Journal of Theoretical Biology 324 (2013) 1–20

Contents lists available at SciVerse ScienceDirect

Journal of Theoretical Biology

journal homepage: www.elsevier.com/locate/yjtbi

Systematic asymmetric nucleotide exchanges produce human mitochondrial RNAs cryptically encoding for overlapping protein coding genes

Hervé Seligmann [a,b,*]

[a] National Natural History Museum Collections, The Hebrew University of Jerusalem, 91904 Jerusalem, Israel
[b] Department of Life Sciences, Ben Gurion University, 84105 Beer Sheva, Israel

HIGHLIGHTS

- Assuming systematic nucleotide exchanges (i.e., A→C→G→A) reveals unknown genes.
- RNA transcripts correspond to different types of systematic exchanges.
- Nucleotide misincorporation rates increase exchange transcribed RNA sizes.
- Nucleotide exchange transcription expands gene density by an order of magnitude.

ARTICLE INFO

Article history:
Received 8 November 2012
Received in revised form
26 January 2013
Accepted 28 January 2013
Available online 8 February 2013

Keywords:
RNA transcription
RNA editing
DNA polymerase gamma
Circular code
Replicational mutation gradients

ABSTRACT

GenBank's EST database includes RNAs matching exactly human mitochondrial sequences assuming systematic asymmetric nucleotide exchange-transcription along exchange rules: A→G→C→U/T→A (12 ESTs), A→U/T→C→G→A (4 ESTs), C→G→U/T→C (3 ESTs), and A→C→G→U/T→A (1 EST), no RNAs correspond to other potential asymmetric exchange rules. Hypothetical polypeptides translated from nucleotide-exchanged human mitochondrial protein coding genes align with numerous GenBank proteins, predicted secondary structures resemble their putative GenBank homologue's. Two independent methods designed to detect overlapping genes (one based on nucleotide contents analyses in relation to replicative deamination gradients at third codon positions, and circular code analyses of codon contents based on frame redundancy), confirm nucleotide-exchange-encrypted overlapping genes. Methods converge on which genes are most probably active, and which not, and this for the various exchange rules. Mean EST lengths produced by different nucleotide exchanges are proportional to (a) extents that various bioinformatics analyses confirm the protein coding status of putative overlapping genes; (b) known kinetic chemistry parameters of the corresponding nucleotide substitutions by the human mitochondrial DNA polymerase gamma (nucleotide DNA misinsertion rates); (c) stop codon densities in predicted overlapping genes (stop codon readthrough and exchanging polymerization regulate gene expression by counterbalancing each other). Numerous rarely expressed proteins seem encoded within regular mitochondrial genes through asymmetric nucleotide exchange, avoiding lengthening genomes. Intersecting evidence between several independent approaches confirms the working hypothesis status of gene encryption by systematic nucleotide exchanges.

1. Introduction

The human transcriptome is much more complex than previously believed (Mercer et al., 2011a). For example, RNAs with sequences matching the 3′-to-5′ sequence of parts of the mitochondrial genome exist (Seligmann, 2012a), suggesting 'unorthodox' transcription. Such inverted RNA has the ability to form a triplex structure with regular complementary strands if it is mainly composed of purines, or mainly of pyrimidines. Indeed, DNA and RNA have the ability to form triplex structures through Hoogsteen pairings (Hoogsteen, 1963). Such structures, formed by DNA:RNA hybridization, have been observed for mitochondrial genomes (Clayton, 2000), in association with DNA and RNA synthesis (Annex and Williams, 1990; Rocher et al., 2002; Takamatsu et al., 2002). At physiological pH, the third strand forms antiparallel Hoogsteen base pairings, located in the major groove of the regular DNA double helix. Antiparallel Hoogsteen pairings require that the third strand's sequence is the inverse (3′-to-5′) of one of the (5′-to-3′) sequences forming the double helix. The third strand pairs with that specific strand. The

* Corresponding author. Tel.: +9272 6585 874.
E-mail address: varanuseremius@gmail.com

0022-5193/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.jtbi.2013.01.024

occasional observation of triplex helices probably implies that 3′-to-5′ inverted sequences (of DNA or RNA) are polymerized. Indeed, RNA corresponding to inverted sequences of the human mitochondrial genome has been found, and inverted sequences seem to code for protein coding genes encrypted in the inverted sequence, increasing the number of genes without lengthening the genome (Seligmann, 2012a).

Recent evidence shows that regular 5′-to-3′ polymerized RNA can be elongated also by 3′-to-5′ nucleotide addition (Jackman et al., 2012), which lends greater plausibility to the hypothesis that also 3′-to-5′ inverted RNA can be produced. However, RNA matching the 3′-to-5′ inverted sequence of DNA does not necessarily imply 3′-to-5′ RNA polymerization. Lets assume a new type of 5′-to-3′ RNA transcription, which systematically exchanges A (at the DNA level) by U at the RNA level, and T at the DNA level by A at the RNA level (A↔U/T), and where C would be systematically exchanged by G and G by C (C↔G). The sequence obtained by this hypothetical symmetric nucleotide exchanging RNA transcription A↔U/T+C↔G of sequence X would be identical to the 3′-to-5′ inverted sequence of the antisense (complementary) sequence of X. Hence combined systematic exchanges of A↔U/T+ C↔G during RNA polymerization and/or editing would produce inverted 3′-to-5′ RNA with the ability to form antiparallel Hoogsteen pairings, resulting in triplex structures.

The hypothesis that systematic nucleotide exchanges occur (assuming that the resulting RNA includes all four regular nucleotides) implies that 9 symmetric exchanges could exist (involving only two nucleotides: A↔C, A↔G, A↔U, C↔G, C↔U, G↔U/T; involving all four nucleotides: A↔C+G↔U, A↔G+C↔U, and A↔U+C↔G, the hypothesis on symmetric exchanges is to be examined elsewhere). There are also 14 potential asymmetric exchanges, 8 involving three nucleotides (A→C→G→A, A→C→U/T→A, A→G→C→A, A→G→U/T→A, A→U/T→C→A, A→U/T→G→A, C→G→U/T→C, C→U/T→G→C) and 6 involving all four nucleotides (A→C→G→U/T→A, A→C→U/T→G→A, A→G→C→U/T→A, A→G→T→C→A, A→U/T→C→G→A, A→U/T→G→C→A). These asymmetric nucleotide exchanges, their effects on RNA and amino acid coding properties are presented in Table 1.

Table 1
The 14 different RNA sequences produced by asymmetric nucleotide exchanging transcription of a single DNA sequence (ACGT) according to the 14 types of symmetric nucleotide exchange rules. Similarity indicates numbers of nucleotides remaining to regular transcription (nucleotides indicated in bold), followed by numbers of nucleotide exchanges that are transitions (A↔G or C↔U/T, nucleotides indicated in italics). Numbers of nucleotides complementing the original sequence are also indicated (complementary nucleotides are underlined). The amino acid coded by the three first nucleotides according to the vertebrate mitochondrial genetic code is also indicated.

Exchange rule	ACGT	Similarity	Compl.	Amino acid exchange
Regular	ACGU	4–0	0	Thr→Thr
A→C→G→A	C*G*A**U**	1–1	1	Thr→→Arg
A→C→U→A	C*U***G**A*	1–1	1	Thr→Leu
A→G→C→A	GA**C**U	1–1	1	Thr→Asp
A→G→U→A	*G***C**U*A*	1–1	1	Thr→Ala
A→U→C→A	U*A***G***C*	1–1	1	Thr→Stop
A→U→G→A	U**C***A*G	1–1	1	Thr→Ser
C→G→U→C	**A**G*U*C	1–1	1	Thr→Ser
C→U→G→C	**A***U*CG	1–1	1	Thr→Ile
A→C→G→U→A	CGUA	0–0	2	Thr→Arg
A→C→U→G→A	C*UA*G	0–2	0	Thr→Leu
A→G→C→U→A	*GU*C*A*	0–2	2	Thr→Val
A→G→U→C→A	GAUC	0–2	0	Thr→Asp
A→U→C→G→A	UGAC	0–2	2	Thr→Trp
A→U→G→C→A	UACG	0–0	2	Thr→Tyr

Here I examine the hypothesis that systematic asymmetric exchange transcription occasionally occurs, and that such RNAs might be coding for overlapping genes in the human mitochondrial genome, for the 14 asymmetric exchange rules described above and in Table 1.

Several recently described 'unorthodox' mechanisms seem to regulate the expression of numerous mitochondrial overlapping genes: (a) translational activity by suppressor (antisense) tRNAs, tRNAs with anticodons matching stop codons (Seligmann, 2010a), coevolves with mitochondrial overlapping off frame protein coding sequences (Seligmann, 2011a; Faure et al., 2011; Seligmann, 2012b, 2012c), enabling to express coding information stored in off frame sequences despite the off frame presence of stop codons (Seligmann and Pollock, 2003, 2004a, 2004b; Seligmann and Pollock, 2003a, 2004), (b) production of 3′-to-5′ inverted RNA enables to unleash the coding information stored in the inverted sequence of the gene (Seligmann, 2012a, 2012c), (c) assuming quadruplet codons, also called tetracodons, several mitochondrial regions apparently include additional protein coding genes, whose number coevolves with translational activity by antisense tRNAs with expanded anticodons (Seligmann, 2012d; evidence for antisense tRNA activity independently of alternative, parallel coding systems exists in Seligmann (2010a, 2010c, 2011b)). Hence genetic systems might also use nucleotide exchanging RNA polymerization (whether through symmetric or asymmetric exchanges) to increase the coding potential of DNA.

The analyses presented below include two main sections, each subdivided into different parts: (1) a search, using GenBank's Blastn tool (Altschul et al., 1997, 2005), for RNAs in GenBank's database of human ESTs (expressed sequence tags) corresponding to each of the 14 potential nucleotide exchange rules presented above; and (2) a search, using Blastp, for proteins existing in GenBank and aligning with hypothetical polypeptides translated from RNA if asymmetric nucleotide exchange RNA polymerization occurred.

Each section includes parts designed to evaluate that alignments are not false positives. Additional analyses show that data for abundances of RNAs produced by the various exchange rules match quantitative estimations of overlap coding in the same exchanged RNAs. Convergence of these independent types of evidences suggests that exchange transcription, and associated protein coding, are natural phenomena that remained undetected until now. The coding structure of these genes indicate functionality, hence exchange-transcription-encoded genes might have adaptive components. The results ask for a total reappraisal of our concept of protein coding genes and their detection.

2. Results and discussion

2.1. RNAs produced by systematic asymmetric exchange transcription

The first step is to search for 'hard' evidence that RNA produced by systematic nucleotide asymmetric exchange exists. For that purpose, the complete human mitochondrial genome was recoded *in silico* along each of the 14 rules of asymmetric exchanges described above. For example, A→C→G→A means that all A's were replaced by Cs, all Cs by Gs, and all Gs by As. Thymidines remain unchanged according to this exchange rule. Exchanging rules examined here keep the number of nucleotide species constant, eight exchange rules imply three nucleotide species, and six exchange rules imply all four nucleotide species. Table 2 presents all the RNAs detected by Blastn and matching these exchanged human mitochondrial genome sequences within GenBank's EST database. Blastn returns hits for a total of 20 ESTs,