

We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900

Open access books available

186,000

International authors and editors

200M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



Directed Mutations Recode Mitochondrial Genes: From Regular to Stopless Genetic Codes

Hervé Seligmann

Additional information is available at the end of the chapter

<http://dx.doi.org/10.5772/intechopen.80871>

Abstract

Mitochondrial genetic codes evolve as side effects of stop codon ambiguity: suppressor tRNAs with anticodons translating stops transform genetic codes to stopless genetic codes. This produces peptides from frames other than regular ORFs, potentially increasing protein numbers coded by single sequences. Previous descriptions of marine turtle Olive Ridley mitogenomes imply directed stop-depletion of noncoding +1 gene frames, stop-creation recodes regular ORFs to stopless genetic codes. In this analysis, directed stop codon depletion in usually noncoding gene frames of the spiraling whitefly *Aleurodicus dispersus*' mitogenome produces new ORFs, introduces stops in regular ORFs, and apparently increases coding redundancy between different gene frames. Directed stop codon mutations switch between peptides coded by regular and stopless genetic codes. This process seems opposite to directed stop creation in HIV ORFs within genomes of immunized elite HIV controllers. Unknown DNA replication/editing mechanisms probably direct stop creation/depletion beyond natural selection on stops. Switches between genetic codes regulate translation of different gene frames.

Keywords: codon-amino acid reassignment, *Lepidochelys olivacea*, antitermination, ribosomal RNA, directional evolution, APOBEC

1. Introduction

Mitochondrial genetic code diversification frequently reassigns stop codons to amino acids [1]. This situation probably reflects ambiguity in roles of stop codons during translation in mitochondria. Indeed, phylogenetic reconstructions of the evolution of genetic codes based on differences between codons in amino acid and stop codon assignments resemble known phylogenies of organisms using these genetic codes, if ambiguity of stop codons in mitochondrial (but not nuclear) genetic codes is accounted [2].

These observations suggest translational activity of mitochondrial tRNAs with anticodons matching stop codons [3] templated by antisense strands of regular mitochondrial tRNAs [4, 5]. Predicted occurrences of mitochondrial stop suppressor (antiterminator) tRNAs coevolve with stop codon usages in predicted off-frame protein coding genes in all analyzed mitochondria (primates and *Drosophila* [6, 7], including their ribosomal RNAs [8], turtles [9] and *Chaetognaths* [10]), and in a peptide detected by specific monoclonal immunolocalization in human mitochondria [11]. Hence, there are probably more mitochondrion-encoded genes than usually accepted [12, 13]. Mitogenome size reduction probably causes gene multifunctionality, including mt tDNAs functioning as replication origins [14–21].

Translation of stop codons occurs in different organisms [22, 23], but seems particularly widespread in mitochondria [2, 3] as reflected by the evolution of mitochondrial genetic codes [1]. Nuclear genetic codes lacking dedicated stop codons have been described in protists [24–27] and fungi [28].

1.1. Alternative coding by expanded codons

Switching between regular and stopless genetic codes is not the only strategy increasing dramatically information encoded by genomes. Isolated tetracodons, codons expanded by a fourth silent nucleotide, are known since the dawn of molecular biology [29]. These are sometimes translated by tRNAs with expanded anticodons [30, 31]. It seems probable that systematic frameshifts produce stretches of tetracodons that code for (yet) undetected peptides ([32–34]). Some evidence suggests that tetracoding occurs especially at high environmental temperatures [35] and is predicted by genetic code optimization [36]: codon-anticodon interactions are more stable when four rather than only three base pairs hybridize. Other theoretical considerations suggest that the mitochondrial vertebrate genetic code evolved from a specific subset of 64 tetracodons, the tesserae, chosen on the basis of symmetry principles [37].

Indeed, analyses of mitochondrial mass spectra searching for peptides matching translations assuming tetra- and pentacodons, codons expanded by one or two silent nucleotides, detected numerous tetra- and pentacoded peptides [38–43].

1.2. Alternative coding by swinger polymerization

Further little known mechanisms increase numbers of proteins potentially coded by single sequences. Polymerization occasionally exchanges systematically between nucleotides during DNA replication [44–46] or RNA transcription [47–53] for long sequence stretches (23 exchange rules are possible, nine symmetric, e.g., $A \diamond C$, and fourteen asymmetric, e.g., $A > C > G > A$), producing swinger sequences. Swinger replication, in particular the double symmetric exchange $A \diamond T + C \diamond G$, seems most frequent for mitochondrial ribosomal RNAs [46]. This increases the coding potential of rRNAs, strengthening the hypothesis that rRNAs are modern remnants of protogenomes that templated for translational molecules (tRNA-like and rRNA-like) and protein coding genes [54–58] by dense overlap coding. This is compatible with the occurrence of protein coding regions within modern rRNAs [8, 59–61].

I stress here that the exchange $A \diamond T + C \diamond G$ is not trivial: this creates the complement of the template sequence, which is not the regular inverse (or reverse) complement. “Complement”

is frequently used as shortcut for inverse complement, but the A \diamond T + C \diamond G transformed sequence is a different sequence because it lacks the 3'-to-5' inversion combined with nucleotide complementarity. The shortcut has been used because 3'-to-5' inverted sequences had not been described previous to the description of A \diamond T + C \diamond G transformed sequences.

It seems regular transcription occasionally switches abruptly to swinger transcription (and vice versa), as indicated by chimeric RNAs. These RNAs correspond in part to regular DNA, and an adjacent part corresponds to DNA only if accounting for swinger transformation [62]. Corresponding chimeric peptides have also been detected [63]. Chimeric DNA also exists: the mitogenome of the stonefly *Kamimuria wangi* (NC_024033) is regular, beside its 16S rRNA, which is entirely swinger transformed along transformation A \diamond T + C \diamond G [46].

1.3. Regulation of alternative transcriptions by secondary structure

Secondary structures are important components of RNA function and evolution [64, 65]. Secondary structures formed by tRNA sequences punctuate posttranscriptional processing of mitochondrial transcripts [66–68]. Palindromes potentially forming secondary structures after sequence swinger-transformation associate with detected mitochondrial swinger RNAs [69]. This is similar to what is known from regular RNA processing in mitochondria and, surprisingly, giant viruses [70] which bear also other striking resemblances with mitogenomes, including similar gene order [71]. Transcription sometimes deletes systematically mono- or dinucleotides after transcribing trinucleotides (del-transcription), translated into peptides that in part converge with peptides translated from regular RNAs by expanded anticodons [38]. Del-transcription, or at least detection of delRNAs, seems downregulated by secondary structures formed after transformation of the sequence by systematic deletions [72].

1.4. Mechanisms that switch between genetic codes?

Phenomena systematically exchanging nucleotides remind more specific mechanisms that alter the genetic code according to which a protein is coded, from a regular genetic code to a stopless genetic code, and/or vice versa. Indeed, previous analyses of mitogenomes revealed that in one GenBank mitogenome (from a marine turtle, the Olive Ridley, *Lepidochelys olivacea*), following GenBank's annotations, several regular protein coding genes do not code for the regular proteins essential for mitochondrial metabolism and usually encoded at these genomic locations. These essential proteins are indeed coded by the corresponding sequences, but only after a frameshift, and only if stops in that frame are translated, explaining the erroneous annotation of stopless ORFs (abbreviation for open reading frames) that do not code for recognized mitochondrial proteins [9].

This was originally interpreted as resulting from directed selection on stop codons. Observations of systematic mutations in contexts creating stop codons in ORFs of HIV genes specifically in elite controllers immune to HIV [73, 74] suggests that enzymatically directed mutagenesis during DNA replication and/or edition could transform ORFs coded according to a regular genetic code into one coded by a stopless genetic code, and vice versa, as observed in several mitochondrial genes of the Olive Ridley. For HIV, introducing stops presumably drastically reduces viral production and contributes to immunity.

This hypothesis of mutations directed at stop codons is in line with observations that polymerase errors are more frequent in stop codon contexts, interpreted as an adaptational bias to introduce mutations in stops [75]. In the next section, GenBank is explored to detect further mitogenomes in which genetic codes were switched by producing stop codons in ORFs and stop-depletion in other frames.

2. Methods

2.1. Exploring GenBank for genetic code switches

Previous Blastp searches found proteins already described in GenBank and aligning with hypothetical peptides translated from randomly chosen frameshifted vertebrate mitochondrial genes. These analyses detected the unusual proteins translated from ORFs of the mitogenome of *Lepidochelys olivacea* [9]. In these cases, the regular mitochondrial proteins are coded in frames that include stops, and hence were not recognized as the regular gene. The annotated frame is stopless, but codes for other, unknown peptides. These other peptides are homologous to peptides translated after frameshift from regular mitochondrial protein coding genes, from other mitogenomes that did not undergo stop codon depletion in non-ORF frames.

2.2. Choice of seed sequences for BLAST searches

The method described above only detects homologies for sequences sufficiently similar to “seed” sequences used for BLAST analyses of GenBank. Therefore, using as seed the human mitogenome, mainly vertebrate proteins were detected, as for the above-mentioned *Lepidochelys olivacea*. A similar situation occurs for detection of swinger DNA/RNA sequences: the original searches using as seed swinger transformed versions of the human mitogenome only detected vertebrate sequences [45], but BLAST analyses using a randomly chosen invertebrate mitogenome (from the North Pacific krill *Euphausia pacifica* (NC_016184)) detected numerous additional swinger sequences, from insect mitogenomes [38].

This search principle for insect nucleotide sequences can also be applied for proteins. I use as seed the five peptides translated from the five “noncoding” frames of the 13 regular protein coding genes of *Euphausia*'s previously randomly chosen invertebrate mitogenome. These 65 peptides were blasted to search GenBank for proteins already described and with high homologies with peptides translated from *Euphausia*'s noncoding frames.

3. Results

3.1. Preliminary results from *Aleurodicus dispersus*

Preliminary BLAST analyses of peptides translated from noncoding frames of *Euphausia*'s mitochondrial protein coding genes detected GenBank proteins from the mitogenome of *Aleurodicus dispersus*, a sap-sucking spiraling whitefly. These have high homology levels with peptides translated from the antisense sequence of several among *Euphausia*'s protein

coding genes. These unusual CDs in this insect mitogenome remind previous descriptions of other unusual CDs in the mitogenome of the marine turtle *Lepidochelys olivacea*. This justifies detailed analyses of peptides translated from the six frames of the 13 protein coding genes of the mitogenomes of *Aleurodicus dispersus* (JX566506), and, for comparative purpose, of its closest relative with a complete mitogenome in GenBank, *Aleurodicus dugesii* (NC_005939), whose predicted proteome seems coded according to regular rules.

3.2. *Aleurodicus dispersus* protein coding genes

All six frames of the 13 mitochondrial protein coding genes of *Aleurodicus dispersus* were translated according to the regular invertebrate mitochondrial genetic code. First, BLAST analyzed peptides translated from GenBank-annotated, stopless ORFs to verify which of these peptides are “normal,” i.e., have regular homologies with the corresponding protein predicted for the regular ORF of the mitogenome of *Aleurodicus dugesii*.

These analyses confirm that GenBank annotations of the six *Aleurodicus dispersus* mitogenes COI, COII, AT6, COIII, ND3, and ND2 code for typical invertebrate proteins homologous with corresponding proteins in regular insect mitogenomes, notably *Aleurodicus dugesii*. The remaining seven genes follow different coding structures described below, based on frame-shifts and/or stop depletion/translation. Blastp does not detect any homologies for proteins predicted according to GenBank annotations for genes AT8, ND1, ND6, ND5, ND4, and ND4L, and only partial homology for CytB.

Mitochondrial metabolism without the regular proteins usually translated from these seven genes seems impossible. Regular analyses of the mitogenome of *Aleurodicus dugesii* detect these proteins as they are annotated in GenBank. The possibility that these genes were transferred in *Aleurodicus dispersus* to the nucleus and that proteins are imported to the mitochondrion seems unlikely as ORFs occur at positions corresponding to gene locations coding for the seven missing proteins in the predicted mitoproteome of *Aleurodicus dispersus*.

3.3. Recoding of mitogenes in *Aleurodicus dispersus*

3.3.1. Two ORFs on the same strand: AT8

The case of the missing ATP synthetase subunit 8 is solved by Blastp analysis of the peptide coded by the +2 frameshifted sequence of gene AT8. Residues at positions 8–48 (the gene has 49 codons including stop codon) in frame +2 have 75% similarity with congeneric mitochondrial ATP synthetase subunit 8 of *Aleurodicus dugesii* (YP_026055, e value 2×10^{-9} , see alignment in **Figure 1**). It seems likely that the regular AT8 gene codes for the corresponding protein. This frame does not contain stops, implying that this gene has two stopless ORFs. The GenBank-annotated ORF does not correspond to the regular AT8, which is the +2 frame-shifted sequence of the annotated sequence.

3.3.2. Stop codon translation after frameshift: ND6

For gene ND6, the stopless ORF annotated in GenBank does not align with any ND6-like protein. This conundrum is solved by Blastp analysis of the peptide translated from the +1

frameshifted sequence of ND6 as it is annotated in GenBank. It aligns with 86% similarity with the mitochondrial NADH dehydrogenase subunit 6 of congeneric *Aleurodicus dugesii* (positions 33–137 in *Aleurodicus dispersus* and 32–129 in *Aleurodicus dugesii*, e value 1×10^{-37} , not shown). Hence, the annotated gene corresponds to a stopless frame that does not translate into a recognizable mitochondrial protein, while the +1 frame, which contains three stop codons codes for ND6. Only one of the stop codons is within the alignment, where it corresponds to a tyrosine in *Aleurodicus dugesii*. This implies translation of at least one stop codon, as previously described for other short mitochondrial protein coding genes where the protein coding region includes a programmed frameshift and translation of stops (ND3 in birds [76] and in turtles [77]). It seems plausible that ND6 translation starts in the 5' region of the frame as annotated in GenBank, and then a programmed frameshift occurs in the vicinity of the 5'-starting point of the alignment. Translation of the stop codon by tyrosine is compatible with translation by tRNAs with near-cognate anticodons [78–80].

3.3.3. Frameshift with stop translation: CytB

The situation in CytB is similar and reminds again known cases of proteins coded by two frames. The ORF as annotated in GenBank has high homology (96% similarity) from residue 137 to 355 with the regular cytochrome B of *Aleurodicus dugesii* (YP_026063, e value 5×10^{-91} , **Figure 2**). The 5' extremity of cytochrome B is coded by the +1 frameshifted sequence of the gene, as indicated by high similarity (88%) in the alignment from residues 6–136 with the regular cytochrome B of *Aleurodicus dugesii* (YP_026063, e value 5×10^{-54} , **Figure 2**). Position 131 is a stop that corresponds to a tyrosine in *Aleurodicus dugesii*. Hence, this gene's coding structure implies frameshift and probable stop translation, potentially by near cognate anticodon.

JX566506	8	<u>MVIFV</u> FFFFFFFFFSLSFNFYYYYKFYTFKNKKIYFD <u>YYKIKW</u>	48
YP_026055	8	MWLVLFFFFFFFFFSLSLNFYYYYKSYNFKMSKSNINYYKMKW	48

Figure 1. Alignment between peptide translated from the +1 frameshifted annotated GenBank sequence of gene AT8 in JX566506 with the protein translated from sequence AGA54141. Bold indicates similar residues, and underlined letters are for identical residues.

Frame +1			
JX566506	6	<u>KNLFLFTNLNGFIIDLGVPSYLYLWNNFSGLLGLVLVIQFLTGLFLTFHYASIVRVAFDRVIHIMRDVNWGWLLRYMHIDGASFFFIFIFIYIHIGR</u>	100
YP_026063	6	KNLFLFKNLNDFIIDLGVPTNLYLWNNFSGLLGLIFMIQFMTGLFLTFHYASNINVAFDSVIHIMRDVNSGWLLRYMHINGASFFFIFIFIYIHIGR	100
JX566506	101	<u>GLYYFSFTKKVWIRGEI</u> LLLLLMKIAFMG* <u>IFPWG</u>	136
YP_026063	101	GLYYFSFKKLVWISGVMILLMMIAFMGYILPWG	136
Frame 0 according to GenBank annotation of JX566506			
JX566506	137	AKNHFEQPTVITNLLSAIPLSGNLIWNWIWGGFSVGNATLNRFYSFHFFFPFILIFLIFHLLFLHIDGSSNSLGLNNTYDKIKFYPPYLLKDYM	231
YP_026063	136	GQMSFWGATVITNLLSAIPIWGDIVSWIWGGFSVGNATLNRFYSFHFFFPFILIFLIFHLLFLHIDGSSNSLGLNNTYDKIKFYPPYLLKDYM	230
JX566506	232	<u>GFMFFFFFFFIIIFNPLILSDSENFI</u> MANSPLITPIHIQPEWYLLFAYAILRSITSKLGGVMALFFSILILLIMLFLKSKFNGLMFYPILKIMFF	326
YP_026063	231	<u>GFMFFFFFFFIIIFNPLILSDSENFI</u> MANSPLITPIHIQPEWYLLFAYAILRSITSKLGGVMALFFSILILLIMLFLKSKFNGLMFYPILKIMFF	325
JX566506	327	<u>FFFFIVLILTLWLGSKQVEYPYMLGSLMT</u>	355
YP_026063	326	FFFFIVLILTLWLGSKQVEYPYMLGQLMT	354

Figure 2. Alignment between peptides translated from frame 0 and the +1 frameshifted annotated GenBank sequence of gene CytB in JX566506 with *Aleurodicus dugesii* cytochrome B, sequence YP_026063. Bold indicates similar residues, and underlined letters are for identical residues. Frame +1 of sequence annotated in GenBank JX566506.

3.3.4. Stop codon depletion in antisense strand: ND1

The annotation in GenBank for ND1 does not correspond to a protein homologous to NADH dehydrogenase subunit 1. However, the peptide translated from the +1 frame of the opposite (antisense) strand has high homology with NADH dehydrogenase subunit 1 from *Aleurodicus dugesii* (YP_026064, 94% similarity for the complete length, e value 0). This implies regular encoding of that protein. The misannotation originates from depletion of all stops in one frame of the antisense strand of that gene. The corresponding antisense frame has 15 TAR stop codons in the regular ND1 of *Aleurodicus dugesii*. **Figure 3** aligns the peptide translated from this presumably noncoding antisense frame in *Aleurodicus dugesii* with the peptide translated from the GenBank-annotated frame. Stop codons correspond in this alignment mainly to serine (seven cases), then to tryptophan (two cases) and once each to leucine, lysine, methionine, and asparagine. This predicted translation is to much lower extents compatible with near cognate translation, and might be due to specific tRNA(s) with anticodon(s) matching stops.

The fact that this antisense frame is stop codon depleted in *Aleurodicus dispersus* so that it does not necessitate any special translational machinery for its expression suggests the possibility that this frame is translated in *Aleurodicus dugesii* (and in other species) and produces an unknown functional protein, and this is due to stop codon translation by antiterminator tRNAs. Indeed, the entirety of both mitochondrial strands is transcribed to RNA; hence, RNA corresponding to this supposedly noncoding strand necessarily exists and could be translated [81]. The alignment suggests that the amino acid most probably inserted by that stop-suppressor tRNA is serine. This is in line with previous observations from cytochrome c oxidase subunit I from the silkworm *Samia ricini* [7, 9], where stop codons in a usually noncoding frame systematically mutated to serine. This finding strengthens serine as the likely residue inserted at stop codons in insect mitochondria. This awaits confirmation by translation and tRNA aminoacylation experiments (as for example done for giant virus tRNAs [82]).

Dug	2	IFK*SLVLSFFLV*YNLNDKLLKKYSIVNGKNLIQVKYINLS*RNRGSVHRVQMNKKHKI	61
Dis	4	IFKSSLVLSFFLVWYNLNDKLLKKYSIVNGKNLIQVKYINLSSRNRGSVHRVQMNKKHKI	63
Dug	62	MNVKKKIIDLIFMLKNIVNVINLMIKMLEYSAKMKAKENLLYSMLKPEINSDSPSEKSN	121
Dis	64	MNVKKKIIDLIFMLKNIVNVINLMIKMLEYSAKMKAKENLLYSMMKPEINSDSPSEKSK	123
Dug	122	GVRLVSAKIVMIQM*NKNIKYK*NIMYFCMMNM*LKFNCVSVKLTIVMMMI IKLISYEIIW	181
Dis	124	GVRLVSAKMVMIQMSNMKIKSMWNDKFFCIKNMLLKFICSVKLMIMMMI IKLISYEIIW	183
Dug	182	AMDRISPPIA*LEFEDHLINMMIMLIDIIILINIKMLNLLKLFQ*NG*IIHIIKEQMMF	241
Dis	184	AMERINPIMASLELEDHLIIMMMIMFMDIILMTIKMLNLLKLFQMNGSIIHIINEQMMF	243
Dug	242	NNGNM**MM*FDLKFKFISLENNLMASLKGWSIPMINVLLGPLRFCM*LKILRSIKVKNA	301
Dis	244	NKGNKSSMMKLDLKFEMFISLVNNLMASLNGCKIPKIMVLLGPLRFCMNLKILRSINVKNA	303
Dug	302	ILKKVLINNKIMIKK	316
Dis	304	ILKKMDNNKMIMFNK	318

Figure 3. Alignment between the peptide translated from the +1 frameshifted antisense sequence of ND1 in *Aleurodicus dugesii* (YP_026063) and the peptide translated from the ORF annotated in GenBank for ND1 in *Aleurodicus dispersus* (JX566506). Underlined asterisks indicate stops in the antisense sequence of *Aleurodicus dugesii*, which frequently correspond to serine in the stop-depleted ORF of *Aleurodicus dispersus*.

3.3.5. Stop codon depletion in antisense strand and stop codon translation: ND4l and ND5

Annotations in GenBank for *Aleurodicus dispersus*' ND4l and ND5 do not match proteins homologous to the corresponding NADH dehydrogenase subunits. For ND4l, the peptide translated from frame 0 of the antisense of that gene (this frame includes five stops) yields a short alignment with the regular protein in *Aleurodicus dugesii* (24 residues, 87% similarity, e value 8.5, not shown). Hence, the annotated ORF in *Aleurodicus dispersus* is probably a stop-codon-depleted antisense sequence (the corresponding frame in *Aleurodicus dugesii* has four stops), which does not code for the regular ND4l gene. This stop codon depletion in *Aleurodicus dispersus* introduced five stops in the sense strand frame that apparently codes for ND4l according to the above-described alignment.

For ND5, the peptide translated from the +1 frame of the antisense of the GenBank-annotated sequence is homologous over its complete length to NADH dehydrogenase subunit 5 of *Aleurodicus dugesii* (89% similarity, e value 0, not shown). This frame has a single stop codon that aligns with serine in *Aleurodicus dugesii* (see discussion of insertion of serine at stops in previous section).

3.3.6. Stop codon depletion in antisense strand, frameshift, and stop translation: ND4

A further mitochondrial gene for which the GenBank annotation does not produce the expected protein for *Aleurodicus dispersus* is ND4. Alignment analyses detect peptides homologous with regular NADH dehydrogenase subunit 4 when translating frames +1 and +2 of the antisense of the GenBank-annotated ND4 gene. Blastp alignment analyses detect homology with the regular ND4-encoded protein of *Aleurodicus dugesii*; the regular protein is encoded in antisense frame +1 until residue 297 (89% similarity, e value 3×10^{-126} , not shown). This alignment includes a single stop, matching tyrosine in *Aleurodicus dugesii*. Part of the remaining protein is coded by a stopless stretch of antisense strand frame +2, where residues 341–427 align with the regular protein from *Aleurodicus dugesii* (73% similarity, e value 2×10^{-21}). Hence, the annotated ORF in *Aleurodicus dispersus* is a stop-codon-depleted antisense frame that codes for an apparently different protein, the actual NADH dehydrogenase subunit 4 is encoded by two frames, one containing a stop, on the opposite strand.

4. General discussion

4.1. Genomic stop codon-depletion: overall analysis

Table 1 presents numbers of stops in all six frames of the mitochondrial genes of *Aleurodicus dugesii*, and of *Aleurodicus dispersus*, according to the strand presented in GenBank. The frame coding for the regular protein in *Aleurodicus dugesii* is always the only stopless frame, the frame(s) coding for the regular proteins are underlined in *Aleurodicus dispersus*, these are not necessarily stopless, and are not necessarily the only stopless frame.

In order to account for slight variations in gene sizes, I compare between percentages of stop codons averaged across all six frames in the two species, gene by gene. Mean stop

	<i>Aleurodicus dugesii</i>							<i>Aleurodicus dispersus</i>						
	N	F0	F1	F2	R0	R1	R2	N	F0	F1	F2	R0	R1	R2
AT6	217	<u>0</u>	40	12	39	15	11	218	<u>0</u>	34	13	38	18	11
AT8	48	<u>0</u>	6	6	4	3	6	51	0	7	<u>0</u>	4	7	2
COI	515	<u>0</u>	65	31	64	61	11	518	<u>0</u>	59	35	68	60	13
COII	222	<u>0</u>	33	15	24	25	3	222	<u>0</u>	29	15	21	21	5
COIII	261	<u>0</u>	34	11	37	14	13	266	<u>0</u>	25	16	33	22	16
CytB	378	<u>0</u>	39	21	50	47	12	380	<u>0</u>	<u>27</u>	24	37	53	20
ND1	320	39	15	45	<u>0</u>	31	47	320	28	<u>0</u>	48	0	38	56
ND2	320	<u>0</u>	59	20	42	37	22	318	<u>0</u>	43	14	35	41	15
ND3	121	<u>0</u>	19	4	20	16	1	121	<u>0</u>	16	4	18	15	1
ND4I	101	8	3	22	<u>0</u>	7	16	100	<u>5</u>	9	11	0	14	16
ND4	426	44	8	83	<u>0</u>	42	60	434	39	<u>19</u>	<u>50</u>	0	43	64
ND5	559	49	16	86	<u>0</u>	53	102	558	64	<u>0</u>	99	0	41	100
ND6	140	<u>0</u>	17	10	25	16	17	142	11	0	<u>3</u>	14	20	16
12S	245	21	27	20	36	22	21	249	28	27	32	20	22	28
16S	396	34	35	51	42	56	50	396	41	56	51	50	36	31

Table 1. TAR stop codons in the six frames of mitogenome-encoded genes of *Aleurodicus dugesii* and *Aleurodicus dispersus*.

codon percentages decrease in 11 among 13 protein coding genes of *Aleurodicus dispersus*, as compared to *Aleurodicus dugesii*. This is a significant majority of cases according to a one tailed sign test using a binomial distribution and assuming equal probability of getting more or less stop codons in any of these mitogenes ($P = 0.00562$). This overall stop codon depletion occurs in all seven “recoded” genes. Stop codon depletion occurs qualitatively in four among the six genes with regular, unchanged coding structure. This tendency is not statistically significant for this subgroup of genes when using the robust, but blunt nonparametric sign test. A paired t test between mean percentages of stop codons averaged across frames indicates also for these six genes a statistically significant decrease in stops in *Aleurodicus dispersus*, as compared to *Aleurodicus dugesii*. This result suggests that stop codon depletion occurred across all or at least most of this genome, and for most frames, not only for genes whose coding structure was altered, and not only for frames who became ORFs.

Presumably, unknown mechanisms associated with replication depleted stop codons in this species’ mitogenome, perhaps cumulatively over several replication or DNA edition cycles. Total stop codon depletion in some frames produced new ORFs. Natural selection against stop codons presumably enhanced unknown enzymatic phenomena, eliminating stop codons in these frames. It seems plausible that these frames in usual mitogenomes code for proteins translated by stop suppressor tRNAs. Specific unknown conditions in *Aleurodicus dispersus* may favor enhanced expression of peptides coded by frames that usually include stops in other mitogenomes, such as *Aleurodicus dugesii*. These constraints would have ultimately caused genomic stop codon depletion in *Aleurodicus dispersus*. In regular mitogenomes, stop codon translation downregulates expression of these unusual peptides in favor of proteins coded by regular ORFs, but in *Aleurodicus dispersus*, this hierarchy may be inexistent (when two stopless ORFs occur in a gene) or reversed (as in several mitogenes of *Lepidochelys olivacea* [9]), with translation of the unusual peptide not necessitating stop codon suppression, and translation of regular mitochondrial proteins requiring tRNAs with anticodons matching stop codons.

4.2. Coding redundancy between frames and tolerating ribosomal frameshifts

The original hypothesis of frame shiftability suggests that different frames of a gene code for somewhat similar peptides, presumably because the genetic code is optimized to tolerate frameshifts [83–85]. This hypothesis suggests that redundancy among frames in *Aleurodicus dispersus* should be greater than in the closely related *Aleurodicus dugesii* where coding seems regular, assuming that changes in coding structure increase redundancy among frames for coding protein variants with similar functions.

I used ClustalX to align the regular protein with peptides coded by the +1 and +2 frames of the same coding strand, for each *Aleurodicus dispersus* and *Aleurodicus dugesii*. Numbers of amino acids that were identical in the alignment were divided by total peptide lengths. This proportion for genes from *Aleurodicus dispersus* is plotted as a function of the corresponding proportion for *Aleurodicus dugesii* for alignments between frame 0 and frame +1 (**Figure 4**). Redundancy between frame 0 and +1 is greater in mitogenes of *Aleurodicus dispersus* in nine among twelve genes (there was no difference between these species for gene COII), a statistically significant majority according to a one tailed sign test ($P = 0.0365$). This tendency however does not exist for alignments between frames 0 and +2 (redundancy in *Aleurodicus dispersus* greater in 6 among thirteen genes).

This analysis tentatively indicates that stop codon depletion and coding by frameshifting and translation of stop codons might associate with a phenomenon increasing tolerance to

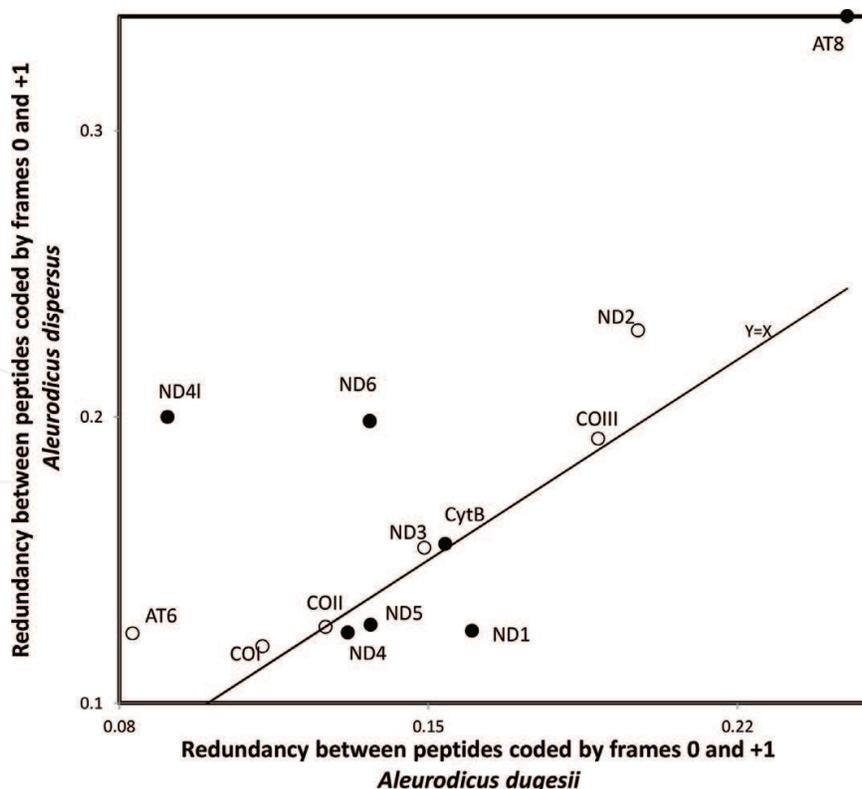


Figure 4. Redundancy between peptides coded by frames 0 and +1 for mitochondrial genes of *Aleurodicus dispersus* as a function of the corresponding redundancy for *Aleurodicus dugesii*. Filled symbols are for genes with unusual coding structure (frameshifts, stop depletion creating new ORFs, stop codon translation).

frameshifts during translation. Indeed, frequencies of off frame stop codons in mitochondrial genes are inversely correlated to predicted ribosomal RNA stability [86–88], suggesting that genes adapt to avoid negative effects of ribosomal frameshifts [44, 89, 90]. Stop codon-depletion may enable coding for more proteins, in addition to increasing redundancies between frames. Several effects could explain that results are not very strong statistically at the level of redundancy between frames. This hypothesis should be further tested, experimentally as done by Wang et al. [83–85] and by other bioinformatics analyses. For example, one can expect that frameshift tolerance biases exist for identity at amino acids that are not easily replaced by other amino acids (e.g., cysteine), but less for mutable ones (leucine, isoleucine, etc.). The preliminary tests presented here are not incompatible with the frameshift tolerance hypothesis [83–85].

It is important to understand in this context that the genetic code's discovery, among the greatest fundamental discoveries, is not over, but only in process. Indeed, coding sequences include much more information than generally believed, even beyond RNA editing (RDD [91]), systematic transformations during replication [44–46] and transcription [39, 47–51], and translation along expanded codons [32–43]. Cryptic codes [92, 93] such as the well-developed theory of the natural circular code [94–112] regulate the ribosomal translation frame [113–116], and protein cotranslational folding [117], remain to be described and decoded.

4.3. Sequencing artifacts and genome annotation

During the redescription of the recoded *Lepidochelys olivacea* mitogenome [9], an anonymous reviewer suggested that sequencing errors mimicked frameshifting mutations (insertion/deletion), producing the impression of frame recoding. This explanation is incompatible with the phenomena described in *Lepidochelys* and *Aleurodicus*, because these involve numerous specific changes/mutations in stop codon-specific nucleotide contexts, totally depleting stop codons in usually noncoding frames, and introducing stop codons in usually stopless, regular ORFs. Frameshifting mutations insert/delete a nucleotide within a regular ORF, which due to the frameshifting mutation is split between two frames. This does not deplete stop codons occurring in noncoding frames, nor introduce stops in the frameshifted ORF. ORF creation in usually noncoding frames by stop codon depletion in *Lepidochelys olivacea* [9] and *Aleurodicus dispersus* probably originates from natural, enzymatic, directed mutations [118] or other processes causing directed mutations, such as transposon-mediated directed mutations [119, 120].

Recoding probably occurs beyond mitogenomes. However, the short highly conserved mitogenomes [121] are most adequate to manual reannotation, a first necessary stage to detect events where genes are recoded from one to another genetic code. I suggest that annotations of genomes, and mitogenomes in particular, take systematically into account phenomena such as swinger sequences [51], and directed stop codon depletions that may result in ORFs that do not code for regular recognized proteins as presented here, especially in genomes/genes that seem unusual and remain in an unverified status in GenBank. **Figure 5** resumes the changes in coding structures that occurred in *Aleurodicus dispersus* due to recoding, as compared to the “ancestral” regular situation in *A. dugesii*. The proposed stopless genetic code in *Aleurodicus dispersus* presumably introduces serine at stops TAR and differs from previously described alternative arthropod mitochondrial genetic codes, which usually recode codons AGR [122, 123].

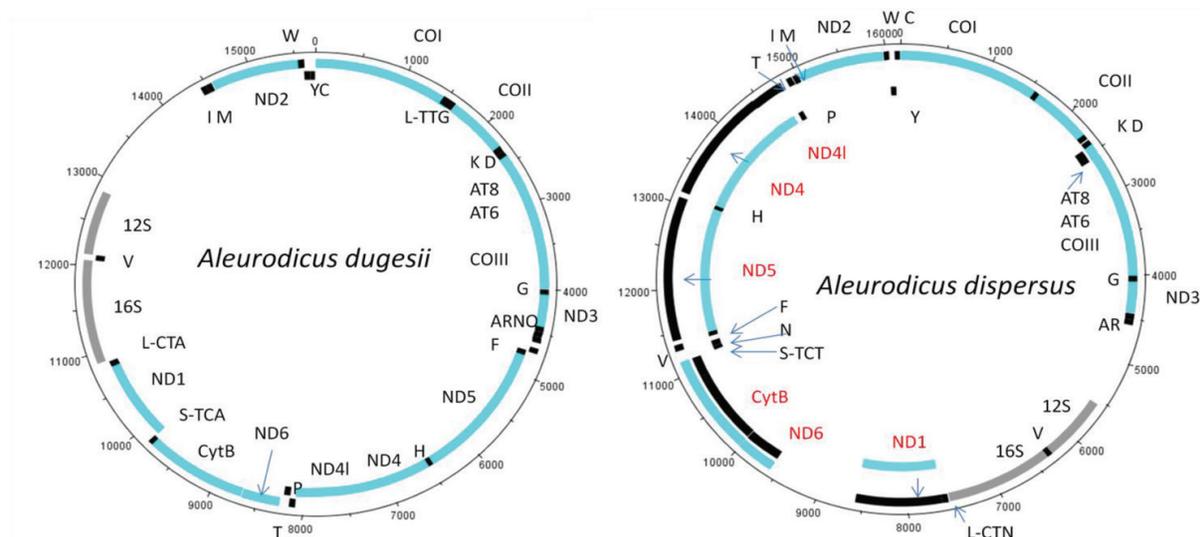


Figure 5. Classical and unusual mitogenome structures of whiteflies (*Aleurodicus dugesii*; *A. dispersus*). In *A. dispersus*, GenBank annotates genes ND1, ND4l, ND4, ND5, ND6 and CytB erroneously stopless frames coding for other proteins. A different frame with stops codes for the metabolically usual protein. CytB: both frames on same strand; other genes: opposite strands.

5. Conclusions

In vertebrate mitochondria, BLAST analyses of peptides translated from frames that are not recognized ORFs and contain stop codons align with high homology with proteins translated from regular mitochondrial ORFs in GenBank. Many such ORFs code for peptides matching usually noncoding sequences and occur in the mitogenome of *Lepidochelys olivacea* [9]. In this case, similar analyses are done for invertebrate mitochondria and a mitochondrial genome (JX566506, from *Aleurodicus dispersus*, Yu and Du, submitted in GenBank 2012, unpublished) considered as unverified are detected, probably because many protein-coding genes are undetectable with usual coding rules.

Several phenomena, and their combination, explain this situation. Alignment analyses detect the coding rules for these genes: frameshift, translation of stop codons, and depletion of stop codons in usually noncoding frames. Previous analyses detected in *Lepidochelys olivacea* CytB two stop-codon-deprived frames on the sense strand, among which one codes for the regular cytochrome B, and the other for an unknown protein. GenBank annotates erroneously the latter frame as coding for cytochrome B. A similar observation is reported here for the gene AT8 in *Aleurodicus dispersus*.

Some mitochondrial protein coding genes in *Aleurodicus dispersus* are unusual in the sense that the stop-codon-depleted frame erroneously annotated as the regular mitochondrial protein coding gene is on the strand opposite to the sense strand coding (with or without stop codons) for the actual usual protein: ND1, ND4l, ND4, ND5. The process depleting stop codons in these antisense frames is unknown and of particular interest. It is a probable combination of natural selection and enzymatically directed mutations to and from stop codons in the adequate nucleotide contexts, perhaps promoted by unknown conditions specific to *Aleurodicus dispersus*. In some genes, the protein usually coded by the regular genetic code

necessitates translating stop codons (a stopless genetic code), while frames including stop codons and therefore not considered as ORFs become stop codon depleted, and hence corresponding peptides are coded by the regular invertebrate mitochondrial genetic code. This situation where peptides coded by regular and stopless genetic codes are swapped might reflect a reversal in hierarchies of needs for the expressions of the respective peptides, specific to *Aleurodicus dispersus*. The requirement for tRNAs translating stop codons would regulate these respective expressions, *de facto* swapping between regular and stopless genetic codes. I suggest that the enzymatically directed stop codon depletion is related to the process that caused directed introductions of stop codons in the coding frames of HIV proteins integrated in the nuclear genomes of “elite” HIV controller individuals [68, 69].

Acknowledgements

This study was supported by Méditerranée Infection and the National Research Agency under the program “Investissements d’avenir,” reference ANR-10-IAHU-03, and the A*MIDEX project (no ANR-11-IDEX-0001-02).

Conflict of interest

The author declares no conflict of interest.

Author details

Hervé Seligmann^{1,2*}

*Address all correspondence to: varanuseremius@gmail.com

1 Aix-Marseille University, URMITE, UM 63, CNRS UMR7278, IRD 198, INSERM U1095, Hospitalo-Universitary Institute, Méditerranée-Infection, Marseille, France

2 The National Natural History Collections, The Hebrew University of Jerusalem, Jerusalem, Israel

References

- [1] Seligmann H. Phylogeny of genetic codes and punctuation codes within genetic codes. *Bio Systems*. 2015;**129**:36-43
- [2] Seligmann H. Alignment-based and alignment-free methods converge with experimental data on amino acids coded by stop codons at split between nuclear and mitochondrial genetic codes. *Bio Systems*. 2018;**167**:33-46

- [3] Seligmann H. Avoidance of antisense antiterminator tRNA anticodons in vertebrate mitochondria. *Bio Systems*. 2010;**101**:42-50
- [4] Seligmann H. Undetected antisense tRNAs in mitochondrial genomes? *Biology Direct*. 2010;**5**:39
- [5] Seligmann H. Pathogenic mutations in antisense mitochondrial tRNAs. *Journal of Theoretical Biology*. 2011;**269**:287-296
- [6] Seligmann H. Two genetic codes, one genome: Frameshifted primate mitochondrial genes code for additional proteins in presence of antisense antitermination tRNAs. *Bio Systems*. 2011;**105**:271-285
- [7] Seligmann H. An overlapping genetic code for frameshifted overlapping genes in *Drosophila* mitochondria: Antisense antitermination tRNAs UAR insert serine. *Journal of Theoretical Biology*. 2012;**298**:51-76
- [8] 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*, Chapter 4. Nova Science Publisher. 2013. pp. 67-86
- [9] Seligmann H. Overlapping genetic codes for overlapping frameshifted genes in Testudines, and *Lepidochelys olivacea* as a special case. *Computational Biology and Chemistry*. 2012;**41**: 18-34
- [10] Barthélémy R-M, Seligmann H. Cryptic tRNAs in chaetognath mitochondrial genomes. *Computational Biology and Chemistry*. 2016;**62**:119-132
- [11] Faure E, Delaye L, Tribolo S, Levasseur A, Seligmann H, Barthélémy R-M. Probable presence of an ubiquitous cryptic mitochondrial gene on the antisense strand of the cytochrome oxidase I gene. *Biology Direct*. 2011;**6**:56
- [12] Breton S, Milani L, Ghiselli F, Guerra D, Stewart DT, Passamonti M. A resourceful genome: Updating the functional repertoire and evolutionary role of animal mitochondrial DNAs. *Trends in Genetics*. 2014;**30**:555-564
- [13] Capt C, Passamonti M, Breton S. The human mitochondrial genome may code for more than 13 proteins. *Mitochondrial DNA Part A, DNA mapping, sequencing, and analysis*; **27**:3098-3101
- [14] Seligmann H, Krishnan NM, Rao BJ. Possible multiple origins of replication in primate mitochondria: Alternative role of tRNA sequences. *Journal of Theoretical Biology*. 2006, 2015;**241**:321-332
- [15] Seligmann H, Krishnan NM, Rao BJ. Mitochondrial tRNA sequences as unusual replication origins: Pathogenic implications for *Homo sapiens*. *Journal of Theoretical Biology*. 2006;**243**:375-385
- [16] Seligmann H, Krishnan NM. Mitochondrial replication origin stability and propensity of adjacent tRNA genes to form putative replication origins increase developmental stability in lizards. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution*. 2006;**306**:433-4449

- [17] 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. *Journal of Molecular Biology*. 2008;**379**:188-199
- [18] 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
- [19] Seligmann H. Mutation patterns due to converging mitochondrial replication and transcription increase lifespan, and cause growth rate-longevity tradeoffs. In: Seligmann H, editor. *DNA Replication-Current Advances*, Chapter 6. Rijeka, Croatia: InTech; 2011. pp. 151-180
- [20] Seligmann H. Replicational mutation gradients, dipole moments, nearest neighbor effects and DNA polymerase gamma fidelity in human mitochondrial genomes. In: Stuart D, editor. *The Mechanisms of DNA Replication*, Chapter 10. Rijeka, Croatia: InTech; 2013. pp. 257-286
- [21] 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**:248-257
- [22] Atkins JF, Baranov PV. Translation: Duality in the genetic code. *Nature*. 2007;**448**:1004-1005
- [23] Ivanova NN, Schwientek P, Tripp HJ, Rinke C, Pati A, Huntemann M, et al. Stop codon reassignments in the wild. *Science*. 2014;**344**:909-913
- [24] Sánchez-Silva R, Villalobo E, Morin L, Torres A. A new noncanonical nuclear genetic code: Translation of UAA into glutamate. *Current Biology*. 2003;**13**:442-447
- [25] Heaphy SM, Mariotti M, Gladyshev VN, Atkins JF, Baranov PV. Novel ciliate genetic code variants including the reassignment of all three stop codons to sense codons in *Condylostoma magnum*. *Molecular Biology and Evolution*. 2016;**33**:2885-2889
- [26] Swart EC, Serra V, Petroni G, Nowacki M. Genetic codes with no dedicated stop codon: Context-dependent translation termination. *Cell*. 2016;**166**:691-702
- [27] Záhonová K, Kostygov AY, Ševčíková T, Yurchenko V, Eliáš M. An unprecedented non-canonical nuclear genetic code with all three termination codons reassigned as sense codons. *Current Biology*. 2016;**26**:2364-2369
- [28] Mühlhausen S, Findeisen P, Plessmann U, Urlaub H, Kollmar M. A novel nuclear genetic code alteration in yeasts and the evolution of codon reassignment in eukaryotes. *Genome Research*. 2016;**26**:945-955
- [29] Riddle DL, Carbon J. Frameshift suppression: A nucleotide addition in the anticodon of a glycine transfer RNA. *Nature: New Biology*. 1973;**242**:230-234
- [30] 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 Research*. 1992;**20**:3463-3469

- [31] 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. *Journal of Molecular Biology*. 1992;**228**:1042-1054
- [32] Seligmann H. Putative mitochondrial polypeptides coded by expanded quadruplet codons, decoded by antisense tRNAs with unusual anticodons. *Biosystems*. 2012;**110**:84-106
- [33] Seligmann H. Pocketknife tRNA hypothesis: Anticodons in mammal mitochondrial tRNA side-arm loops translate proteins? *Biosystems*. 2013;**113**:165-175
- [34] Seligmann H. Putative anticodons in mitochondrial tRNA sidearm loops: Pocketknife tRNAs? *Journal of Theoretical Biology*. 2014;**340**:155-163
- [35] Seligmann H, Labra A. Tetracoding increases with body temperature in Lepidosauria. *Biosystems*. 2013;**114**:155-166
- [36] Baranov PV, Venin M, Provan G. Codon size reduction as the origin of the triplet genetic code. *PLoS One*. 2009;**4**:e5708
- [37] 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. *Nature Precedings*; 2012. Available at: <http://dx.doi.org/10.1038/npre.2012.7136.1>
- [38] Seligmann H. Codon expansion and systematic transcriptional deletions produce tetra-, pentacoded mitochondrial peptides. *Journal of Theoretical Biology*. 2015;**387**:154-165
- [39] Seligmann H. Translation of mitochondrial swinger RNAs according to tri-, tetra- and pentacodons. *Biosystems*. 2016;**140**:38-48
- [40] Seligmann H. Natural chymotrypsin-like-cleaved human mitochondrial peptides confirm tetra-, pentacodon, non-canonical RNA translations. *Biosystems*. 2016;**147**:78-93
- [41] Seligmann H. Unbiased mitoproteome analyses confirm non-canonical RNA, expanded codon translations. *Computational and Structural Biotechnology Journal*. 2016;**14**:391-403
- [42] Seligmann H. Natural mitochondrial proteolysis confirms transcription systematically exchanging/deleting nucleotides, peptides coded by expanded codons. *Journal of Theoretical Biology*. 2017;**414**:76-90
- [43] Seligmann H. Reviewing evidence for systematic transcriptional deletions, nucleotide exchanges, and expanded codons, and peptide clusters in human mitochondria. *Biosystems*. 2017;**160**:10-24
- [44] Seligmann H. Species radiation by DNA replication that systematically exchanges nucleotides? *Journal of Theoretical Biology*. 2014;**363**:216-222
- [45] Seligmann H. Mitochondrial swinger replication: DNA replication systematically exchanging nucleotides and short 16S ribosomal DNA swinger inserts. *Biosystems*. 2014;**125**:22-31
- [46] 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 Part A, DNA mapping, sequencing, and analysis*. 2016;**27**:2440-2446

- [47] Seligmann H. Overlapping genes in the 3'-to-5'-direction in mitochondrial genes and 3'-to-5' polymerization of non-complementary RNA by an 'invertase'. *Journal of Theoretical Biology*. 2012;**315**:38-82
- [48] Seligmann H. Triplex DNA:RNA, 3'-to-5' inverted RNA and protein coding in mitochondrial genomes. *Journal of Computational Biology*. 2013;**20**:660-671
- [49] Seligmann H. Polymerization of non-complementary RNA: Systematic symmetric nucleotide exchanges mainly involving uracil produce mitochondrial RNA transcripts coding for cryptic overlapping genes. *Biosystems*. 2013;**111**:156-174
- [50] Seligmann H. Systematic asymmetric nucleotide exchanges produce human mitochondrial RNAs cryptically encoding for overlapping protein coding genes. *Journal of Theoretical Biology*. 2013;**324**:1-20
- [51] Michel CJ, Seligmann H. Bijective transformation circular codes and nucleotide exchanging RNA transcription. *Biosystems*. 2014;**118**:39-50
- [52] Seligmann H, Raoult D. Stem-loop RNA hairpins in giant viruses: Invading rRNA-like repeats and a template free RNA. *Frontiers in Microbiology*. 2018;**9**:101
- [53] Warthi G, Seligmann H. Swinger RNAs in the human mitochondrial transcriptome. In: mitochondrial DNA-new insights, Seligmann H, Warthi G, editors. Chapter 4, InTech. ISBN: 978-953-51-6167-7
- [54] Bloch DP, McArthur B, Widdowson R, Spector D, Guimaraes RC, Smith J. tRNA-rRNA sequence homologies: Evidence for a common evolutionary origin? *Journal of Molecular Evolution*. 1983;**19**:420-408
- [55] Bloch DP, McArthur B, Widdowson R, Spector D, Guimaraes RC, Smith J. tRNA-rRNA sequence homologies: A model for the origin of a common ancestral molecule, and prospects for its reconstruction. *Origins of Life*. 1983;**14**:571-578
- [56] Bloch DP, McArthur B, Guimaraes RC, Smith J, Staves MP. tRNA-rRNA sequence matches from inter- and intraspecies comparisons suggest common origins for the two RNAs. *Brazilian Journal of Medical and Biological Research*. 1983;**22**:931-944
- [57] Root-Bernstein M, Root-Bernstein R. The ribosome as a missing link in the evolution of life. *Journal of Theoretical Biology*. 2015;**367**:130-158
- [58] Root-Bernstein M, Root-Bernstein R. The ribosome as a missing link in the evolution II: Ribosomes encode ribosomal proteins that bind to common regions of their own mRNAs and rRNAs. *Journal of Theoretical Biology*. 2015;**397**:115-127
- [59] Kermekchiev M, Ivanova L. Ribin, a protein encoded by a message complementary to rRNA, modulates ribosomal transcription and cell proliferation. *Molecular and Cellular Biology*. 2001;**21**:8255-8263
- [60] Barthélémy R-M, Jaafari N, Galea P, Jule Y, Faure E. Immunodetection of ribin-like proteins in neuron-based cellular models. *Journal of Medical Genetics and Genomics*. 2010;**2**:44-56
- [61] Caetano-Anollés D, Caetano-Anollés G. Piecemeal buildup of the genetic code, ribosomes, and genomes from primordial tRNA building blocks. *Life (Basel, Switzerland)*. 2016;**6**:E43

- [62] Seligmann H. Swinger RNAs with sharp switches between regular transcription and transcription systematically exchanging ribonucleotides: Case studies. *Biosystems*. 2015;**135**:1-8
- [63] Seligmann H. Chimeric mitochondrial peptides from contiguous regular and swinger RNA. *Computational and Structural Biotechnology Journal*. 2016;**14**:283-297
- [64] Seligmann 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. *Current Opinion in Microbiology*. 2016;**31**:1-8
- [65] El Houmami N, Seligmann H. Evolution of nucleotide punctuation marks: From structural to linear signals. *Frontiers in Genetics*. 2017;**8**:36
- [66] Ojala D, Montoya J, Attardi G. tRNA punctuation model of RNA processing in human mitochondria. *Nature*. 1981;**290**:470-474
- [67] Stewart JB, Beckenbach AT. Characterization of mature mitochondrial transcripts in *Drosophila*, and the implications for the tRNA punctuation model in arthropods. *Gene*. 2009;**445**:49-57
- [68] Shimpi GG, Vargas S, Polisen A, Wörheide G. Mitochondrial RNA processing in absence of tRNA punctuations in octocorals. *BMC Molecular Biology*. 2017;**18**:16
- [69] Seligmann H. Swinger RNA self-hybridization and mitochondrial non-canonical swinger transcription, transcription systematically exchanging nucleotides. *Journal of Theoretical Biology*. 2016;**399**:84-91
- [70] Claverie JM, Abergel G. Mimivirus and its viroplasm. *Annual Review of Genetics*. 2009;**43**:49-66
- [71] Seligmann H. Giant viruses as protein-coated mitochondria? *Virus Research*. 2018;**253**:77-86
- [72] 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;**142-143**:43-51
- [73] Colson P, Ravaux I, Tamalet C, Glazunova O, Baptiste E, Chabrière E, et al. HIV infection en route to endogenization: Two cases. *Clinical Microbiology and Infection*. 2014;**20**:1280-1288
- [74] Tamalet C, Colson P, Decroly E, Dhiver C, Ravaux I, Stein A, et al. Reevaluation of possible outcomes of infections with human immunodeficiency virus. *Clinical Microbiology and Infection*. 2016;**22**:299-311
- [75] Jestin J-L, Kempf A. Chain termination codons and polymerase-induced frameshift mutations. *FEBS Letters*. 1997;**419**:153-156
- [76] Mindell DP, Sorenson MD, Dimcheff DE. An extra nucleotide is not translated in mitochondrial ND3 of some birds and turtles. *Molecular Biology and Evolution*. 1998;**15**:1568-1571

- [77] Russell RD, Beckenback AT. Recoding of translation in turtle mitochondrial genomes: Programmed frameshift and evidence of a modified genetic code. *Journal of Molecular Evolution*. 2008;**67**:682-695
- [78] Seligmann H. Do anticodons of misacylated tRNAs preferentially mismatch codons coding for the misloaded amino acid? *BMC Molecular Biology*. 2010;**11**:41
- [79] Seligmann H. Error compensation of tRNA misacylation by codon-anticodon mismatch prevents translational amino acid misinsertion. *Computational Biology and Chemistry*. 2011;**35**:81-95
- [80] Seligmann H. Coding constraints modulate chemically spontaneous mutational replication gradients in mitochondrial genomes. *Current Genomics*. 2012;**13**:37-54
- [81] Chang DD, Clayton DA. Precise identification of individual promoters for transcription of each strand of human mitochondrial DNA. *Cell*. 1984;**36**:635-643
- [82] Abergel C, Rudinger-Thirion J, Giegé R, Claverie JM. Virus-encoded aminoacyl-tRNA synthetases: Structural and functional characterization of mimivirus TyrRS and MetRS. *Journal of Virology*. 2007;**81**:12406-12417
- [83] Wang X, Wang X, Chen G, Zhang J, Liu Y, Yang C. The shiftability of protein coding genes: The genetic code was optimized for frameshift tolerating. *PeerJ PrePrints*. 2015;**3**:e806v1
- [84] Wang X, Dong Q, Chen G, Zhang J, Liu Y, Zhao J, et al. Why are frameshift homologs widespread within and across species? *bioRxiv*. 2016:067736. DOI: 10.1101/067736
- [85] Geyer R, Mamlouk A. On the efficiency of the genetic code after frameshift mutations. *PeerJ*. 2018;**6**:e4825
- [86] Seligmann H, Pollock DD. The ambush hypothesis: Hidden stop codons prevent off-frame gene reading. The MidSouth Computational Biology and Bioinformatics Society. 2003. Abstract 36
- [87] Seligmann H, Pollock DD. The ambush hypothesis: Hidden stop codons prevent off-frame gene reading. *DNA and Cell Biology*. 2004;**23**:701-705
- [88] Seligmann H. The ambush hypothesis at the whole-organism level: Off frame, 'hidden' stops in vertebrate mitochondrial genes increase developmental stability. *Computational Biology and Chemistry*. 2010;**34**:80-85
- [89] Itzkovitz S, Alon U. The genetic code is nearly optimal for allowing additional information within protein-coding sequences. *Genome Research*. 2007;**17**:405-412
- [90] Tse H, Cai JJ, Tsoi HW, Lam EP, Zuen KZ. Natural selection retains overrepresented out-of-frame stop codons against frameshift peptides in prokaryotes. *BMC Genomics*. 2010;**11**:491
- [91] Bar-Yaacov D, Avital G, Levin L, Richards AL, Hachen N, Rebolledo Jaramillo B, et al. RNA-DNA differences in human mitochondria restore ancestral form of 16S ribosomal RNA. *Genome Research*. 2013;**23**:1789-1796. DOI: 10.1101/gr.161265.113

- [92] Popov O, Segal D, Trifonov EN. Linguistic complexity of protein sequences as compared to texts of human languages. *Biosystems*. 1996;**38**:65-74
- [93] Seligmann H. Bijective codon transformations show genetic code symmetries centered on cytosine's coding properties. *Theory in Biosciences*. 2018;**137**:17-31. DOI: 10.1007/s12064-017-0258-x
- [94] Arquès DG, Michel CJ. A complementary circular code in the protein coding genes. *Journal of Theoretical Biology*. 1996;**182**:45-58
- [95] Arquès DG, Lacan J, Michel CJ. Identification of protein coding genes in genomes with statistical functions based on the circular code. *Bio Systems*. 2002;**66**:73-92
- [96] Frey G, Michel CJ. Circular codes in archaeal genomes. *Journal of Theoretical Biology*. 2003;**223**:413-431
- [97] Ahmed A, Frey G, Michel CJ. Frameshift signals in genes associated with the circular code. *In Silico Biology*. 2007;**7**:155-168
- [98] Ahmed A, Frey G, Michel CJ. Essential molecular functions associated with the circular code evolution. *Journal of Theoretical Biology*. 2010;**264**:613-622. DOI: 10.1016/j.jtbi.2010.02.006
- [99] Gonzalez DL, Giannerini S, Rosa R. Circular codes revisited: A statistical approach. *Journal of Theoretical Biology*. 2011;**275**:21-28. DOI: 10.1016/j.jtbi.2011.01.028
- [100] Michel CJ. Circular code motifs in transfer and 16S ribosomal RNAs: A possible translation code in genes. *Computational Biology and Chemistry*. 2012;**37**:24-37. DOI: 10.1016/j.combiolchem.2011.10.002
- [101] Michel CJ, Pirillo G. A permuted set of a trinucleotide circular code coding the 20 amino acids in variant nuclear codes. *Journal of Theoretical Biology*. 2013;**319**:116-121. DOI: 10.1016/j.jtbi.2012.11.023
- [102] Michel CJ. A genetic scale of reading frame coding. *Journal of Theoretical Biology*. 2014;**355**:83-94. DOI: 10.1016/j.jtbi.2014.03.029
- [103] Fimmel E, Strüngmann L. On the hierarchy of trinucleotide n-circular codes and their corresponding amino acids. *Journal of Theoretical Biology*. 2015;**364**:113-120. DOI: 10.1016/j.jtbi.2014.09.011
- [104] Michel CJ. An extended genetic scale of reading frame coding. *Journal of Theoretical Biology*. 2015;**365**:164-174. DOI: 10.1016/j.jtbi.2014.09.040
- [105] Michel CJ. The maximal C(3) self-complementary trinucleotide circular code X in genes of bacteria, eukaryotes, plasmids and viruses. *Journal of Theoretical Biology*. 2015;**380**:156-177. DOI: 10.1016/j.jtbi.2015.04.009
- [106] Fimmel E, Strüngmann L. Codon distribution in error-detecting circular codes. *Life (Basel, Switzerland)*. 2016;**6**(1). pii: E14. DOI: 10.3390/life6010014
- [107] El Soufi K, Michel CJ. Circular code motifs in genomes of eukaryotes. *Journal of Theoretical Biology*. 2016;**408**:198-212. DOI: 10.1016/j.jtbi.2016.07.022

- [108] El Soufi K, Michel CJ. Unitary circular code motifs in genomes of eukaryotes. *Biosystems*. 2017;**153-154**:45-62. DOI: 10.1016/j.biosystems.2017.02.001
- [109] Michel CJ. The maximal C³ self-complementary trinucleotide circular code X in genes of bacteria, archaea, eukaryotes, plasmids and viruses. *Life (Basel, Switzerland)*. 2017;**7**. pii: E20. DOI: 10.3390/life7020020
- [110] Michel CJ, Ngoune VN, Poch O, Ripp R, Thompson JD. Enrichment of circular code motifs in the genes of the yeast *Saccharomyces cerevisiae*. *Life (Basel, Switzerland)*. 2017;**7**(4). pii: E52. DOI: 10.3390/life7040052
- [111] Fimmel E, Michel CJ, Strüngmann L. Strong comma-free codes in genetic information. *Bulletin of Mathematical Biology*. 2017;**79**:1796-1819. DOI: 10.1007/s11538-017-0307-0
- [112] Fimmel E, Strüngmann L. Mathematical fundamentals for the noise immunity of the genetic code. *Biosystems*. 2018;**164**:186-198. pii: S0303-2647(17)30237-X. DOI: 10.1016/j.biosystems.2017.09.007
- [113] Michel CJ. Circular code motifs in transfer RNAs. *Computational Biology and Chemistry*. 2013;**45**:17-29. DOI: 10.1016/j.compbiolchem.2013.02.004
- [114] El Soufi K, Michel CJ. Circular code motifs in the ribosome decoding center. *Computational Biology and Chemistry*. 2014;**52**:9-17. DOI: 10.1016/j.compbiolchem.2014.08.001
- [115] El Soufi K, Michel CJ. Circular code motifs near the ribosome decoding center. *Computational Biology and Chemistry*. 2015;**59 Pt A**:158-176. DOI: 10.1016/j.compbiolchem.2015.07.015
- [116] Fimmel E, Michel CJ, Starman M, Strüngmann L. Self-complementary circular codes in coding theory. *Theory in Biosciences*. 2018;**137**:51-65
- [117] Seligmann H, Warthi G. Genetic code optimization for cotranslational protein folding: Codon directional asymmetry correlates with antiparallel beta-sheets, tRNA synthetase classes. *Computational and Structural Biotechnology Journal*. 2017;**15**:412-424. DOI: 10.1016/j.csbj.2017.08.001
- [118] Hall BG. Spontaneous point mutations that occur more often when advantageous than when neutral. *Genetics*. 1990;**126**:5-16
- [119] Li G-D. Natural site-directed mutagenesis might exist in eukaryotic cells. arXiv. 2016. arXiv:1603.06538
- [120] Saier MH Jr, Zhang Z. Transposon-mediated directed mutation controlled by DNA binding proteins in *Escherichia coli*. *Frontiers in Microbiology*. 2014;**5**:390
- [121] Satoh TP, Sato Y, Masuyama N, Miya M, Nishida M. Transfer RNA gene arrangement and codon usage in vertebrate mitochondrial genomes: A new insight into gene order conservation. *BMC Genomics*. 2010;**11**:479
- [122] Abascal F, Posada D, Knight RD, Zardova R. Parallel evolution of the genetic code in arthropod mitochondrial genomes. *PLoS Biology*. 2006;**4**:e127
- [123] Abascal D, Posada D, Zardova R. The evolution of the mitochondrial genetic code in arthropods revisited. *Mitochondrial DNA*. 2012;**23**:84-91. DOI: 10.3109/19401736.2011.653801

