;31:1–8.
- [112] Sharma A, Sharma A. Plasmodium falciparum mitochondria import tRNAs along with an active phenylalanyl-tRNA synthetase. Biochem J 2015;465:459–69.
- [113] Shi H, Dong J, Irwin DM, Zhang S, Mao X. Repetitive transpositions of mitochondrial DNA sequences to the nucleus during the radiation of horseshoe bats (*Rhinolophus*, Chiroptera). Gene 2016;581:161–9.
- [114] Sieber F, Duchêne AM, Maréchal-Drouard L. Mitochondrial RNA import: from diversity of natural mechanisms to potential applications. Int Rev Cell Mol Biol 2011;287:145–90.
- [115] Smigrodzki RM, Khan SM. Mitochondrial microheteroplasmy and a theory of aging and age-related disease. Rejuvenation Res 2005;8:172–98.
- [116] Song S, Jiang F, Yuan J, Guo W, Miao Y. Exceptionally high cumulative percentage of NUMTs originating from linear mitochondrial DNA molecules in the *Hydra magnipapillata* genome. BMC Genomics 2013;14:447.
- [117] Song H, Moulton MJ, Whiting MF. Rampant nuclear insertion of mtDNA across diverse lineages within Orthoptera (Insecta). PLoS One 2014;9:e110508.
- [118] Soto-Calderón ID, Clark NJ, Wildschutte JV, DiMattio K, Jensen-Seaman MI, Anthony NM. Identification of species-specific nuclear insertions of mitochondrial DNA (numts) in gorillas and their potential as population genetic markers. Mol Phylogenet Evol 2014;81:61–70.
- [119] Soto-Calderón ID, Lee EJ, Jensen-Seaman MI, Anthony NM. Factors affecting the relative abundance of nuclear copies of mitochondrial DNA (numts) in hominoids. J Mol Evol 2012;75:102–11.
- [120] Sroga GE, Nemoto F, Kuchino Y, Bjork GR. Insertion (sufB) in the anticodon loop or base substitution (sufC) in the anticodon stem of tRNA(Pro)2 from Salmonella typhimurium induces suppression of frameshift mutations. Nucleic Acids Res 1992;20:3463–9.
- [121] Stefano GB, Kream RM. Mitochondrial DNA heteroplasmy in human health and disease. Biomed Rep 2016;4:259–62.
- [122] Taylor EB, Rutter J. Mitochondrial quality control by the ubiquitin-prooteasome system. Biochem Soc Trans 2011;39:1509–13.
- [123] Thalman O, Hebler J, Poinar HN, Päabo S, Vigilant L. Unreliable mtDNA data due to nuclear insertions: a cautionary tale from analysis of humans and other great apes. Mol Ecol 2004;13:321–35.

- [124] Thalman O, Serre D, Hofreiter M, Lukas D, Eriksson J, Vigilant L. Nuclear insertions help and hinder inference of the evolutionary history of gorilla mtDNA. Mol Ecol 2005;14:179–88.
- [125] Tourmen Y, Baris O, Dessen P, Jacques C, Malthièry Y, Reynier P. Structure and chromosomal distribution of human mitochondrial pseudogenes. Genomics 2002;80: 71–7.
- [126] Tsuji J, Frith MC, Tomii K, Horton P. Mammalian NUMT insertion is non-random. Nucleic Acids Res 2012;40:9073–88.
- [127] Tuohy TM, Thompson S, Gesteland RF, Atkins JF. Seven, eight and nine-membered anticodon loop mutants of tRNA(2Arg) which cause + 1 frameshifting. Tolerance of DHU arm and other secondary mutations. J Mol Biol 1992;228:1042–54.
- [128] Verscheure S, Backeljau T, Desmyter S. In silico discovery of a nearly complete mitochondrial genome Numt in the dog (*Canis lupus familiaris*) nuclear genome. Genetica 2015;143:453–8.
- [129] Wang B, Zhou X, Shi F, Liu Z, Roos C, Garber PA, et al. Full-length Numt analysis provides evidence for hybridization between the Asian colobine genera *Trachypithecus* and *Semnopithecus*. Am J Primatol 2015;77:901–10.

- [130] Wang IX, Core LJ, Kwak H, Brady L, Bruzel A, McDaniel L, et al. RNA-DNA differences are generated in human cells within seconds after RNA exits polymerase II. Cell Rep 2014;6:906–15.
- [131] Xie B, Yang W, Chen L, Jiang H, Liao Y, Liao DJ. Two RNAs or DNAs may artificially fuse together at a short homologous sequence (SHS) during reverse transcription or polymerase chain reactions, and thus reporting an SHS-containing chimeric RNA requires extra caution. PLoS One 2016;11:e0154855.
- [132] Yang W, Wu J-M, Bi A-D, Ou-yang Y, Shen H-H, Chim G-W, et al. Possible formation of mitochondrial-RNA containing chimeric or trimeric RNA implies a posttranscriptional and post-splicing mechanism for RNA fusion. PLoS One 2013;8: e77016.
- [133] Yao YG, Kong QP, Salas A, Bandelt HJ. Pseudomitochondrial genome haunts disease studies. J Med Genet 2008;45:769–72.
- [134] Zhang DX, Hewitt GM. The long and short of nuclear mitochondrial DNA (Numt) lineages Reply from D-X. Zhang and G.M. Hewitt. Trends Ecol Evol 1997;12:114.