



# Evolution of Nucleotide Punctuation Marks: From Structural to Linear Signals

Nawal El Houmami and Hervé Seligmann\*

URMITE, Aix Marseille Université UM63, CNRS 7278, IRD 198, INSERM 1095, IHU – Méditerranée Infection, Marseille, France

## OPEN ACCESS

### Edited by:

Maria Poptsova,  
Moscow State University, Russia

### Reviewed by:

David B. Wilson,  
Cornell University, USA  
Hermona Soreq,  
Hebrew University of Jerusalem, Israel

### \*Correspondence:

Hervé Seligmann  
timonuslepidus@gmail.com

### Specialty section:

This article was submitted to  
Bioinformatics and Computational  
Biology,  
a section of the journal  
Frontiers in Genetics

**Received:** 14 October 2016

**Accepted:** 13 March 2017

**Published:** 27 March 2017

### Citation:

El Houmami N and Seligmann H  
(2017) Evolution of Nucleotide  
Punctuation Marks: From Structural  
to Linear Signals. *Front. Genet.* 8:36.  
doi: 10.3389/fgene.2017.00036

We present an evolutionary hypothesis assuming that signals marking nucleotide synthesis (DNA replication and RNA transcription) evolved from multi- to unidimensional structures, and were carried over from transcription to translation. This evolutionary scenario presumes that signals combining secondary and primary nucleotide structures are evolutionary transitions. Mitochondrial replication initiation fits this scenario. Some observations reported in the literature corroborate that several signals for nucleotide synthesis function in translation, and vice versa. (a) Polymerase-induced frameshift mutations occur preferentially at translational termination signals (nucleotide deletion is interpreted as termination of nucleotide polymerization, paralleling the role of stop codons in translation). (b) Stem-loop hairpin presence/absence modulates codon-amino acid assignments, showing that translational signals sometimes combine primary and secondary nucleotide structures (here codon and stem-loop). (c) Homopolymer nucleotide triplets (AAA, CCC, GGG, TTT) cause transcriptional and ribosomal frameshifts. Here we find in recently described human mitochondrial RNAs that systematically lack mono-, dinucleotides after each trinucleotide (delRNAs) that delRNA triplets include 2x more homopolymers than mitogenome regions not covered by delRNA. Further analyses of delRNAs show that the natural circular code X (a little-known group of 20 translational signals enabling ribosomal frame retrieval consisting of 20 codons {AAC, AAT, ACC, ATC, ATT, CAG, CTC, CTG, GAA, GAC, GAG, GAT, GCC, GGC, GGT, GTA, GTC, GTT, TAC, TTC} universally overrepresented in coding versus other frames of gene sequences), regulates frameshift in transcription and translation. This dual transcription and translation role confirms for X the hypothesis that translational signals were carried over from transcriptional signals.

**Keywords:** stem-loop hairpin, secondary structure, nucleotide motif, transcription signals, codon-amino acid assignment

## INTRODUCTION

Punctuation marks are inherent to written systems by providing a critical framework for specifying information. Spread along nucleotide sequences, the study of punctuation signals is relatively neglected in genetics and deserves interdisciplinary attention combining molecular biology, linguistics, and coding theory. Involving 64 nucleotide triplets called codons (Elzanowski and Ostell, 2013), the genetic code is a system coding the set of rules by which information is

translated from RNA into proteins by living cells and viruses, by specifying which amino acid will be added during protein synthesis. Information encoded within genetic material also possesses superimposed cryptic messages (Popov et al., 1996), revealing highly complex semiotics (e.g., in circular virusoid RNAs, AbouHaidar et al., 2014). The rules of DNA punctuation vary among 25 recognized genetic codes, suggesting these constantly evolve. Codon–amino acid assignment evolved mainly by changes in punctuation codons, namely initiation (start) and termination (stop) codons (Seligmann, 2015b), impacting length and structures of coding and non-coding DNA sequences.

Here analyses focus on the evolution of punctuation signals, assuming two core principles. First, primitive punctuation of nucleotide sequences consists of multidimensional structures (such as stem-loop hairpins) allowing form recognition by DNA and RNA polymerases. This is illustrated by hairpins that are signals formed by self-hybridization of nucleotides indicating where DNA and RNA polymerizations initiate (e.g., in vertebrate mitochondria, Clayton, 1992). Second, we presume that multi- and unidimensional punctuation marks used for nucleotide synthesis were secondarily hijacked for translation. Thus, we propose a multistep model where multidimensional structures later evolved into linear signals, in parallel to book page earmarking using structural recognition, versus memorizing page numbers requiring a consensual code. In this model, we assume that protein synthesis emerged after that of nucleotides.

## From Multidimensional to Unidimensional Punctuation Signals

Hairpins are structural signals spread within genomes of all organisms. In human mitochondria, stem-loop DNA structures define replication origins (Hixson et al., 1986; Clayton, 1992, 2000; Seligmann and Krishnan, 2006; Seligmann et al., 2006a,b; Seligmann, 2008, 2010a, 2011; Seligmann and Labra, 2014). They guide RNA processing in mitochondria (Ojala et al., 1981), and in giant viruses and their viroplasmids (Byrne et al., 2009; Claverie and Abergel, 2009). This ubiquitous structural signaling also applies to RNA:DNA hybrids, which play a role in the origin-independent replication priming in eukaryotic cells (Stuckey et al., 2015) and in transcription termination in human mitochondria (Zheng et al., 2014).

However, different nucleotide sequences can form similar secondary structures. This generates structural ambiguity responsible for versatile and non-specific signals. For example, in prokaryote RNA-based defense systems against genome invasion by parasites (viruses and plasmids), structural ambiguity of dual-RNAs that guide nucleases to degrade invading DNA is used by bacteria possessing orthologous type II CRISPR-Cas defense systems, for which functional exchangeability was recently demonstrated (Fonfara et al., 2014). The dual RNA duplexes consist of hybridization between tracrRNA and crRNA. The tracrRNAs are trans-encoded RNAs that complement crRNAs, short palindromic repeats. Also, analysis of structural RNA similarities between rRNA and RNA viruses unraveled an ancient transition from cellular organisms possessing ribosomes to viruses (Seligmann and Raoult, 2016), which is undetectable

when considering only unidimensional (linear) sequence information.

Synthesis of short consensus signals, such as Shine–Dalgarno sequences, is more cost-effective than that of nucleotide sequences forming secondary structures. Hence avoidance of metabolic costs (Akashi and Gojobori, 2002; Seligmann, 2003, 2012b; Brocchieri and Karlin, 2005; Warringer and Blomberg, 2006; Heizer et al., 2011; Chen and Bundschuh, 2012; Raiford et al., 2012; Krick et al., 2014; Chen W.-H. et al., 2016) should favor evolution of linear consensus signals. Consequently, linear signals presumably evolved more recently to become punctuation marks with higher accuracy, specialization, and metabolic efficiency than structural signals.

In some cases, enzymatic recognition requires both primary and secondary structures. In mitochondria, initiation of DNA polymerization requires a short specific sequence close to the 3' extremity of the light strand replication origin hairpin (Hixson et al., 1986; Clayton, 2003; Wanrooij et al., 2012). We herein propose a three-step model where multidimensional structures later evolved into linear signals.

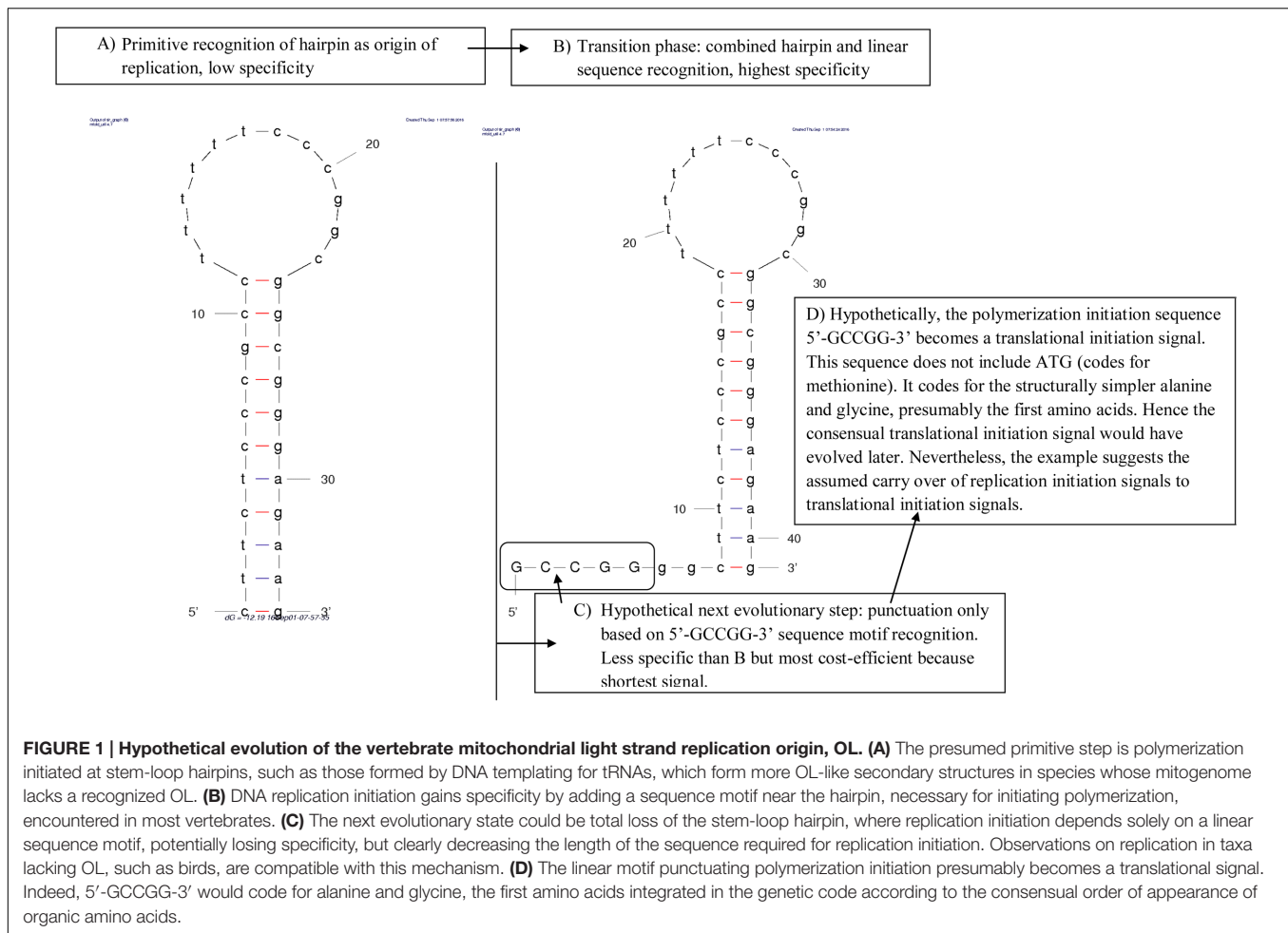
## Evolution of Origin of Replication: From Structural to Linear Signals

We illustrate this evolutionary scenario by applying it to the light strand origin of replication (OL) of vertebrate mitochondria (Figure 1). In phase A, polymerases only recognize stem-loop hairpins. Indeed, in mitochondria, heavy strand DNA templating for tRNAs form OL-like structures that occasionally function as light strand replication origin (Seligmann et al., 2006b). The polyT sequence in the modern OL loop marks mitochondrial RNA polymerase binding of the OL (Fusté et al., 2010).

Several functional homologies between tRNAs and OL corroborate the hypothesis that tDNA functions as OL. This is in line with functional homologies suggesting that the translation apparatus evolved from DNA/RNA replication (Maizels and Weiner, 1994). First, the OL might function as tRNA because aminoacylated RNA corresponding to the OL was observed (Yu et al., 2008). Second, some tRNAs resemble the OL structure: short armless OL-like tRNAs occur in nematode mitochondria (He et al., 2005; Jühling et al., 2012; Wende et al., 2014). Third, modern tRNAs apparently result from fusion of OL-like tRNA halves (Di Giulio, 2009, 2012a,b, 2013; Branciamore and Di Giulio, 2011). Furthermore, OL-like isolated tRNA arms (Hirose et al., 2015) could have tRNA functions (Seligmann, 2013, 2014): they resemble the OL-like armless nematode tRNAs and their anticodons match genomic codon abundances.

Assuming that Figure 1A presents the primitive system, this system acquired higher specificity when combined with recognition of a primary sequence motif, 5'-GCCGG-3' (Figure 1B), which is the typical situation in modern genomes.

In phase B, replication initiation requires both stem-loop hairpin and consensus signal as observed in most vertebrate mitochondria. In phase C, we hypothesize that the structural component of the signal was replaced by a cost-effective consensus linear sequence signal. Hence the signal became less specific by losing the stem-loop hairpin, but less costly in terms



of numbers of required constitutive nucleotides, consisting of a shorter linear sequence signal. This evolutionary step could correspond to our incomplete understanding of mitochondrial replication organisms that lack a recognized OL, most birds (Desjardins and Morais, 1991), and some lepidosaurians (Macey et al., 1997; Seligmann and Labra, 2014). Indeed, in bird mitochondria, light-strand replication initiates synchronously at several locations lacking secondary structures (Reyes et al., 2005). Finally, we suggest a hypothetical functional shift from initiation of DNA replication to protein translation (phase D).

## Hijacking of Punctuation Marks from DNA to Protein Synthesis

The principle of evolution from secondary to primary sequence structure punctuation could also be seen at the level of the translational apparatus. Translation involves tRNAs and rRNAs, molecules whose function is inherently based on their multidimensional structure. tRNAs can be interpreted as a combined structural and linear signal, where the anticodon is the linear signal. tmRNAs (transfer-messenger RNA, Janssen and Hayes, 2012; Himeno et al., 2014), which rescue stalled ribosomes on mRNAs lacking proper translational stop signal also combine structural and linear signals, with extremities

forming tRNA-like structures, and the rest of the sequence reminding mRNAs (Di Giulio, 2015; Macé and Gillet, 2016). This suggests that tmRNAs are remnants of ancient termination signals. It seems that translation of linear sequence signals (codons) is based on their interplay with multidimensional RNA structures (Brown et al., 2015). This principle applies to dual functions of codons: presence of a stem-loop hairpin on the mRNA determines which amino acid is inserted at the codon, changing the translational signal of the codon (Lobanov et al., 2010).

## From Non-ribosomal Peptide Synthesis to mRNA Translation

Translation by ribosomes is complex and was probably absent at molecular evolution's first steps (Baranov et al., 2009; Root-Bernstein and Root-Bernstein, 2015, 2016). Hence proteins were probably first produced by ribosome-free systems, such as non-ribosomal peptide synthesis (Roy and Ibbá, 2010; Chen Y. et al., 2016). We propose that when ribosomal protein translation evolved, it used the same punctuation signals as nucleotide polymerization, especially RNA transcription. This assumption yields the testable prediction that signals punctuating DNA/RNA polymerizations are the same as those in protein translation. This

second principle assumes that synthesis of nucleotide sequences evolved before translation. Hence, we postulate that punctuation signals for transcription were included into the genetic code's classical translation system.

We applied this to the example in **Figure 1**. This yields the evolutionary step D, where the sequence signal 5'-GCCGG-3' used for initiating DNA synthesis after OL-binding could have become the punctuation mark that initiates codon translation. By considering known initiation codons, mainly represented by ATG, the sequence 5'-GCCGG-3' does not fit any one (Elzanowski and Ostell, 2013). However, this sequence, when translated as codons, codes for alanine and glycine, the structurally simplest amino acids. These were the first integrated in the genetic code according to any hypotheses considering emergence of life (Trifonov, 2000, 2004), and according to protein sequence comparisons (Trifonov, 2009). In this case the sequence 5'-GCCGG-3', a consensual signal for initiating replication, would reflect, at the translational level, an ancestral coding sequence for initiation of a primitive form of translation.

### Termination Signals for Nucleotide and Amino Acid Synthesis

The nucleotide triplets TAA, TAG, TGA, which function as termination codons in translation, are hotspots for single nucleotide deletions during polymerization (Jestin and Kempf, 1997). Putatively, these frameshift mutation signals could have become translation termination signals.

### Translation Termination: Structural Signal Versus Codon

Translation termination is not always based on stop codons. The genetic code of some ciliates lacks punctuation signals dedicated to the termination of peptide chain elongation (Swart et al., 2016). Similarly, the termination codon may be missing in some mRNAs (Schaub et al., 2012). As mentioned above, tmRNAs sometimes take this termination role. In some cases, translational termination combines secondary and primary structures, as shown in **Figure 1B** for polymerization signals. This principle is in line with an otherwise unexplained observation of the architecture of the genetic code: codon assignments maximize potential for hairpin formation (Itzkovitz and Alon, 2007), and in parallel maximize numbers of off frame stops, which prevent translation after ribosomal frameshifts (Seligmann and Pollock, 2004; Seligmann, 2007, 2010b, 2012a; Tse et al., 2010; Krizek and Krizek, 2012). This link between translation termination and hairpin formation might reflect that current termination codons replaced hairpins.

### Parallels between Transcription and Translation Frameshifts

A further convergence exists between transcription and translation signals, regarding effects of homopolymer nucleotide triplets (AAA, CCC, GGG, TTT). These homopolymers do not confer any information regarding the reading frame (Crick et al., 1957). They induce frameshift mutations, meaning that nucleotides are 'missed' during DNA and RNA polymerizations

(Atkins et al., 2016). These triplets also cause ribosomal slippage during protein translation (Klobutcher and Farabaugh, 2002; Ketteler, 2012; Advani and Dinman, 2016). In other words, they mark programmed frameshifts for nucleotide polymerizations (DNA replication and RNA transcription) and translation. This classical parallel between nucleotide polymerizations and translation frameshift by homopolymers is tested here on recently described mitochondrial RNAs, so called delRNAs (Seligmann, 2015a, 2016). These RNAs reflect an unusual transcription pattern called del-transcription, in which one or two nucleotides are consistently deleted after each transcribed nucleotide triplet. Here we take advantage of delRNAs to test whether known translational frameshifting signals also function as transcriptional deletion signals. This would enable to test, for frameshifting-deletion signals, the hypothesis that nucleotide polymerization signals were carried over to translation.

### Transcription that Systematically Deletes Nucleotides

The systematic deletions characterizing delRNAs might result from processes including, but not restricted to posttranscriptional editing (Li et al., 2011; Bar-Yaacov et al., 2013; Wang et al., 2014).

Mechanisms regulating del-transcriptions are unknown. We herein tested the hypothesis that detected human delRNAs are enriched in homopolymer nucleotide triplets, as compared to other human mitogenome regions not covered by delRNA. We test a further prediction of the hypothesis that punctuations for DNA/RNA polymerizations were carried over to ribosomal translation of mRNAs into proteins. In this case, we test whether a known system that punctuates ribosomal translation, the natural circular code that enables translation frame retrieval (see below explanations), also regulates systematic mono- and dinucleotide deletions occurring after transcription of nucleotide triplets in mitochondrial delRNA<sub>3-1</sub> and delRNA<sub>3-2</sub> (Seligmann, 2015a).

### The Natural Circular Code X

Initiation and termination codons are translational signals that punctuate boundaries of protein coding sequences. The genetic code also includes a little known system of signals that punctuates the coding frame within protein coding sequences, by regulating the ribosomal translation frame. This intra-gene punctuation system was discovered by analyses of protein coding genes that identified one specific set of 20 codons overrepresented in the protein coding frame as compared to the other, non-coding frames of the genes. These 20 codons constitute a circular code (Arquès and Michel, 1996). Here we will test whether this translational circular code also applies to transcription of delRNAs.

Briefly, within the genetic code, circular codes are sets of codons that allow retrieving the frame of any circular word built by these codons (Lacan and Michel, 2001; Fimmel and Strüngmann, 2016; Fimmel et al., 2016). The notion of punctuation of such a word is inherent to codons composing the circular code. This is because any combination of two codons from such a code produces non-redundant text.

Circular codes by definition can not include the four homopolymer nucleotide triplets AAA, CCC, GGG, and TTT that frequently cause frameshifts. Circular codes can only include the 60 remaining codons. For any codon XYZ in a circular code, its permutations YZX and ZXY can not be part of that circular code. Therefore, for any codon included in a circular code, the two codons formed by its permutations can't be included in that circular code. This means that circular codes include at most 20 codons (60/3). Such circular codes of 20 codons are maximal, and are called 'maximal circular codes' because a circular code with more than 20 codons is impossible in the context of the genetic code. Among 60 codons,  $3^{20} \sim 349$  million potential combinations of 20 codons include  $\sim 13$  million maximal circular codes.

There are only 221554 maximal  $C^3$  circular codes. The property  $C^3$  means that the 20 permutations of the 20 circular code codons XYZ to YZX (example AAC- $\rightarrow$ ACA) form a maximal circular code, and that the 20 permutations of these 20 circular code codons XYZ to ZXY (example AAC- $\rightarrow$ CAA) also form a maximal circular code. There are only 216 circular codes of 20 codons that are self-complementary and have the  $C^3$  property (Gonzalez et al., 2011; Michel, 2014; Fimmel and Strüngmann, 2015). Self-complementarity means that for any codon of the circular code, its inverse complement occurs among the remaining X codons (example AAC/GTT).

The 20 codons {AAC, AAT, ACC, ATC, ATT, CAG, CTC, CTG, GAA, GAC, GAG, GAT, GCC, GGC, GGT, GTA, GTC, GTT, TAC, TTC} overrepresented within protein coding frames of genes constitute one of these 216 above-defined codes (Arquès and Michel, 1996). Indeed, this set of 20 codons has the following properties: (a) it is a circular code (for any codon XYZ, the 19 remaining codons do not include its permutations YZX and ZXY); (b) it is maximal (20 codons); it has self-complementarity (for any codon, the 19 remaining codons include its inverted complement); and (c) it is  $C^3$ . The latter property means that the permutations YZX and ZXY of XYZ produce each sets of 20 codons that are circular codes. These two circular codes are maximal and  $C^3$ , but lack self-complementarity.

Note that the 20 trinucleotides that Arquès and Michel (1996) detected overrepresented in the +1 and the +2 frames of protein coded genes are indeed permutations ZXY and YZX of the the circular code detected in the coding frame. Obtaining this empirical result within natural gene sequences, considering all possible combinations of 20 codons, has  $P = 6.2 \times 10^{-8}$ . These codons code for their assigned amino acid, and also constitute the natural circular code 'X', which enables to retrieve the ribosomal translational frame (Arquès and Michel, 1996; Ahmed et al., 2007, 2010; Michel, 2012; El Soufi and Michel, 2014, 2015).

Codons of X code for thirteen of the twenty natural amino acids. X does not include stop codons, nor homopolymers of nucleotide triplets. X enables to detect the ribosomal translation frame, but mechanisms by which this occurs remain unknown. Codons of X can be seen as opposite to homopolymer nucleotide triplets in terms of signaling the translation frame.

The human mitochondrial delRNAs are analyzed here in terms of the circular code X. This is done in order to test whether X regulates del-transcription. We expect that punctuation by

X is common to transcription and translation, as predicted in **Figure 1**.

## MATERIALS AND METHODS

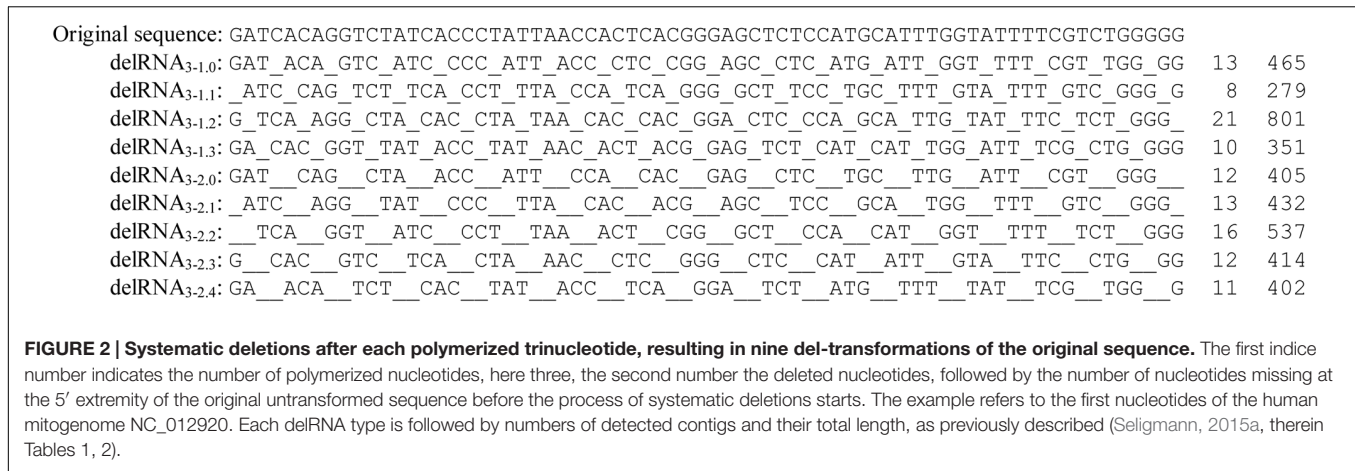
Here we present for the reader's convenience the methods used by Seligmann (2015a) to detect delRNAs. The human mitochondrial mitogenome NC\_012920 was transformed according to systematic mono- and dinucleotide deletions after each trinucleotide. This produces four versions of the mitogenome missing every fourth nucleotide, and five mitogenome versions missing every fourth and fifth nucleotides. **Figure 2** shows these nine transformations for a given sequence.

Note that in principle, systematic deletions might follow patterns that differ from those described in **Figure 2**. Systematic deletions of more than two nucleotides, and after polymerization of less or more than three nucleotides, might exist. This means that delRNAs are characterized by numbers of polymerized nucleotides, followed by numbers of deleted/missing nucleotides. For example, systematic deletions of mononucleotides after each polymerized trinucleotide are noted delRNA<sub>3-1</sub>. Dinucleotide deletions are noted delRNA<sub>3-2</sub>, regular RNAs lacking deletions could be noted in this system delRNA<sub>3-0</sub>. According to this annotation system, the still untested hypothesis of systematic mononucleotide deletions after each polymerized dinucleotide would be noted delRNA<sub>2-1</sub>.

Systematic deletions are characterized by an additional variable, which is the frame of deletion, as compared to the first nucleotide of the regular, untransformed template sequence. Systematic mononucleotide deletions after each polymerized trinucleotide can follow four frames: the process starts at the first nucleotide of the regular, untransformed sequence, polymerizing the three first nucleotides and deleting the fourth, then polymerizing nucleotides five to seven, deleting the eighth, and so on. This case is noted delRNA<sub>3-1.0</sub>. The process can also start at the second, or at the third or the fourth nucleotide, resulting in notations delRNA<sub>3-1.1</sub>, delRNA<sub>3-1.2</sub>, and delRNA<sub>3-1.3</sub>. For systematic dinucleotide deletions after each trinucleotide, corresponding annotations for the five potential deletion frames are delRNA<sub>3-2.0</sub>, delRNA<sub>3-2.1</sub>, delRNA<sub>3-2.2</sub>, delRNA<sub>3-2.3</sub>, and delRNA<sub>3-2.4</sub> (see the nine del-transformations of a sequence in **Figure 2**).

The nine del-transformations of the human mitogenome were analyzed by blastn (Altschul et al., 1997), by comparing them to 72 samples of human transcripts sequenced by RNA-Seq, Illumina HiSeq 2500 technology, from Genbank's Sequence Read Archive (SRA), entries SRX768406-SRX768476 (Garzon et al., 2014).

These blast analyzes detected RNAs matching the nine del-transformed versions of the mitogenome, presented by Seligmann (2015a), therein Tables 1, 2 for delRNAs<sub>3-1</sub> and delRNAs<sub>3-2</sub>, respectively. Here, **Figure 2** indicates the numbers of distinct mitogenome regions (contigs) covered by these RNAs, and the total number of nucleotide sites covered by detected delRNAs, for each of the nine del-transformations examined by Seligmann (2015a). These results from previous analyzes



are presented here within this section because they consist the 'materials' used here for further analyzes.

We counted numbers of nucleotide triplets for each of the nine del-transformations of the human mitogenome, separately for sequences covered by detected delRNAs, and for the rest of the corresponding del-transformed mitogenome (not covered by detected delRNA). Note that the nucleotide triplets counted are for del-transformed versions of the human mitogenome, not for the original untransformed mitogenome. This count is done only for contiguous nucleotide triplets, meaning nucleotide triplets between deleted nucleotide(s): the triplets do not cover nucleotide(s) deleted from the untransformed mitogenome. Separate counts were done for each of the four homopolymer triplets, and for the 20 nucleotide triplets of the natural circular code X.

We calculate the ratios between frequencies of nucleotide triplets within detected delRNAs and their frequencies in the rest of the human mitogenome (not covered by delRNAs). These ratios indicate whether a specific nucleotide triplet is overrepresented within detected delRNAs, as compared to the

rest of the mitogenome when the ratio is >1, or underrepresented when the ratio is <1. We used chi-square statistics to test whether nucleotide triplet counts differ between sequences covered by detected delRNAs, versus the rest of the mitogenome.

## RESULTS

### Excess Homopolymer Nucleotide Triplets in delRNAs

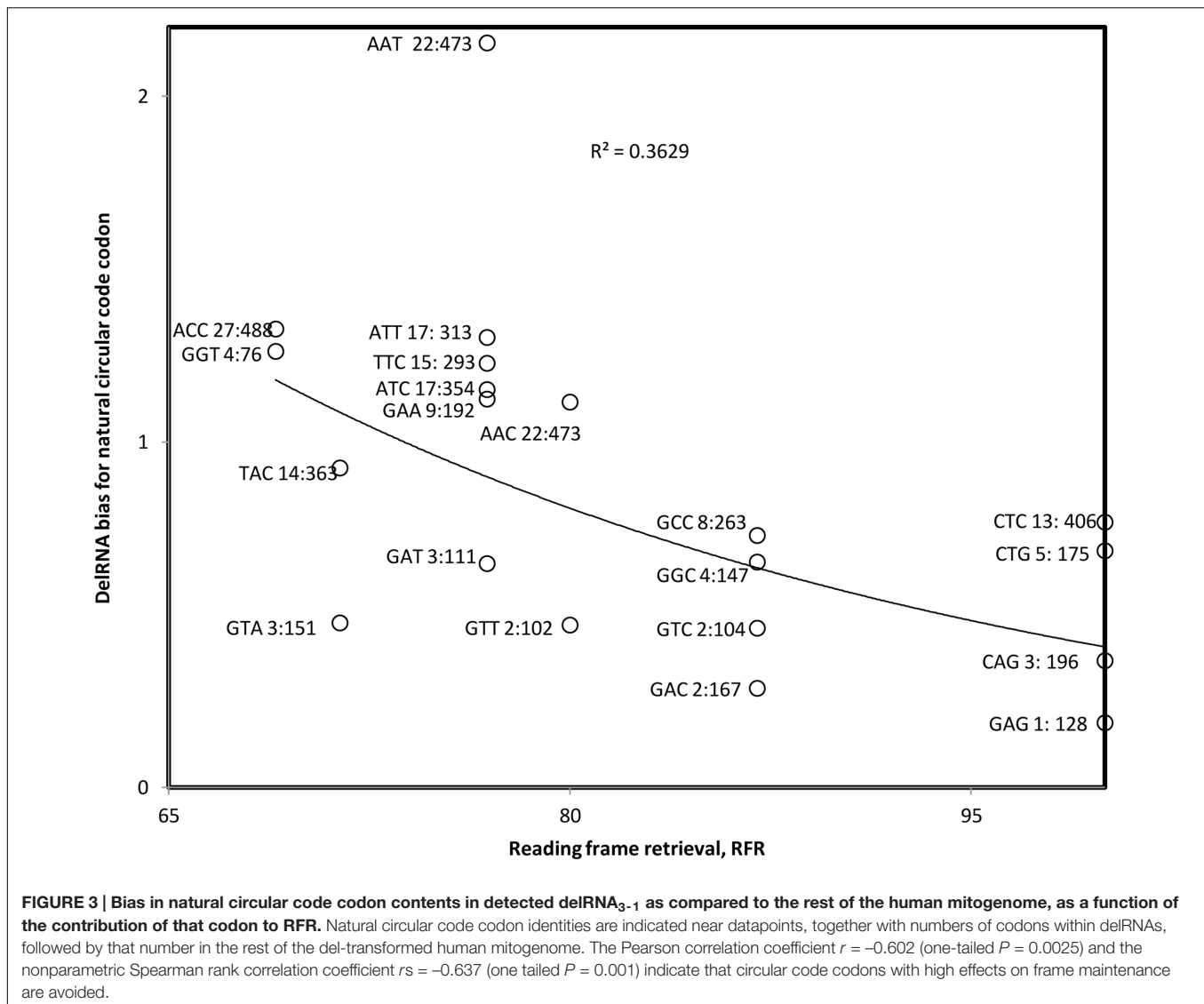
Previous results detected delRNAs for each type of del-transcription (Seligmann, 2015a), meaning RNAs matching the human mitogenome at the condition that one assumes systematic deletions after each transcribed nucleotide triplet. **Figure 2** shows the number of nucleotide sites of the human mitogenome covered by delRNAs, for each delRNA type. Our working hypothesis is that homopolymers contribute to the systematic deletions that produce delRNAs. This would show that del-transcription occurs as a result of signals that are known common to replicational/transcriptional and translational frameshifting, the homopolymer nucleotide triplets. This would put del-transcription in the context of the evolutionary model of carryover of nucleotide polymerization signals to translational signals (**Figure 1**).

**Table 1** compares numbers of homopolymer nucleotide triplets within the deletion frames of detected delRNAs [delRNAs described by Seligmann (2015a), therein Tables 1, 2] and their corresponding number in the same deletion frame in the rest of the mitogenome (for which no delRNA was detected). Detected delRNAs<sub>3-1</sub> and delRNAs<sub>3-2</sub> include 1.51 and 2.35 times more homopolymers, respectively, than del-transformed versions of the human mitogenome for which no delRNA was detected. This difference in homopolymer contents between del-transformed mitogenome regions covered by delRNAs, and the rest of the del-transformed mitogenome, is statistically significant for each delRNAs<sub>3-1</sub> and delRNAs<sub>3-2</sub> ( $P < 0.001$ , chi-square tests). Hence, homopolymer nucleotide triplets are significantly associated with delRNAs. This observation that delRNAs are not random sequences, but specifically enriched in triplets that cause frameshifts excludes that delRNA detections result

**TABLE 1 | Numbers of homopolymers (AAA, CCC, GGG, TTT) among trinucleotides within del-transformed versions of the human mitogenome, for detected delRNAs (as described by Seligmann, 2015a, therein Tables 1, 2), versus corresponding numbers in remaining human mitogenome regions, assuming the same del-transformation (columns headed by 'other').**

Trinucleotide	delRNA <sub>3-1</sub>	Other	delRNA <sub>3-2</sub>	Other
AAA	37	487	61	463
CCC	24	600	47	577
GGG	0	72	1	71
TTT	22	229	17	234
All homopol	83	1388	126	1345
Total	632	15934	730	15837
Percent	13.14	8.71	19.94	8.49

The last line indicates percentages of homopolymer nucleotide triplets in del-transformed sequences. Detected delRNAs have 1.51 and 2.35 times more homopolymers than del-transformed versions of the human mitogenome for which no delRNA has been detected.



from spurious alignments due to the sheer large quantity of transcripts compared with the del-transformed mitogenome. Results confirm the role of homopolymers as transcriptional frameshifting signals in the specific context of systematic deletions during transcription.

### Nucleotide Triplets in delRNAs that Belong to the Natural Circular Code X

We also analyze delRNA nucleotide triplet contents in relation to a known system signaling the translational frame, the natural circular code X, which regulates ribosomal translation. The evolutionary scenario in **Figure 1** assumes that translational signals (such as the circular code) were carried over from nucleotide polymerizations to translation. This hypothesis predicts non-random associations between delRNAs and the natural circular code X.

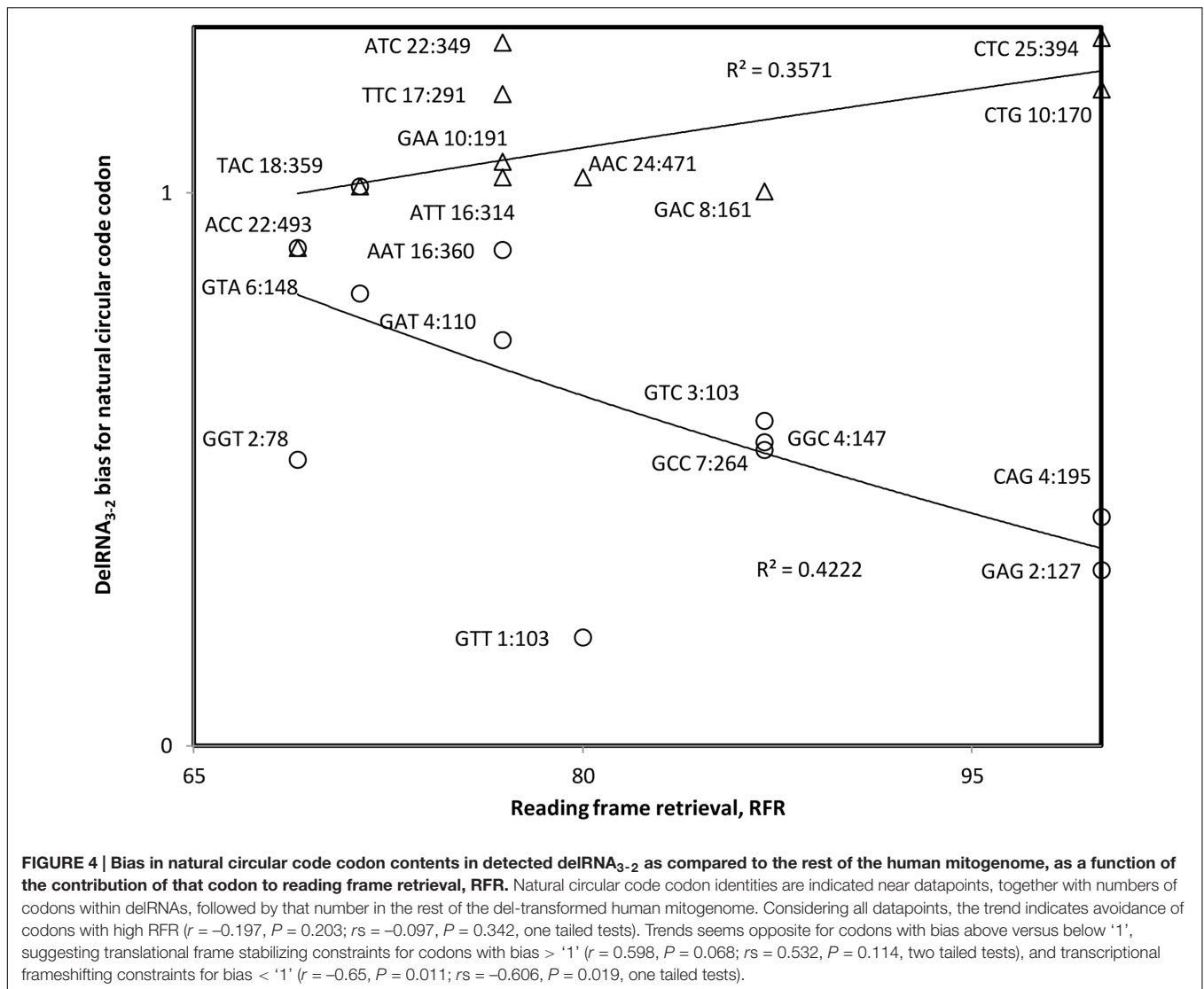
We counted codons of X within previously detected delRNAs, and compared their frequency with that in the rest of the human

mitogenome, not covered by delRNAs. The total frequency of X codons does not differ between del-transcribed and other sequences.

However, if X has a role in both transcription and translation, these roles may be conflicting for delRNAs. At the level of translation, X should maintain the ribosomal frame, expecting overrepresentation of codons of X in detected delRNAs. For del-transcription, codons of X, if they affect transcription, would prevent transcriptional frameshifts that characterize delRNAs. Hence, X should be underrepresented in delRNAs. Therefore, the lack of bias regarding X in detected human mitochondrial delRNAs could be due to opposite transcriptional versus translational constraints.

### Codon-Specific Reading Frame Retrieval (RFR)

In order to test whether codons of X have opposite roles in the transcription versus translation for delRNAs we apply analyses



that differentiate X codons in function of their ability to detect the programmed coding frame. RFR, the reading frame retrieval (RFR) score, estimates for each codon of X its contribution to frame detection, as previously defined (Ahmed et al., 2010; Michel and Seligmann, 2014). For example, triplets ACC and GGT contribute to frame detection in 69% of the cases. CAG, CTC, CTG, and GAG contribute to frame detection in 100% of cases. Their RFR is 69 and 100, respectively; the RFR of the remaining 14 codons is between these two extremes. In the context of delRNAs, we expect that codons with low RFR should be overrepresented and those with high RFR underrepresented if transcription constraints prevail in delRNAs over translation constraints. We expect the opposite if translation constraints prevail over transcriptional ones.

We calculate the ratio between the frequency of codons of X in delRNAs and their frequency in the rest of the mitogenome. This ratio is plotted as a function of the corresponding RFR for delRNA<sub>3-1</sub> (Figure 3). The Pearson correlation coefficient

$r = -0.602$  (one-tailed  $P = 0.0025$ ) and the nonparametric Spearman rank correlation coefficient  $r_s = -0.637$  (one tailed  $P = 0.001$ ) show avoidance of X codons with high RFR, and overrepresentation of X codons with relatively low RFR. This result is in line with a balance between transcriptional and translational effects of the natural circular code X and suggests that X might affect systematic transcriptional deletions. The pattern in Figure 3 suggests that codons with high RFR are avoided in delRNAs, to enable transcriptional frameshifting, but that codons of X with low RFR are overrepresented, enabling some relatively weak, yet existing, translational frame regulation. Hence this result indicates a dual role of the natural circular code in del-transcription and the regulation of translation of its products, the delRNAs.

We performed the same test for delRNA<sub>3-2</sub>. Results overall confirm those described above that transcription constraints decrease bias for X codons in relation to the RFR of X codons, at least for the underrepresented half of X codons. However,



they also indicate that the overrepresented half of X codons follow translational constraints (**Figure 4**). These results that show dual transcriptional and translational roles for the natural circular code are in line with the evolutionary hypothesis that nucleotide polymerization signals were hijacked for translation. They suggest that the natural circular code has its origins in the regulation of nucleotide polymerizations.

## CONCLUSION

Some examples suggest an evolutionary scenario where DNA punctuation evolved from secondary structures signaling polymerization initiation, termination, and/or processing to linear sequence motifs, which further evolved to translational signals (**Figure 1**). Presumably, primitive low-specificity structural signals evolved into a transition state where both structural and linear nucleotide sequence signals confer high specificity to the punctuation system. Presumably, signals consisting only of linear sequences (evolutionary phase C in **Figure 1**) are more derived and metabolically more efficient because these depend on shorter sequences, but with specificity intermediate between that of the presumed first and the second evolutionary phases. The vertebrate mitochondrial light strand replication origin, OL, and stem-loop hairpins formed by DNA that templates for mitochondrial tRNAs, seem to fit this evolutionary scenario.

Stem-loop hairpins, which punctuate mitochondrial RNA processing (Ojala et al., 1981), also regulate codon-amino

acid assignments (Lobanov et al., 2010), suggesting that transcriptional punctuation marks evolved into translational ones. Analyses of mitochondrial delRNA codon content show that codons belonging to the natural circular code X (which regulates ribosomal frame retrieval during translation) affect del-transcription. This first empirical evidence for effects of the natural circular code X on frameshifting deletions during nucleotide synthesis fits the complex predictions of dual functions of X by maintaining translation frame and promoting transcriptional frameshifts. Punctuation signals common to translation and transcription are compatible with the scenario that the former evolved from the latter. Similar scenarios could apply to the evolution of some of the genetic code's codon-amino acid assignments.

## AUTHOR CONTRIBUTIONS

HS designed the research and MS, NEH critically revised the MS.

## 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).

## REFERENCES

- AbouHaidar, M. G., Venkataraman, S., Golshani, A., Liu, B., and Ahmad, T. (2014). Novel coding, translation, and gene expression of a replicating covalently closed circular RNA of 220 nt. *Proc. Natl. Acad. Sci. U.S.A.* 111, 14542–14547. doi: 10.1073/pnas.1402814111
- Advani, V. M., and Dinman, J. D. (2016). Reprogramming the genetic code: The emerging role of ribosomal frameshifting in regulating cellular gene expression. *Bioessays* 38, 21–26. doi: 10.1002/bies.201500131
- Ahmed, A., Frey, G., and Michel, C. J. (2007). Frameshift signals in genes associated with the circular code. *In Silico Biol.* 7, 155–168.
- Ahmed, A., Frey, G., and Michel, C. J. (2010). Essential molecular functions associated with the circular code evolution. *J. Theor. Biol.* 264, 613–622. doi: 10.1016/j.jtbi.2010.02.006
- Akashi, H., and Gojobori, T. (2002). Metabolic efficiency and amino acid composition in the proteomes of *Escherichia coli* and *Bacillus subtilis*. *Proc. Natl. Acad. Sci. U.S.A.* 99, 3695–3700. doi: 10.1073/pnas.062526999
- Altschul, S. F., Madden, T. L., Schaeffer, A. A., Zhang, J., Zhang, Z., Miller, W., et al. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25, 3389–3402. doi: 10.1093/nar/25.17.3389
- Arquès, D. G., and Michel, C. J. (1996). A complementary circular code in the protein coding genes. *J. Theor. Biol.* 182, 45–58. doi: 10.1006/jtbi.1996.0142
- Atkins, J. F., Loughran, G., Bhatt, P. R., Fith, A. E., and Baranov, P. V. (2016). Ribosomal frameshifting and transcriptional slippage: from genetic steganography and cryptography to adventitious use. *Nucleic Acids Res.* 44, 7007–7078. doi: 10.1093/nar/gkw530
- Baranov, P. V., Venin, M., and Provan, G. (2009). Codon size reduction as the origin of the triplet genetic code. *PLoS ONE* 4:e5708. doi: 10.1371/journal.pone.0005708
- Bar-Yaacov, D., Levin, A. G., Richards, A. L., Hachen, N., Rebolledo Jaramillo, B., Nebrutenko, A., et al. (2013). RNA-DNA differences in human mitochondria restore ancestral form of 16S ribosomal RNA. *Genome Res.* 23, 1789–1796. doi: 10.1101/gr.161265.113
- Branciamore, S., and Di Giulio, M. (2011). The presence in tRNA molecule sequences of the double hairpin, an evolutionary stage through which the origin of this molecule is thought to have passed. *J. Mol. Evol.* 72, 352–363. doi: 10.1007/s00239-011-9440-9
- Brocchieri, L., and Karlin, S. (2005). Protein length in eukaryotic and prokaryotic proteomes. *Nucleic Acids Res.* 33, 3390–3400. doi: 10.1093/nar/gki615
- Brown, A., Shao, S., Murray, J., Hegde, R. S., and Ramakrishnan, V. (2015). Structural basis for stop codon recognition in eukaryotes. *Nature* 524, 496–496. doi: 10.1038/nature14896
- Byrne, D., Grzela, A., Larigue, A., Audic, S., Chevnesse, S., Encinas, S., et al. (2009). The polyadenylation site of *Mimivirus* transcripts obeys a stringent ‘hairpin rule’. *Genome Res.* 19, 376–384. doi: 10.1101/gr.091561.109
- Chen, C., and Bundschuh, R. (2012). Systematic investigation of insertional and deletional RNA-DNA differences in the human transcriptome. *BMC Genomics* 13:616. doi: 10.1186/1471-2164-13-616
- Chen, W.-H., Lu, G., and Bork, P. (2016). Energy efficiency trade-offs drive nucleotide usage in transcribed regions. *Nat. Commun.* 7:11334. doi: 10.1038/ncomms11334
- Chen, Y., McClure, R. A., and Kelleher, N. L. (2016). Screening for expressed nonribosomal peptide synthetases and polyketide synthetases using LC-MS/MS-based proteomics. *Methods Mol. Biol.* 1401, 135–147. doi: 10.1007/978-1-4939-3375-4\_9
- Claverie, J. M., and Abergel, C. (2009). *Mimivirus* and its viroplasm. *Annu. Rev. Genet.* 43, 49–66. doi: 10.1146/annurev-genet-102108-134255
- Clayton, D. A. (1992). Transcription and replication of animal mitochondrial DNAs. *Int. Rev. Cytol.* 141, 217–232. doi: 10.1016/S0074-7696(08)62067-7
- Clayton, D. A. (2000). Transcription and replication of mitochondrial DNA. *Hum. Reprod.* 2, 11–17. doi: 10.1093/humrep/15.suppl\_2.11

- Clayton, D. A. (2003). Mitochondrial DNA replication: what we know. *IUBMB Life* 55, 213–217. doi: 10.1080/1521654031000134824
- Crick, F. H., Griffith, J. S., and Orgel, L. E. (1957). Codes without commas. *Proc. Natl. Acad. Sci. U.S.A.* 43, 416–421. doi: 10.1073/pnas.43.5.416
- Desjardins, P., and Morais, R. (1991). Nucleotide sequence and evolution of coding and noncoding regions of a quail mitochondrial genome. *J. Mol. Evol.* 32, 153–161. doi: 10.1007/BF02515387
- Di Giulio, M. (2009). Formal proof that the split genes of tRNAs of *Nanoarchaeum equitans* are an ancestral character. *J. Mol. Evol.* 69, 505–511. doi: 10.1007/s00239-009-9280-z
- Di Giulio, M. (2012a). The origin of the tRNA molecule: Independent data favor a specific model of its evolution. *Biochimie* 94, 1464–1466. doi: 10.1016/j.biochi.2012.01.014
- Di Giulio, M. (2012b). The 'recently' split transfer RNA genes may be close to merging the two halves of the tRNA rather than having just separated them. *J. Theor. Biol.* 310, 1–2. doi: 10.1016/j.jtbi.2012.06.022
- Di Giulio, M. (2013). A polyphyletic model for the origin of tRNAs has more support than a monophyletic model. *J. Theor. Biol.* 318, 124–128. doi: 10.1016/j.jtbi.2012.11.012
- Di Giulio, M. (2015). A model for the origin of the first mRNAs. *J. Mol. Evol.* 81, 10–17. doi: 10.1007/s00239-015-9691-y
- El Soufi, K., and Michel, C. J. (2014). Circular code motifs in the ribosome decoding center. *Comput. Biol. Chem.* 52, 9–17. doi: 10.1016/j.compbiolchem.2014.08.001
- El Soufi, K., and Michel, C. J. (2015). Circular code motifs near the ribosome decoding center. *Comput. Biol. Chem.* 59(Pt A), 158–176. doi: 10.1016/j.compbiolchem.2015.07.015
- Elzanowski, A., and Ostell, J. (2013). *The Genetic Codes*. Available at: <http://www.ncbi.nlm.nih.gov/Taxonomy/taxonomyhome.html/index.cgi?chapter=tgencodes#SG2> [accessed September 2016].
- Fimmel, E., Michel, C. J., and Struengmann, L. (2016). n-Nucleotide circular codes in graph theory. *Philos. Trans. A Math. Phys. Eng. Sci.* 374:20150058. doi: 10.1098/rsta.2015.0058
- Fimmel, E., and Strümgmann, L. (2015). On the hierarchy of trinucleotide n-circular codes and their corresponding amino acids. *J. Theor. Biol.* 364, 113–120. doi: 10.1016/j.jtbi.2014.09.011
- Fimmel, E., and Strümgmann, L. (2016). Codon distribution in error-detecting circular codes. *Life* 6:e14. doi: 10.3390/life6010014
- Fonfara, I., Le Rhun, A., Chylinski, K., Makarova, K. S., Lécrivain, A. L., Bzdrenga, J., et al. (2014). Phylogeny of Cas9 determines functional exchangeability of dual-RNA and Cas9 among orthologous type II CRISPR-Cas systems. *Nucleic Acids Res.* 42, 2577–2590. doi: 10.1093/nar/gkt1074
- Fusté, J. M., Wanrooij, S., Jemt, E., Granycome, C. E., Cluett, T. J., Shi, Y., et al. (2010). Mitochondrial RNA polymerase is needed for activation of the origin of light-strand DNA replication. *Mol. Cell* 37, 67–78. doi: 10.1016/j.molcel.2009.12.021
- Garzon, R., Volinia, S., Papaioannou, D., Nicolet, D., Kohlschmidt, J., Yan, P. S., et al. (2014). Expression and prognostic impact of lncRNAs in acute myeloid leukemia. *Proc. Natl. Acad. Sci. U.S.A.* 111, 18679–18684. doi: 10.1073/pnas.1422050112
- Gonzalez, D. L., Giannerini, S., and Rosa, R. (2011). Circular codes revisited: a statistical approach. *J. Theor. Biol.* 275, 21–28. doi: 10.1016/j.jtbi.2011.01.028
- He, Y., Jones, J., Armstrong, M., Lamberti, F., and Moens, M. (2005). The mitochondrial genome of *Xiphinema americanum* sensu stricto (Nematoda: Enopleia): considerable economization in the length and structural features of encoded genes. *J. Mol. Evol.* 61, 819–833. doi: 10.1007/s00239-005-0102-7
- Heizer, E. M. Jr., Raymer, M. L., and Krane, D. E. (2011). Amino acid biosynthetic cost and protein conservation. *J. Mol. Evol.* 72, 466–473. doi: 10.1007/s00239-011-9445-4
- Himeno, H., Kurita, D., and Muto, A. (2014). tmRNA-mediated trans-translation as the major ribosome rescue system in a bacterial cell. *Front. Genet.* 5:66. doi: 10.3389/fgene.2014.00066
- Hirose, Y., Ikeda, K. T., Noro, E., Hiraoka, K., Tomita, M., and Kanai, A. (2015). Precise mapping and dynamics of tRNA-derived fragments (tRFs) in the development of *Triops cancriformis* (tadpole shrimp). *BMC Genetics* 16:83. doi: 10.1186/s12863-015-0245-5
- Hixson, J. E., Wong, T. W., and Clayton, D. A. (1986). Both the conserved stem-loop and divergent 5'-flanking sequences are required for initiation at the human mitochondrial origin of light-strand replication. *J. Biol. Chem.* 261, 2384–2390.
- Itzkovitz, S., and Alon, U. (2007). The genetic code is nearly optimal for allowing additional information within protein-coding sequences. *Genome Res.* 17, 405–412. doi: 10.1101/gr.5987307
- Janssen, B. D., and Hayes, C. S. (2012). The tmRNA ribosome-rescue system. *Adv. Protein Chem. Struct. Biol.* 86, 151–191. doi: 10.1016/B978-0-12-386497-0.00005-0
- Jestin, J. L., and Kempf, A. (1997). Chain termination codons and polymerase-induced frameshift mutations. *FEBS Lett.* 419, 153–156. doi: 10.1016/S0014-5793(97)01422-1
- Jühling, F., Pütz, J., Florentz, C., and Stadler, P. F. (2012). Armless mitochondrial tRNAs in Enopleia (Nematoda). *RNA Biol.* 9, 1161–1166. doi: 10.4161/rna.21630
- Ketteler, R. (2012). On programmed ribosomal frameshifting: the alternative proteomes. *Front. Genet.* 3:242. doi: 10.3389/fgene.2012.00242
- Klobutcher, L. A., and Farabaugh, P. J. (2002). Shifty ciliates: frequent programmed translational frameshifting in euplotids. *Cell* 111, 763–766. doi: 10.1016/S0092-8674(02)01138-8
- Krick, T., Verstraete, N., Alonso, L. G., Shub, D. A., Ferreira, D. U., Shub, M., et al. (2014). Amino acid metabolism conflicts with protein diversity. *Mol. Biol. Evol.* 31, 2905–2912. doi: 10.1093/molbev/msu228
- Krizek, M., and Krizek, P. (2012). Why has nature invented three stop codons of DNA and only one start codon? *J. Theor. Biol.* 304, 183–187. doi: 10.1016/j.jtbi.2012.03.026
- Lacan, J., and Michel, C. J. (2001). Analysis of a circular code model. *J. Theor. Biol.* 213, 159–170. doi: 10.1006/jtbi.2001.2416
- Li, M., Wang, I. X., Li, Y., Bruzel, A., Richards, A. L., Toung, J. M., et al. (2011). Widespread RNA and DNA sequence differences in the human transcriptome. *Science* 333, 53–58. doi: 10.1126/science.1207018
- Lobanov, A. V., Turanov, A. A., Hatfield, D. L., and Gladyshev, V. N. (2010). Dual functions of codons in the genetic code. *Crit. Rev. Biochem. Mol. Biol.* 45, 257–265. doi: 10.3109/10409231003786094
- Macé, K., and Gillet, R. (2016). Origins of tmRNA: the missing link in the birth of protein synthesis? *Nucleic Acids Res.* 44, 8041–8051. doi: 10.1093/nar/gkw693
- Macey, J. R., Larson, A., Ananjeva, N. B., Fang, Z., and Papenfuss, T. J. (1997). Two novel gene orders and the role of light-strand replication in rearrangement of the vertebrate mitochondrial genome. *Mol. Biol. Evol.* 14, 91–104. doi: 10.1093/oxfordjournals.molbev.a025706
- Maizels, N., and Weiner, A. M. (1994). Phylogeny from function: evidence from the molecular fossil record that tRNA originated in replication, not translation. *Proc. Natl. Acad. Sci. U.S.A.* 91, 6729–6734. doi: 10.1073/pnas.91.15.6729
- Michel, C. J. (2012). Circular code motifs in transfer and 16S ribosomal RNAs: a possible translation code in genes. *Comput. Biol. Chem.* 37, 24–37. doi: 10.1016/j.compbiolchem.2011.10.002
- Michel, C. J. (2014). A genetic scale of reading frame coding. *J. Theor. Biol.* 355, 83–94. doi: 10.1016/j.jtbi.2014.03.029
- Michel, C. J., and Seligmann, H. (2014). Bijective transformation circular codes and nucleotide exchanging RNA transcription. *Biosystems* 118, 39–50. doi: 10.1016/j.biosystems.2014.02.002
- Ojala, D., Montoya, J., and Attardi, G. (1981). tRNA punctuation model of RNA processing in human mitochondria. *Nature* 290, 470–474. doi: 10.1038/290470a0
- Popov, O., Segal, D. M., and Trifonov, E. N. (1996). Linguistic complexity of protein sequences as compared to texts of human languages. *Biosystems* 38, 65–74. doi: 10.1016/0303-2647(95)01568-X
- Raiford, D. W., Heizer, E. M. Jr., and Miller, R. V. (2012). Metabolic and translational efficiency in microbial organisms. *J. Mol. Evol.* 74, 206–216. doi: 10.1007/s00239-012-9500-9
- Reyes, A., Yang, M. Y., Bowmaker, M., and Holt, I. J. (2005). Bidirectional replication initiates at sites throughout the mitochondrial genome of birds. *J. Biol. Chem.* 280, 3242–3250. doi: 10.1074/jbc.M411916200
- Root-Bernstein, M., and Root-Bernstein, R. (2015). The ribosome as a missing link in the evolution of life. *J. Theor. Biol.* 367, 130–158. doi: 10.1016/j.jtbi.2014.11.025
- Root-Bernstein, R., and Root-Bernstein, M. (2016). 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.* 397, 115–127. doi: 10.1016/j.jtbi.2016.02.030

- Roy, H., and Ibba, M. (2010). Bridging the gap between ribosomal and nonribosomal protein synthesis. *Proc. Natl. Acad. Sci. U.S.A.* 107, 14517–14518. doi: 10.1073/pnas.1009939107
- Schaub, R. E., Poole, S. J., Garza-Sánchez, F., Benbow, S., and Hayes, C. S. (2012). Proteobacterial ArfA peptides are synthesized from non-stop messenger RNAs. *J. Biol. Chem.* 287, 29765–29775. doi: 10.1074/jbc.M112.374074
- Seligmann, H. (2003). Cost minimization of amino acid usage. *J. Mol. Evol.* 56, 151–161. doi: 10.1007/s00239-002-2388-z
- Seligmann, H. (2007). Cost minimization of ribosomal frameshifts. *J. Theor. Biol.* 249, 162–167. doi: 10.1016/j.jtbi.2007.07.007
- Seligmann, H. (2008). 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.* 379, 188–199. doi: 10.1016/j.jmb.2008.03.066
- Seligmann, H. (2010a). 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* 99, 85–93. doi: 10.1016/j.biosystems.2009.09.003
- Seligmann, H. (2010b). The ambush hypothesis at the whole-organism level: off frame, 'hidden' stops in vertebrate mitochondrial genes increase developmental stability. *Comput. Biol. Chem.* 34, 80–85. doi: 10.1016/j.compbiolchem.2010.03.001
- Seligmann, H. (2011). "Mutation patterns due to converging mitochondrial replication and transcription increase lifespan, and cause growth rate-longevity tradeoffs," in *DNA Replication-Current Advances*, Vol. 6, ed. H. Seligmann (Rijeka: InTech), 151–180. doi: 10.5772/24319
- Seligmann, H. (2012a). Coding constraints modulate chemically spontaneous mutational replication gradients in mitochondrial genomes. *Curr. Genom.* 13, 37–54. doi: 10.2174/138920212799034802
- Seligmann, H. (2012b). Positive and negative cognate amino acid bias affects compositions of aminoacyl-tRNA synthetases and reflects functional constraints on protein structure. *BIO* 2, 11–26. doi: 10.5618/bio.2012.v2.n1.2
- Seligmann, H. (2013). Pocketknife tRNA hypothesis: anticodons in mammal mitochondrial tRNA side-arm loops translate proteins? *Biosystems* 113, 165–176. doi: 10.1016/j.biosystems.2013.07.004
- Seligmann, H. (2014). Putative anticodons in mitochondrial tRNA sidearm loops: Pocketknife tRNAs? *J. Theor. Biol.* 340, 155–163. doi: 10.1016/j.jtbi.2013.08.030
- Seligmann, H. (2015a). Codon expansion and systematic transcriptional deletions produce tetra-, pentacoded mitochondrial peptides. *J. Theor. Biol.* 387, 154–165. doi: 10.1016/j.jtbi.2015.09.030
- Seligmann, H. (2015b). Phylogeny of genetic codes and punctuation codes within genetic codes. *Biosystems* 129, 36–43. doi: 10.1016/j.biosystems.2015.01.003
- Seligmann, H. (2016). Systematically frameshifting by deletion of every 4th or 4th and 5th nucleotides during mitochondrial transcription: RNA self-hybridization regulates delRNA expression. *Biosystems* 14, 43–51. doi: 10.1016/j.biosystems.2016.03.009
- Seligmann, H., and Krishnan, N. M. (2006). Mitochondrial replication origin stability and propensity of adjacent tRNA genes to form putative replication origins increase developmental stability in lizards. *J. Exp. Zool. B* 306B, 433–449. doi: 10.1002/jez.b.21095
- Seligmann, H., Krishnan, N. M., and Rao, B. J. (2006a). Mitochondrial tRNA sequences as unusual replication origins: pathogenic implications for *Homo sapiens*. *J. Theor. Biol.* 243, 375–385.
- Seligmann, H., Krishnan, N. M., and Rao, B. J. (2006b). Possible multiple origins of replication in primate mitochondria: alternative role of tRNA sequences. *J. Theor. Biol.* 241, 321–332.
- Seligmann, H., and Labra, A. (2014). The relation between hairpin formation by mitochondrial WANCY tRNAs and the occurrence of the light strand replication origin in Lepidosauria. *Gene* 542, 248–257. doi: 10.1016/j.gene.2014.02.021
- Seligmann, H., and Pollock, D. D. (2004). The ambush hypothesis: hidden stop codons prevent off-frame gene reading. *DNA Cell Biol.* 23, 701–705. doi: 10.1089/dna.2004.23.701
- Seligmann, H., and Raoult, D. (2016). 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.* 31, 1–8. doi: 10.1016/j.mib.2015.11.004
- Stuckey, R., Garcia-Rodriguez, N., Aguilera, A., and Wellinger, R. E. (2015). Role for RNA:DNA hybrids in origin-independent replication priming in a eukaryotic system. *Proc. Natl. Acad. Sci. U.S.A.* 112, 5779–5784. doi: 10.1073/pnas.1501769112
- Swart, E. C., Serra, V., Petroni, G., and Nowacki, M. (2016). Genetic codes with no dedicated stop codon: context-dependent translation termination. *Cell* 166, 691–702. doi: 10.1016/j.cell.2016.06.020
- Trifonov, E. N. (2000). Consensus temporal order of amino acids and evolution of the triplet code. *Gene* 261, 139–151. doi: 10.1016/S0378-1119(00)00476-5
- Trifonov, E. N. (2004). The triplet code from first principles. *J. Biomol. Struct. Dyn.* 22, 1–11. doi: 10.1080/07391102.2004.10506975
- Trifonov, E. N. (2009). The origin of the genetic code and of the earliest oligopeptides. *Res. Microbiol.* 160, 481–486. doi: 10.1016/j.resmic.2009.05.004
- Tse, H., Cai, J. J., Tsoi, H.-W., Lam, E. P., and Yuen, K. Y. (2010). Natural selection retains overrepresented out-of-frame stop codons against frameshift peptides in prokaryotes. *BMC Genomics* 11:491. doi: 10.1186/1471-2164-11-491
- Wang, I. X., Core, L. J., Kwak, H., Brady, L., Bruzel, A., McDaniel, L., et al. (2014). RNA-DNA differences are generated within seconds after RNA exits polymerase II. *Cell Rep.* 6, 906–915. doi: 10.1016/j.celrep.2014.01.037
- Wanrooij, S., Miralles Fusté, J., Stewart, J. B., Wanrooij, P. H., Samuelsson, T., Larsson, N. G., et al. (2012). In vivo mutagenesis reveals that OriL is essential for mitochondrial DNA replication. *EMBO Rep.* 13, 1130–1137. doi: 10.1038/embor.2012.161
- Warringer, J., and Blomberg, A. (2006). Evolutionary constraints on yeast protein size. *BMC Evol. Biol.* 6:61. doi: 10.1186/1471-2148-6-61
- Wende, S., Platzer, E. G., Jühling, F., Pütz, J., Florentz, C., Stadler, P. F., et al. (2014). Biological evidence for the world's smallest tRNAs. *Biochimie* 100, 151–158. doi: 10.1016/j.biochi.2013.07.034
- Yu, C.-H., Liao, J.-Y., Zhou, H., and Qu, L.-H. (2008). The rat mitochondrial Ori L encodes a novel small RNA resembling and ancestral tRNA. *Biochem. Biophys. Res. Commun.* 372, 634–638. doi: 10.1016/j.bbrc.2008.05.092
- Zheng, K.-W., Wu, R.-Y., He, Y.-D., Xiao, S., Zhang, J.-Y., Liu, J.-Q., et al. (2014). A competitive formation of DNA:RNA hybrid G-quadruplex is responsible to the mitochondrial transcription termination at the DNA replication priming site. *Nucleic Acids Res.* 42, 10832–10844. doi: 10.1093/nar/gku764

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 El Houmami and Seligmann. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.