

# 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](mailto:book.department@intechopen.com)

Numbers displayed above are based on latest data collected.  
For more information visit [www.intechopen.com](http://www.intechopen.com)



---

# True Mitochondrial tRNA Punctuation and Initiation Using Overlapping Stop and Start Codons at Specific and Conserved Positions

---

Eric Faure and Roxane Barthélémy

Additional information is available at the end of the chapter

<http://dx.doi.org/10.5772/intechopen.75555>

---

## Abstract

In all the taxa and genomic systems, numerous *trn* genes (specifying tRNA) exhibit at specific conserved positions nucleotide triplets corresponding to stop codons (TAG/TAA). Similarly, relatively high frequencies of start codons (ATG/ATA) occur in fungi/metazoan mitochondrial-*trn* genes. The last nucleotide of these triplets is the first involved in the 5'-D- or 5'-T-stem, respectively. Their frequencies are tRNA species dependent. The products of these genes which bear one or two types of these codons are called ss-tRNAs (for stop/start). Metazoan mt-genomes are generally very compact, and many same strand overlapping sequences may simultaneously code for tRNAs and mRNAs. However, this study suggests that overlaps are not a direct mechanism to substantially reduce genome size. For protein-encoding genes, occulting possible overlaps, there are only alternative start codons and/or truncated stop codons, but the first putative in-frame standard initiation codon or complete stop codon is in the upstream or downstream overlapping *ss-trn* sequences, respectively. Even if, to date, experimental data are missing, stress signals might regulate producing extended or not proteins. Finally, possible implications of tRNA/mRNA hybrid molecules in the "RNA world" to "RNA/protein world" transition will be discussed.

**Keywords:** mitochondrion, tRNA origin, start codon, stop codon, overlap, origin of life, RNA world

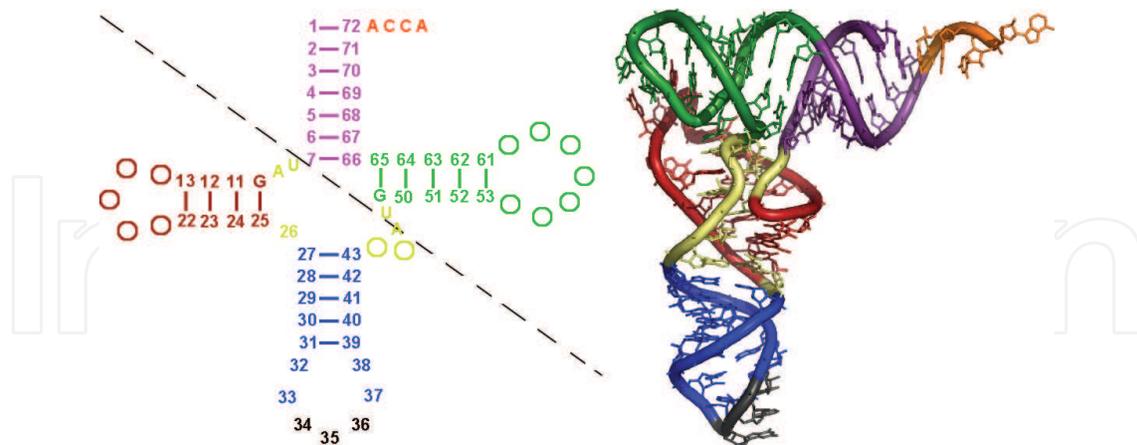
---

## 1. Introduction

Transfer RNAs are key partners in the ribosome-translation machinery. Generally, they are composed of c.70–90 nucleotides (nts). Moreover, they are the most abundant nucleic acid

---

species, constituting up to 10% of all cellular RNAs [1]. Therewith, the number of tRNA molecules is, e.g., about  $2 \times 10^5$  in *Escherichia coli* and  $3 \times 10^6$  in yeast cell [2]. Due to their anticodon, they read genetic information on mRNAs and deliver codon specified amino acids attached to their distal 3'-extremity for peptide bond synthesis on the ribosome. In this sense, tRNA is a key molecule which makes it possible to pass from a covalent bond between a RNA and an amino acid (fossil trace of the RNA world to the RNA/protein world transition) to peptide bonds (RNA/protein world). Genes specifying tRNAs (noted *trn*) are present in prokaryotic and nuclear genomes and in most of the DNAs of organelles (chloroplasts and mitochondria). Usually, tRNAs have a characteristic canonical cloverleaf secondary structure made up of the aminoacyl acceptor-stem and the D-arms (as it contains dihydrouridine), anticodon-arms, and T-arms (for the sequence TΨC where Ψ is pseudouridine), the hairpins, or "arms" consisting of a stem (helicoidal region in 3D) ending in a loop (**Figure 1**). The lengths of each arm, as well as the loop "diameter," vary from the tRNA type and from species to species. Furthermore, deduced *trn* sequences and even sequenced mature tRNAs exhibit reduced D-arms or T-arms or even lacking at least one of them, and in the extreme situation such as in *Enoplea* (nematodes) mitochondrial (mt)-*trn* genes are totally armless [3]. However, around 90% of the mt-tRNAs fold into the canonical cloverleaf structure [4]. In all the genetic systems, the tRNAs can carry a myriad of idiosyncratic posttranscriptional chemical modifications (e.g., <http://modomics.genesilico.pl/>; <http://www.genesilico.pl/rnapathwaysdb/>), and the total number of modified nts is nearly 120 [1]. Moreover, tRNAs become functional by postprocessing addition of the 3'-terminal CCA sequence. Modifications can also be necessary to ensure correct folding [4]. The tRNA folds into an L-shaped 3D structure in which two helical domains (acceptor/T and D/anticodon) are perpendicularly arranged. This particular juxtaposition of the two functional centers, the anticodon and the acceptor terminus, is essential for tRNA function. The



**Figure 1.** Typical cloverleaf secondary structure of a metazoan mt-ss-tRNA (left) with 3D image of an L-shaped tRNA (right). In 2D structure, the standard numbering was applied [5]. The first two nucleotides of the variable region and those of the D-loops and T-loops were represented by circles. The diagonal dashed line indicates the approximate separation between the "top half" and the "cherry-bob"/"bottom half". Nucleotide types were given for UAG10 and AUG49 triplets, the discriminator base (which is preferentially an A), and the CCA tail at the 3'-end. Short lines connect nucleotides forming Watson-Crick pairing within stems. Coloring: acceptor-stem in purple, D-arm in red, anticodon-arm in blue with the anticodon in black, T-arm in green, and CCA tail in orange. The yellow segments represented respectively in descending order of size, the variable region (connector 2), the connector 1 and the nt 26. 3D structure reproduced with the kind permission of Prof. N.R. Voss (Roosevelt University, Ill.) [https://commons.wikimedia.org/wiki/File:3d\\_tRNA.png](https://commons.wikimedia.org/wiki/File:3d_tRNA.png).

two domains are linked together by connector regions, one between the acceptor- and D-stems (connector 1) and the second between the anticodon- and T-stems (connector 2 which has a variable length from 0 to 21 nts and is also named variable region hereafter V-R) [5].

The ribosome allows the best possible spatial arrangement of the various partners and ensures catalysis, but the adaptor molecule which acts as a link between codes of mRNA and amino acids of polypeptides is the tRNA. In order to fill this major role, tRNAs have two distinct characteristics corresponding to two different genetic codes, the anticodon and the operational codes. The latter which is mainly embodied in the acceptor-stem allows to bind covalently and with high specificity an amino acid to a tRNA, a reaction catalyzed by a specific aminoacyl-tRNA synthetase (aaRS) [6]. The operational code might have actually predated the "classic" code associated with anticodons [7]. Moreover, the tRNAs exhibit diversity in uniqueness, all of them must be similar for entering the ribosome machinery; therefore, they generally look structurally homogeneous, especially in their secondary and tertiary structures even if "non-classical" tRNAs are known [3]. Moreover, cloverleaf structure and especially the tertiary interaction network governing the L-shaped tRNA architecture imply conserved and semiconserved bps and nts. On the other side, each type of tRNA structures must interact specifically with aaRSs and posttranscriptional modification enzymes, which implies that parts of their sequences and of their structures (as the V-R size) allow to distinguish them.

Reduced bacterial and most organelle genomes do not encode the full set of 32 tRNA species required to read all triplets of the standard genetic code according to the conventional wobble rules. Superwobbling where a single tRNA species contains modifications of the anticodon-loop, such as an hypermodified uridine at the wobble position 34 of the anticodon, reads all 4 nts at third codon position and has been suggested as a possible mechanism for how reduced tRNA sets may be functional [8]. Indeed, many metazoan mtDNAs have only a total of 22 tRNAs, apparently sufficient to recognize all codons (two tRNAs each for serine and leucine and one tRNA for each of the other 18 amino acids). However, superwobbling induces a reduced translational efficiency, which could explain why most organisms have adopted pairs of isoaccepting tRNAs over the superwobbling mechanism [9]. Moreover, e.g., in Cnidaria (sea anemones, corals, etc.) or Chaetognatha (marine invertebrates), current mtDNAs have lost several of their *trn* genes, and the absence of an apparently full set of mt-*trn* genes has also been mentioned [10]. Studies have investigated the fate of missing tRNAs and their corresponding aaRSs [11], and in many cases, the lost tRNAs are functionally replaced by imported nucleus-encoded tRNAs [10]. However, recent search strategies suggest that efficient reanalyses detect several tRNA-like structures (TLS), which can be efficient tRNAs [12].

Compared to mitochondria found in other eukaryotic kingdoms, those of metazoa are massively reduced in their genetic structure [4]. Their mtDNA is a short, circular molecule that generally contains about 13 intronless, protein-coding genes, all of which are involved in aerobic respiration (also called oxidative phosphorylation) [13]. Moreover, the coding sequences of genes are usually separated by at most a few nts and long polycistronic precursor transcripts may be processed into mature mRNA and rRNA by precise cleavage of the 5' and 3'-termini of the flanking tRNAs. This processing, which is known as the tRNA punctuation model [14], is mediated by RNase P and Z endonucleases, respectively [15]. However, this model is not always applicable, genes are not bound by *trn* genes or these latter may not be involved in the

processing of precursor RNAs. Besides, in several taxa mt-mRNAs, rRNAs and even tRNAs may be oligoadenylated or polyadenylated [16]. This has numerous consequences with potentially dual and opposite roles: this promotes transcript stability or offers a target for initiating degradation. Overlapping genes on the same DNA strand occur throughout metazoa [17]. Therefore, the termination points of the protein-encoding genes could be difficult to infer as stop codons (generally UAA or UAG) may be absent. It is accepted that abbreviated stop codons (U or UA) are converted to UAA codons by polyadenylation after transcript cleavage, and this has been confirmed by analyzes of transcripts in some cases [18]. Sometimes, the initiation codon may also not have been detected. For several protein-encoding genes, the question of a possible overlapping with adjacent downstream or upstream *trn* genes is often raised [19]. Moreover, overlaps between adjacent mt-*trn* genes are frequent, but it is out of our topic [19, 20].

Incidentally, in 2004, searching for chaetognath mt-*trn* genes [21], it was observed that tRNAs bear nt triplets corresponding to stop or start codons at precise conserved positions, and this constitutes the original topic of this chapter.

## 2. Material and methods

Most of the research was done in two databases which include primary sequences and graphical representations of tRNA 2D structures, tRNAdb (<http://trna.bioinf.uni-leipzig.de/DataOutput/>) contained more than 12,000 *trn* genes from 579 species belonging to prokaryotes and eukaryotes, whereas in mitotRNAdb (<http://mttrna.bioinf.uni-leipzig.de/mtDataOutput/>), 30,525 metazoan mt-*trn* genes belonging to 1418 species were recorded [22]. Despite a bias for metazoa, these two databases provide powerful and fast search engines. Alignments were generated by Clustal W ([www.ebi.ac.uk/clustalw/](http://www.ebi.ac.uk/clustalw/)), whereas secondary structures were predicted by tRNAscan-SE (<http://lowelab.ucsc.edu/tRNAscan-SE/>) [23]. BLAST analyzes were conducted using the website: <https://blast.ncbi.nlm.nih.gov/Blast.cgi>.

## 3. Results and discussion

### 3.1. Frequencies of TAR10 and ATR49 triplets in various taxa

Visual observations of tRNA deduced 2D structures suggested that nt triplets which could correspond to stop or start codons seemed to be particularly represented at specific positions. The UAR (R for purine) triplets at position 8–10 in the standard numbering and therefore will be named UAR10, whereas the potential initiation codons whose last nt is at the position 49 will be called AUR49 (**Figure 1**). We chose to number the codons only according to their last nt because the nt 47 is frequently missing in the metazoan mt-tRNAs. Analyzes focus on DNA; hence, these are usually annotated TAR or ATR instead of UAR or AUR. All the tRNAs which bear one or both of these codons are named ss-tRNAs (ss for stop and start) or ss-*trn* for the corresponding genes. Using tRNAdb and mitotRNAdb databases, these triplets' frequencies were investigated in different taxa including nuclear and organelle genomes for eukaryotes (**Table 1**). Excluding taxa for which the number of *trn* genes is too low for statistical

	Nb	TAG	TAA	TAR	ATG	ATA	ATR	TAR+ ATR	TAR+ATR/ATR
<b>Eubacteria</b>	6383	62.2	0.63	62.8	1.61	0.20	1.82	0.58	31.9
Proteobacteria	2602	63.2	0.81	64.0	1.00	0.12	1.11	0.35	31.0
<i>α</i> -proteobacteria	613	67.7	0.16	67.9	0.65	0.16	0.81	0.65	80
<b>Archaea</b>	1081	49.2	0.93	50.1	0.00	0.09	0.09	0	0
<b>Eukaryotes</b>									
nuclear	2226	40.0	0.36	40.3	0.18	0.13	0.31	0.18	57.1
plastids	399	50.4	13.85	64.2	4.01	3.01	7.02	5.76	82.1
Viridiplantae*	160	59.4	7.50	66.9	3.16	1.25	4.38	3.75	85.7
Fungi*	299	32.8	8.36	41.1	4.01	13.04	17.06	11.71	68.6
<i>Porifera</i> *	491	53.4	14.46	67.8	8.15	14.66	22.81	18.94	83.0
<i>Acoelomata</i> *	418	45.7	9.09	54.8	17.22	7.18	24.40	13.16	53.93
<i>Pseudocoelomata</i> *	277	70.0	11.55	81.6	1.08	1.08	2.17	1.44	66.7
<i>Protostomia</i> *	4854	38.4	12.46	50.8	14.61	25.48	40.09	20.00	49.9
<i>Deuterostomia</i> *	24398	49.6	2.25	51.8	18.89	7.99	26.89	14.46	53.8

Metazoan taxa are in italics. *Abbreviation:* Nb, number of *trn* genes. TAR + ATR for % of *trn* genes bearing the two types of triplet. \* Mitochondria.

**Table 1.** Percentage of TAR10 and ATR49 triplets in various taxa independently of the V-R size.

analysis, TAR10 always occurs at high frequencies, whether in prokaryotic, nuclear, or organelle genomes. Values range from 41.1% for fungi to 81.6% for pseudocoelomates. In all the taxa and all tRNA species combined, the percentage of TAG10 triplets is always significantly higher than those of TAA10. The differences are very important in prokaryotic and nuclear genomes, since the percentage of TAA10 is always less than 1, while that of TAG10 is at least 40%. Within the organelle genomes, the difference is smaller but can vary by a factor of 2.5–22.

As the TAR10 triplet (principally TAG) is present in at least 40% of the *trn* genes for all taxa and genomic systems combined, it could have been present in *trn* genes of the Last Unicellular Common Ancestor (LUCA), which presumably lived some 3.5–3.8 billion years ago [24]. It is probably an ancestral character which was present in proto-*trn* sequences. As the percentage of TAA10 strongly increases in *trn* genes of organelles, one can ask whether this character was not already present in their bacterial ancestor. It is now assumed that despite their diversity, all mitochondria derive from an endosymbiotic *α*-proteobacterium which has been integrated into a host cell related to Asgard Archaea approximately 1.5–2 billion years ago [25]. However, the earliest fossils possessing features typical of fungi date to 2.4 billion years ago [26]. Moreover, the eukaryotic cells would be chimeras constituted of an archaebacterium and one or more Eubacteria [27]. In addition, all current models for the origin of eukaryotes suggest that the eukaryotic common ancestor had mitochondria. Therefore, as the level of TAA10 is very low in *trn* genes of *α*-proteobacteria, it could therefore be a derived trait that may be related to the increase in AT% in mtDNA and/or recognition constraints by mt-aaRSs and modification enzymes. Similarly, it is generally accepted that all chloroplasts and their derivatives are derived from a single cyanobacterial ancestor [28], and in current cyanobacteria, the respective percentage of TAG10 and TAA10 triplets are 62.5 and 3.6, respectively. The increase in the percentage of TAA10 characterizes organelles.

In all the taxa for all tRNA species combined, the ATR49 triplets are always present in smaller numbers than TAR10. Moreover, their numbers are negligible except in organelle genomes,

mainly mitochondria. The low level of ATR49 triplets in Pseudocoelomata is due to the frequent absence of T-arm in their mt-tRNAs. In mitochondria, in some taxa, the frequency of ATG49 is higher than of ATA49, while in others, the opposite occurs. The variability is not surprising, given approximately 2 billion years of mtDNA evolution [29]. It must be noted that the nt G is overrepresented at the 5'-end of the 5'-acceptor- and D-stems, quite often at the 5'-end of the T-stem but rarely at the equivalent position of the anticodon-stem. In taxa where the percentage of ATA49 is higher than those of ATG49, G is most often not the nt majority at the 5'-end of the T-stem. Moreover, differences between the relative percentages of ATG49 and ATA49 could be due, at least in part, to variations in the AT% in organelle DNAs. The percentage of ATR49 is very low in  $\alpha$ -proteobacteria and weaker in this last taxon compared to all Proteobacteria or Eubacteria, and it is also very weak in cyanobacteria, and so the significant rate of ATR49 triplets would seem to be a derived condition of organelle DNAs rather than a conserved primitive state lost in current prokaryotes. This trait probably appeared during the transition from endosymbiotic bacterium to permanent organelle that implied massive evolutionary changes including genome reduction, endosymbiotic and lateral gene transfers, and emergence of new genes and the retargeting of proteins [25]. The timing of the mt-endosymbiosis and of the proto-mitochondria to mitochondria transition is uncertain, but one might trace the origin of the ATR49 triplets between at least the first eukaryotic common ancestor (FECA) and the last eukaryotic common ancestor (LECA). A second event occurred, at least, in the mitochondria of the ancestors of Opisthokonta (i.e., Metazoa and Fungi), which would have led to a net increase in numbers of ATR49. ATR49 means that the last two nts of the V-R are AT. It turns out that this mainly concerns the mt-*trn* genes, whose V-R has only 4 nts, which are almost exclusively present in the Fungi/Metazoa clade.

There are large differences in the frequencies of the TAR10 and ATR49 triplets depending on the species of *trn* genes (**Table 2**) and taxa (data not shown), and the selective variations in some taxa suggest that the increase in frequency for some types of triplets would be much more recent than mentioned above; in addition, decreases are also observed. There are, however, very conservative trends such as the presence of ATR49 triplets in genes specifying tRNA-Ala. Analyses on mt-*trn* genes of Deuterostomia for which a great number of sequences for each type are available (from 1085 to 1382) show that only the tRNA-Cys and tRNA-Glu species have intermediate TAR10 percentages (**Table 2**). In all other tRNA species, the values are extreme, 9 and 10 tRNA species with values ranging from 0.4 to 9.8% or greater than 82.4%, respectively (**Table 2**). In contrast, half of the tRNA species have low ATR49 percentages ( $\leq$  to 10.8), and for only four types percentages are  $\geq$ 77.8. There would also be a tendency suggesting that tRNA species with very high or very low percentages of TAR10 most often have low ATR49 (the tRNA species with the 7 highest and the 8 lowest TAR10 percentages exhibit 10 out of 11 of the lower percentages for ATR49).

### 3.2. Examples of putative implications of TAR10 and ATR49 as stop and start codons

In order to investigate possible implications of TAR10 and ATR49 triplets in translation, analyses were performed in GenBank using as keywords: "TAA stop codon is completed by the addition of 3' A residues to the mRNA", "alternative start codon" or "start codon not determined" and mitochondrion (or mitochondrial DNA) complete genome. Then, it was researched whether upstream (for start codon) or downstream (for stop codon) of the protein-encoding gene was a *trn* gene. When a *trn* gene was found, TAR10 or ATR49 triplets were searched, and the same investigation

	Trp	Asp	Arg	His	Gly	Phe	Gln	Ala	Thr	Pro	Val
TAG	95.3	98.0	93.1	91.8	97.0	97.5	97.5	94.2	92.6	91.7	81.9
TAA	4.1	1.0	5.5	6.7	1.4	0.8	0.0	1.7	3.0	2.2	0.5
TAR	99.4	99.0	98.6	98.5	98.4	98.3	97.5	95.9	95.6	93.9	82.4
ATG	1.1	1.2	0.0	6.1	0.1	70.6	0.0	89.4	20.4	11.8	8.7
ATA	9.7	0.7	0.5	50.4	0.0	7.2	0.0	7.5	0.7	5.2	16.1
ATR	10.8	1.9	0.5	56.5	0.1	77.8	0.0	96.9	21.1	17.0	24.8
	Glu	Cys	Met	Leu1	Tyr	Leu2	Ser1	Lys	Ser2	Asn	Ile
TAG	16.6	17.0	0.7	8.1	7.1	5.9	0.1	2.4	1.2	1.1	0.4
TAA	8.8	0.6	9.1	0.3	0.1	0.0	3.0	0.2	0.3	0.2	0.0
TAR	25.4	17.6	9.8	8.4	7.2	5.9	3.1	2.6	1.5	1.3	0.4
ATG	2.0	17.7	95.6	0.1	9.8	0.2	0.0	8.1	42.7	1.9	32.5
ATA	2.2	3.1	0.3	1.1	14.4	3.2	0.0	0.1	0.5	0.6	52.0
ATR	4.2	20.8	95.9	1.2	24.2	3.4	0.0	8.2	43.2	2.5	84.5

The *trn* genes are represented by three-letter codes of amino acids. The tRNAs are ordered by decreasing TAR10 percentages. The percentage values  $\geq$  to 77.8, between 17.0 to 56.5 and  $\leq$  to 10.8 are underlined in yellow, blue, and green, respectively.

**Table 2.** Percentages of TAR10 and ATR49 by mt-*trn* gene species in Deuterostomia.

was then made in conspecific mt-genomes. Using this strategy, these triplets have been only found in metazoan mtDNAs, in which overlapping mt-*trn* genes have long been known.

An example of putative uses of TAR10 triplets as stop codons is presented in **Table 3** for a subclass of parasitic flatworms (Platyhelminthes : Eucestoda). Their mt-genetic code has only UAG and UAA as stop codons, avoiding possible bias due to use of other types of termination codons. In 51 among 66 complete mt-genomes, the first in-frame potential stop codon of the *cox1* gene is in the downstream *trnT* gene (24 cases with TAG10 suggesting a 10 nt overlap between *cox1* and *trnT* genes). Authors considering that this long overlap would be impossible have proposed a number of alternative options favoring overlap avoidance (e.g., [30]). (1) *cox1* might use an earlier atypical stop codon. (2) The 3'-end of the *cox1* mRNA could have an abbreviated stop codon (U or UA instead of UAG10) upstream the *trnT* gene which is completed by polyadenylation. (3) If in the potential long transcript, the cleavage would occur just after G10, the *cox1* mRNA would end with the complete UAG10 as stop codon and the first 10 nts of the *trnT* gene would be added by an unknown editing process. (4) The *trnT* gene would be shorter in its 5'-end lacking the nts from 1 to 8 or 9, e.g., this has been proposed for the mt-*trnT* of Cyclophyllidea (*Echinococcus granulosus*, *Hymenolepis diminuta*, and *Taenia crassiceps*). If the full stop codon is used, then there is only a single nt (G10) overlap between *cox1* and *trnT*. Moreover, if the end of the *cox1* gene is at the level of T9, the stop codon would complete by polyadenylation; whereas if the protein gene has a complete stop codon, the nt G10 would be added by edition. In the alternative structures, the D-arm is absent, whereas it is typical for this tRNA in digeneans (a class of Platyhelminthes) and in other phyla. However, mt-*trnT* genes issuing from Cyclophyllidea for which the first potential stop codon is at different positions (upstream or downstream the *trnT* gene, or in this last gene but upstream or downstream TAG10 or at this last position) exhibit

Order	Family	Species and accession numbers	TAR10	stop cod.
Bothriocephalidea	Bothriocephalidae	<i>Schyzocotyle nayarensis</i> (KX060589*)	TAG <sup>x</sup>	
		<i>Schyzocotyle acheilognathi</i> (KX589243, KX060588, KX060587, KX060590-KX060595), <i>Senga ophiocéphalina</i> (KX434430)	TAG	TAG <sup>§</sup>
Caryophyllidea	Capingentidae	<i>Atractolytocestus huronensis</i> (KY486754*), <i>Breviscolex orientalis</i> (KY486752*),	TAG <sup>x</sup>	
	Lytocestidae	<i>Khawia sinensis</i> (KY486753*)		
	Caryophyllaeidae	<i>Caryophyllaeus brachycollis</i> (KT028770)	none	TAA <sup>&amp;</sup> .
Cyclophyllidea	Anoplocephalidae	<i>Anoplocephala magna</i> (KU236385), <i>A. perfoliata</i> (KR054960)	TAG	TAG <sup>§</sup>
		<i>Moniezia benedeni</i> (KX121040), <i>M. expansa</i> (KX121041)	TAG	TAG <sup>μ</sup>
	Hymenolepididae	<i>Hymenolepis diminuta</i> (AF314223*), <i>H. nana</i> (KT951722), <i>P. c.</i> (KR611041*)	TAG <sup>x</sup>	
		<i>Drepanidotaenia lanceolata</i> (KR817910)	none	TAG <sup>§</sup>
		<i>Cloacotaenia megalops</i> (KU641017*)	TAG	TAA <sup>&amp;</sup>
	Paruterinidae	<i>Cladotaenia vulturi</i> (KU559932*)	TAG <sup>x</sup>	
	Taeniidae	<i>Echinococcus granulosus</i> (AF297617, KJ559023), <i>Hydatigera parva</i> (NC_021141), <i>Taenia arctos</i> (NC_024590), <i>T. crassiceps</i> (AF216699), <i>T. solium</i> (AB086256), <i>Versteria mustelae</i> (NC_021143)	TAG <sup>x</sup>	
		<i>Hydatigera krepkogorski</i> (NC_021142)	TAG	TAA <sup>§</sup>
		<i>Echinococcus canadensis</i> (AB208063, AB235847, AB235848), <i>E. equinus</i> (AF346403, NC_020374), <i>E. multilocularis</i> (AB018440*), <i>E. oligarthrus</i> (AB208545), <i>E. ortleppi</i> (AB235846), <i>E. shiquicus</i> (AB208064)	TAG	TAG <sup>§</sup>
		<i>Echinococcus vogeli</i> (AB208546*), <i>Taenia asiatica</i> (NC_004826), <i>T. crocutae</i> (NC_024591), <i>T. hydatigena</i> (FJ518620, GQ228819), <i>T. multiceps</i> (FJ495086, GQ228818), <i>T. pisiformis</i> (GU569096), <i>T. regis</i> (NC_024589), <i>T. saginata</i> (AY684274), <i>T. taeniaeformis</i> (JQ663994, FJ597547)	TAG	TAA <sup>&amp;</sup>
<i>Echinococcus granulosus</i> (KU601616)		TAG	TAA <sup>μ</sup>	
Diphyllobothriidea	Diphyllobothriidae	<i>Diphyllobothrium latum</i> (AB269325, DQ985706), <i>D. nihonkaiense</i> (EF420138, AB268585), <i>Diplogonoporus balaenopterae</i> (NC_017613), <i>D. grandis</i> (NC_017615), <i>Spirometra erinaceieuropaei</i> (KU852381, AB374543)	TAG <sup>x</sup>	
		<i>Spirometra decipiens</i> (KJ599679*)	TAG	TAA <sup>&amp;</sup>
Proteoce.	Proteocephalidae	<i>Testudotaenia</i> sp. (KU761587)	TAG <sup>x</sup>	

Species names are followed by their accession number(s). \*: sequences for which the authors of these latter considered that there was an abbreviated stop codon and this latter was upstream the *trn* sequence. Symbols: <sup>x</sup> TAR10 was the first in-frame putative stop codon; <sup>§</sup>, <sup>§</sup>, <sup>&</sup> and <sup>μ</sup>, the putative stop codon was upstream the *trn* gene (*tg*), in the *tg* but upstream or downstream TAR10 or nts 8–10, downstream the *tg*, respectively. Abbreviations: *P. c.*, *Pseudanoplocephala crawfordi*; Proteoce., Proteocephalidea; stop cod., putative stop codon according to the authors of the sequences.

**Table 3.** Position of the first complete in-frame stop codon of the *cox1* gene versus the following *trnT* gene in Cestoda (Platyhelminthes).

similar secondary structures, including a D-arm. In addition, the high level of nt conservation in the 5'-end of the *trnT* genes of cestoda (i.e., G1, G2, T7, T8, A9, G10, T11, T12 and A14) suggests strongly that the 5'-acceptor-stem and the D-stem are under positive selection. All this implies that the hypothesis of D-armless tRNAs is, according to us, improbable.

Concerning the putative ATR49 start codon, in GenBank, the number of complete mt-genomes found using the keywords previously mentioned was relatively low; moreover, in some cases, the upstream gene encoded a protein, specified a rRNA and/or there was only one mention for a given taxon. A significant example within Deuterostomia (frogs) is presented in **Table 4**. In the superfamily Hyloidea, the ATA49 triplet is frequently the first potential complete start codon at the level of the gene pair encoding and specifying NAD1 and tRNA-Leu2, respectively. In two families (Bufonidae, Hylidae), for all the sequences (16 belonging to 14 different species), the first ATR triplet found in frame in the ORF of the *nd1* gene is ATA49. For four sequences belonging to three other frog families, the ATR49 triplet is missing from the *trnL2*

Family	Species and accession numbers	ATA49 triplet	Alternative start codon
Dendro.	<i>Anomaloglossus baeobatrachus</i> (NC_030054)	none	AAA <sup>§</sup>
Bufo- nidae	<i>Bufo gargarizans</i> (NC_008410*), <i>B. japonicus</i> (NC_009886*), <i>B. melanostictus</i> (NC_005794*), <i>B. stejnegeri</i> (NC_027686*), <i>B. tibetanus</i> (NC_020048*), <i>Bufo raddei</i> (NC_028424*)	ATA <sup>x</sup>	
	<i>Bufo gargarizans</i> (KM587710, KU321581)	ATA <sup>x</sup>	TAA
Cerato.	<i>Telmatobius chusmisenis</i> (KT949346), <i>T. bolivianus</i> (NC_020002)	none	TTG
Heleo.	<i>Heleophryne regis</i> (NC_019998*)	none	GTG
Hylidae	<i>Dryophytes suweonensis</i> (NC_032380*), <i>Hyla annectans</i> (NC_025309*), <i>H. chinensis</i> (NC_006403*), <i>H. japonica</i> (NC_010232*), <i>H. suweonensis</i> (NC_034238*), <i>H. tsinlingensis</i> (KU601448*, NC_026524*), <i>H. ussuriensis</i> (NC_029410*)	ATA <sup>x</sup>	

<sup>x</sup>: ATA49 as the first putative in-frame start codon. \*: "start codon not determined" according to the authors of the sequences. Alternative start codons are given by the authors of the sequences. <sup>§</sup>: alternative start codon in the *trn* gene. Abbreviations: Dendro., Dendrobatidae; Cerato., Ceratophryidae; Heleo., Heleophrynidae.

**Table 4.** Position of the first putative start codon of the *nad1* gene versus the upstream gene specifying tRNA-Leu2 in Hyloidea, a superfamily of frogs.

gene; moreover, an ATA triplet is integrally present in the V-R of the *trnL2* gene of *Heleophryne regis*, but it is not in frame with the following gene. For these last four cases, the authors of the sequences proposed alternative start codons. This seems obligatory, but this has not been experimentally verified. For several authors who have sequenced parts of mtDNAs of Hylidae, the *nd1* gene would start at ATA49 for about 140 sequences (e.g., Roelants and Bossuyt [31]).

In the two studied taxa, Blast analyzes of the NCBI ESTs and SRA (SequenceRead Archive) databases have been performed, but no result supports the proposed hypotheses: transcripts starting at an ATR49 or terminating at a TAR10 were not found. However, for each taxon, few mt-transcripts occur, and fully matured transcripts are even rarer.

### 3.3. Why mt-ss-*trn* genes with TAG10 and ATR49 triplets as putative stop and start codons only occur in Metazoa?

Foremost, biases in the search strategy cannot be excluded, but the important point to note is that mt-genomes of animals, fungi, protists, and plants differ drastically in all major characteristics including gene content and large size variation. Generally, metazoans have ultra-compact mtDNAs (from c.10,000 to c.50,000 bp); usually, nonfunctional sequences are rapidly eliminated, and there are short intergenic regions and frequent overlaps [13]. However, nonbilaterian mt-genomes have higher variation in size, gene content, shape, and genetic code [32]. The mtDNA size range is from 30,000 to 90,000 bp in fungi, and generally, intergenic regions are relatively long. A broader range of mtDNA size is found in higher plants (from  $0.2 \times 10^6$  to about  $11.3 \times 10^6$  bp [33]), and the largest known mt-genome in this lineage exceeds sizes of reduced bacterial and nuclear genomes [34]. The increased sizes of plant mtDNAs are mostly due to noncoding DNA sequences, large inserted nuclear regions, and many introns and not to a large increase

in gene numbers. The nuclear-derived sequences amount to up nearly half of their size as in melon [35], and so presence of mt-*trn* genes with nuclear origin cannot be excluded. Although not directly correlated, intergenic distances are generally much higher in larger genomes, reducing the number of overlaps. In addition, the situation of plant mt-tRNAs is very complex. Indeed, they contain few “native” tRNAs expressed from true mt-*trn* genes. They possess “chloroplast-like” *trn* genes inserted into the mtDNA. They compensate the loss of mt-*trn* genes by importing several nucleus-encoded tRNAs [36]. In addition, most often in plants, the standard code applies to the reading of organelle genomes, even if ATA is frequently used as start codon. Metazoan mt-genomes are generally small, very constrained and exhibit several gene overlaps between *trn* and protein-encoding genes or between *trn* genes. Their tRNAs have sequence and structural peculiarities and tend to shortening [19]. Our exploration is not exhaustive, but this might explain the presence of putative stop or start codons specifically within mt-*ss-trn* genes of this taxon.

One may wonder why the hypotheses concerning the TAR10 and ATR49 triplets were not proposed before? At least, the presence of these characteristic triplets could have been observed by some authors but considered as having no connection with the translation of neighboring protein genes. Among the first sequenced and the most studied were mt-genomes of *Homo sapiens* (J01415) and *Mus musculus* (J01420). In these latter, no putative start or stop codon occurs at these positions within *trn* sequences adjacent to protein genes (data not shown).

### 3.4. Some known structures in the living world bearing a start or stop codon at least in part of a stem-loop structure

In the living world, many nucleic sequences with secondary structures playing a physiological role involving stop and/or start codons have been discovered. Some representative examples are briefly presented here. (1) The tropism switching of the bacteriophage BPP-1 is mediated by a phage-encoded diversity-generating retroelement, which introduces nt substitutions in a gene that specifies a host cell-binding protein (*Mtd*) [37]. The nt substitutions are introduced in a variable repeat located at the 3'-end of this gene. Two nts after this region, the UAG stop codon is present, and its last nt is situated at the 5' beginning of the 5'-stem of a hairpin. Both the UAG codon and hairpin are required for phage tropism switching. (2) Programmed translational bypassing is a process, whereby ribosomes “ignore” a substantial interval of mRNA sequence. In a bacteriophage T4 gene, bypassing requires translational blockage at a “takeoff codon” immediately upstream of the UAG stop codon, and both codons are in the 5'-stem of a hairpin; moreover, this region is mobile [38]. (3) The operon *flgFG* of the bacterium *Campylobacter jejuni* can encode two genes (*flgF* and *flgG*). Its expression in *E. coli* produces a fusion protein probably due to ribosomal frameshifting (translational hopping) [39]. The putative hop region contains, among others, a hairpin beginning by the last nt of the UAA stop codon of the first mRNA. The AUG start codon of the second gene is in the loop of the following hairpin. (4) In Eubacteria, riboswitches are regulatory segments of DNA or mRNA that can bind a small molecule (the effector), which repress or activate their cognate genes at transcriptional and/or translational levels. In the riboflavin and *cob* operons, conformational changes can form a stem loop which sequesters the translational start site, consisting of the Shine-Dalgarno (SD) sequence plus start codon thus preventing gene translation [40]. (5) Bacterial transfer-messenger RNAs (tmRNAs) have dual TLS and mRNA-like properties. They rescue stalled ribosomes on mRNAs lacking proper translational stop signal; the tRNA-like

structure acts first as an alanine-tRNA, and then the short mRNA reading frame is translated and the product is released [41]. This *trans*-translation terminates at the stop codon terminating the tmRNA reading frame. This stop can be in a little loop or totally or partially integrated in the stem of a hairpin-like structure. In eukaryotes, structurally reduced tmRNAs (no mRNA-like domain) rarely occur in chloroplasts [42] and in mt-genomes (in Jakobids, presumably close to the most ancient living eukaryotes with bacterial-like mt-genome) [41]. Moreover, tmRNA TLSs function even without any canonical initiation factors. These examples show that start or stop codons located in hairpin may have various functions, as we suggest for TAR10 and ATR49.

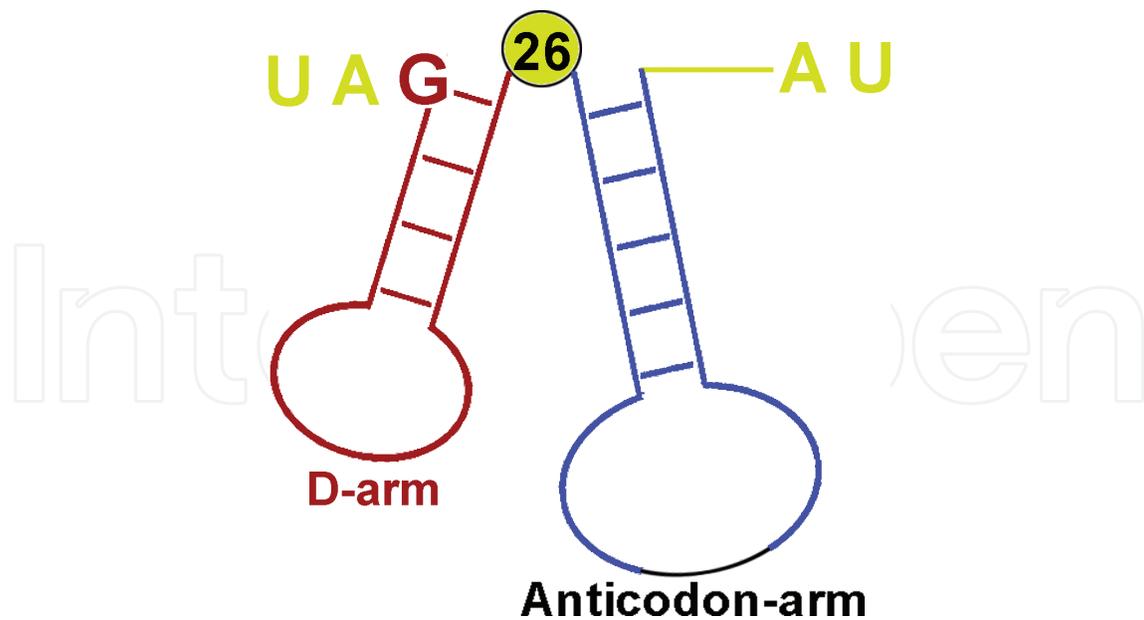
### 3.5. Multifunctionality of tRNAs

Ancient tRNAs probably had diverse functions in replication and proto-metabolism before protein translation [43] and modern tRNAs have also various functions in all the living organisms [1]. These functions include cell wall synthesis, protein N-terminal modification, nutritional stress management, porphyrin biosynthesis (heme and chlorophyll), lipid remodeling, and initiation of retrovirus reverse transcription. Accumulating experimental evidence suggests also that they have important regulatory roles in translation, viral infections, and tumor development (reviewed in [44]). Mt-tRNAs interfere with a cytochrome *c*-mediated apoptotic pathway and promote cell survival [45] and function as replication origins [46]. Moreover, nuclear-tRNA abundance and modifications are dynamically regulated, and tRNAs and their tRNA-derived RNA fragments (tRFs) are centrally involved in stress signaling and adaptive translation [47, 48]. This suggests that the choice of cleavage sites of mRNA transcripts with or not part of the neighboring ss-tRNA could be dynamic and also respond to environmental changes. Some of the noncanonical translation functions of tRNAs can also be driven or enhanced by their ability to adopt different complex three-dimensional structures, and these conformational changes can be linked to functional states [49]. Moreover, the tRNA multifunctionality has also been considered to be, at least in part, random due to the high amount of tRNA species within the cell [1]. In addition, the mt-*trn* genes represent natural pause sites for replication forks and could also prone double-strand breaks [50], and their role, as “punctuation signals,” for processing of mtDNA polycistronic transcripts has already been mentioned.

Enormous numbers of tRFs in all domains of life were found in the last decade [44]. In the plant *Arabidopsis thaliana*, nucleus-, plastid-, and mt-encoded tRNAs can produce tRFs [36]. The tRFs are not randomly degraded tRNAs. Experiments showed several functions including regulation of tumor development and viral infections [44]. Degradations resulting from cleavages at TAR10 and ATR49 triplets could produce a conformation exhibiting two loops linked by a forked-stem structure, roughly resembling a pair of cherries, so called “cherry-bob” (**Figure 2**). Our hypothesis predicts this structure which however has never been observed [51].

### 3.6. Other putative roles of TAR10 and ATR49

Metazoan mt-genomes are believed optimized for rapid replication and transcription. Potentially, TAG10 and ATR49 make transcription/translation more complex but perhaps more efficient. Examples in the Section 3.2 (i.e., Eucestoda) suggest mt-overlaps appeared 100s millions of years (MY) ago, enabling co-evolution between protein-encoding genes and those specifying tRNA.



**Figure 2.** 2D “cherry-bob”/“bottom half” structure. AUG/UAG triplets are discussed in text. Colors as **Figure 1**.

Overlaps involve numerous constraints for genes including sequence bias. Constraints are probably less stringent for *trn* genes, which can evolve rapidly because relatively standard secondary structure coupled with a specific anticodon might suffice for tRNA function [52]. Incomplete cloverleaf structures may also be repaired post-transcriptionally [53].

Alternative processing might be possible for the production of either a supposed complete mRNA or a complete tRNA. In the first case, the synthesis of new complete mRNAs could be promoted by high mitosomal tRNA numbers. Moreover, amino acid starvation can regulate mt-tRNA levels [54]. However, if mt-tRNAs already present are not destroyed, translation would not immediately stop because mt-tRNA half-life which is lower than that of their cytosolic counterparts can nevertheless exceed 10 h [54]. Moreover, aberrant mt-tRNAs can be corrected by RNA editing during or after transcription, and this process appeared independently several times in a wide variety of eukaryotes [5]. As an extreme example, due to large overlaps between *trn* genes, up to 34 nts are added post-transcriptionally during the editing process to the mt-tRNA sequences encoded in an onychophora species, rebuilding the acceptor-stem, the T-arm, and in some extreme cases, the V-R and even a part of the anticodon-stem [55]. In that species, several edition types must be combined, including template-dependent editing [55]. This last example suggests that complete tRNA could be restored after a cleavage just upstream of ATR49. However, edition of parts of the 5'-end of tRNAs seems more problematic. Besides, mRNAs with upstream or downstream ss-tRNA can form a partially double strand region with a homologous ss-tRNA at the level of the acceptor-stem. This might induce mRNA degradation via antisense mechanisms. In bacteria, uncharged tRNAs cause antisense RNA inhibition [56], and small interfering cytosolic tRNA-derived RNAs exist [57]. Modifications (methylation, edition, etc.) of incomplete tRNAs generated after cleavages of polycistronic transcripts at TAR10 or ATR49 triplets would indicate regulatory functions.

Putative use of TAR10 or ATR49 triplets affects protein length. When in frame, this could generate a protein at least 3 or 9 amino acids longer, respectively. Extension length depends

on positions of upstream stop codons completed by polyadenylation and/or on downstream (alternative) initiator codons. Not only complete proteins may be functional. Depending on cleavage positions in polycistronic transcripts, consequences may be neutral, disadvantageous, or favorable in specific contexts. In yeast, extended proteins can increase fitness under stress conditions [58]. In addition, in bacteria and in organelles, alternative initiation codons decrease efficiency [5], and it must be noted that ATR49 triplets are “canonical” start codons.

In other conditions, incomplete mRNAs could be favored. Mitosolic mRNA accumulations can be due to lack of translation because of tRNA paucity. Thus, high mRNA levels might indirectly promote cleavage of entire tRNA transcripts while reducing the synthesis of new functional mRNAs and favoring translation of those which are already present into proteins. Presence/absence of hairpins involving stop or start codons might regulate translation. This regulation could involve proteins that stabilize the hairpins or posttranscriptional modifications. Moreover, translational products of “incomplete” mRNAs might have housekeeping functions.

Regulation of alternative processing producing either complete tRNAs or complete mRNAs requires elucidation. Factors, probably proteins, need characterization. Note that metazoan *mt-atp8* and *atp6* genes overlap (mainly by 10 bp in vertebrates) and are transcribed as joint bicistronic transcript [59]. This proven overlap is inherent to mt-metabolism. Hence, similar overlaps assumed for TAR10 triplets are plausible.

Overlap conservation might reflect the need to produce bicistronic transcripts (5'-tRNA-mRNA-3' or 5'-mRNA-tRNA-3') or functional constraints at protein level (i.e., preserving specific amino acid patterns upstream or downstream the ORF). When overlap regions have conserved, amino acid sequences at the protein N- or C-terminal functional constraints at protein level for overlaps are probable [19]. In viruses, mutation rates are low in DNA regions coding for multiple protein products in separate reading frames (called overprinted genes) because point mutations compatible with functional products from all frames are rare. In these regions, the frame is said “close off.” Partial overlap between protein-encoding genes and *ss-trn* genes would present similar situations explaining greater conservation of extremities of protein and tRNA sequences when the corresponding genes overlap. This lock almost only concerns the *ss-tRNA*'s “top half,” limiting changes in the region interacting with many processing enzymes. The *ss-trn* genes could also regulate translation upstream, bicistronic mRNA/*ss-tRNA* transcripts could be more stable, and likewise, *ss-trn* genes could also play roles in replication and transcription.

### **3.7. Methylation of *trn* genes and tRNAs and their possible roles in transcription and translation**

Methylation is much rarer in mt- than nuclear-DNA [60]. However, these might occur at *trn* genes (particularly around TAR10 and ATR49) and might have deleterious consequences especially because differential mtDNA methylations are linked to aging and diseases (including diabetes and cancers) [60]. Methylation of nts of UAR10 and AUR49 is known as those of A9 and G10 which can be important for correct tRNA foldings [61]. We are unaware whether posttranscriptional modifications occur on bicistronic mt-transcripts containing complete or partial tRNAs. This would be worth investigating including possible consequences on maturation and translation.

### 3.8. Reassignments of codons and ss-tRNA

Several codon-amino acid reassignments are known, mainly from mitochondria [62, 63]. In 11 different mt-genetic codes, UGA stops code for tryptophan and AUA codes for methionine instead of isoleucine in 8 and 5 mt-genetic codes, respectively [63]. Both reassignments avoid potential errors along traditional wobble rules. Reassigning UGA-stop to UGA-Trp fits the “capture” hypothesis, and UGA codons mutate first to synonymous UAA codon in AT-rich mt-genomes. Then, UGA reappears occasionally by mutations, free for “capture” by an amino acid, like Trp [64]. AUA is frequently used as alternative initiation codon. Its reassignment to internal sense Met codon could also have evolved in AT-rich genomes. Moreover, the standard genetic code assigns six codons to arginine, whereas two would fit arginine’s relatively low frequency in current proteins [65]. In 8 out of 11 mt-codes, different strategies reduce Arg codons to four, AGR reassignments to other amino acids (in six genetic codes), lack of two Arg codons (CGA and CGC yeast mt-code), and AGR as terminators in vertebrates. These AGR codons were believed mt-stop codons since early vertebrate evolution [66]. However, at least in humans, AGRs are not recognized terminators [67], suggesting that AGRs have no assignment. Hence, the vertebrate mt-genetic code could be the most optimized known genetic code (that of yeast was not retained because four Leu codons were reassigned to Thr). Characteristics of the nt triplets at the position 8–10 and ending at position 49 should be analyzed for each mt-genetic code.

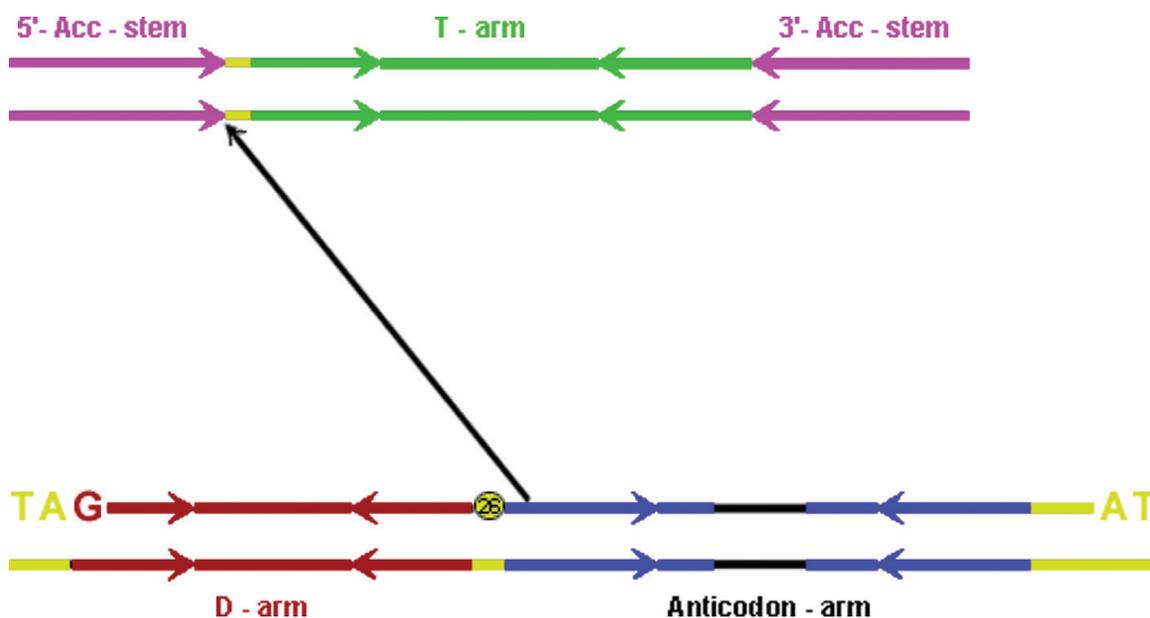
### 3.9. Origin of the cloverleaf structure of tRNA and ss-tRNA

Various models could explain tRNA origins (see reviews [68–70]). The modern tRNA cloverleaf structure might result from direct duplication of primordial RNA hairpins (e.g., [68]). However, studies lend strong support to the “two halves” hypothesis [43], in which tRNAs consist of two coaxially stacked helices with presumed independent structural and functional domains. These correspond to the “top half” containing the acceptor-stem and the T-arm and the “bottom half” with the D-arm and anticodon-arm (**Figure 1**). The 2D representation of the latter corresponds to the “cherry-bob” structure (**Figure 2**). The “top half” of modern tRNA embeds the “operational code” in the identity elements of the acceptor-stem that interacts with the catalytic domain of specific aaRSs and is recognized by RNases P and Z and the CCA-adding enzyme (therefore mainly RNA end processing reactions) [70, 71]. This domain also interacts with translation elongation factor Tu and one rRNA subunit [71]. The importance of this domain in most macromolecular interactions involving tRNAs (including *in vitro* even when it is detached from the “bottom half”) suggests that these half’s specificities were established before the tRNA’s “bottom half,” presumably incorporated later [72]. Growing evidence for tRNA elements involved in both RNA and DNA replication with the 3'-end playing a determinant role has led to the idea that the “top half” initially evolved for replication in the RNA world before the advent of protein synthesis [73]. The supposed evolutionarily recent tRNA “bottom half” provides genetic code specificity. This suggests late implementation of the standard genetic code and late appearance of interactions between the tRNA “bottom half” and ribosomes [74]. Whether the “bottom half” derived from a loop or extra loop belonging to the “top half” or was an independent structural and functional domain that was subsequently incorporated into the “top half” remains unresolved [71]. Some authors suggest independent evolutionary origins [71, 72].

The study of ss-tRNAs suggests a model partially explaining canonical tRNA origins (**Figure 3**). The DNA region specifying the “bottom half” would be integrated in a sequence that can specify the “top half” but at the junction between the parts corresponding to the 3’-end of the 5’-acceptor stem and the 5’-end of the 5’-T-stem.

On the other hand, the “bottom half”/“cherry bob” structure could also be integrated at RNA level, either in the RNA world by intermolecular RNA-RNA recombination or template switches or later with retrotranscription events. Fujishima and Kanai [70] also proposed an equivalent model where a long hairpin corresponding to about the “top half” region merged with a viral RNA element corresponding to the “bottom half” to give the TLS found in modern viral genomes (who however possessed a pseudoknotted acceptor-stem). Besides, rare pre-tRNA molecules from the three domains of life exhibit an intron. The intron’s origin is debated. The “introns-early” scenario assumes most of them were lost during evolution, and the opposite scenario theorizes that introns were inserted into some *trn* genes after their emergence [75]. To date, our hypothesis would rather favor the second scenario, even though it could be considered that the “cherry bob” structure could be an ancestral intron becoming unspliceable.

In tRNAs, the two first nts of both UAR10 and AUR49 belong to connector 1 and 2, respectively. They are thus at the junction between the top and bottom halves and are very close physically in the 3D structure (**Figure 1**). The belonging of some of the nts of the TAR10 and ATR49 triplets to either of the two parts is not discussed here because the theoretical model of **Figure 3** is applicable independently of “bottom half” extremities. However, as the V-R is important for aminoacylation [76], ATR49 triplets could rather integrally belong to the “top half.” The tRNA L-shape is stabilized by various tertiary interactions of the V-R with the D-arm and between the D- and T-loops. Nucleotides of the connectors form contacts with the D-arm, and in some



**Figure 3.** Proposed model for the origin of genes specifying tRNAs with canonical cloverleaf structure. It can be summarized by insertion (follow the arrow) of the region specifying the “bottom half” into those specifying the “top half”. Here, the entire TAG10 triplet presumably belonged to the “bottom half” region as well as the first 2 nts of ATG49. Colors as **Figure 1**.

tRNAs, the G10 can establish potential tertiary interactions with a nt of the V-R upstream the putative start codon [77]. At least in cytosolic tRNAs, frequently U8 and sometimes U48 form noncanonical pairs. Moreover, generally, base pair 15–48 is more conserved in mt-tRNAs than 8–14, and this is probably due to the fundamental role played by the first in maintaining the tRNA L-shape [5]. UAR10 and AUR49 had to play first only a role in the L-shaped tertiary structure of tRNAs, and their implication as codons, if it exists, would be only a derived character. It was hypothesized that DNA punctuation evolved from 2D structures signaling polymerization initiation, termination, and/or processing to linear sequence motifs, which further evolved to translational signals [78]. In ss-tRNA, UAR10 triplet probably already plays a structural role in proto-tRNAs, whereas AUR49 would have appeared only during the evolution of organelle tRNAs and was related to L-shaped tertiary structures of organelle tRNAs and due to severe genome reduction and extreme base compositions. The opposite hypothesis would imply that the AUR49 triplet would have been a plesiomorphic character counter-selected in large genomes but kept in certain bacterial genomes up to mt-ancestors.

### 3.10. tRNAs at the origin of all the nucleic members of the RNA/protein world

Some authors have hypothesized that tRNAs may be the precursors of mRNAs, rRNAs (and therefore proto-ribosomes), and also of the first genomes. Several suggested similar origins for tRNA and rRNA [79]. Analyzes of sequences and secondary structures of ribosome suggested that the ribosomal peptidyl transferase center (PTC) which forms peptide bonds between adjacent amino acids originates from fused proto-tRNAs [80]. Strikingly, the ribosome is a ribozyme, since only RNA catalyzes peptide bond formation [81]. Otherwise, current eubacterial rRNAs themselves could encode several tRNAs [82] and chaetognath 16S rRNA genes appear as tRNA nurseries [12] (or the opposite). Eubacterial 5S rRNAs contain TLSs similar to alanine and arginine tRNAs [82], exhibiting tRNA-like 2D structures [83]. Some suggest that rRNAs are fused tRNA molecules [80].

Molecular biology dogmatically assumes that “tRNA genes are of course entirely noncoding” [84]. But in 1981, Eigen and Winkler-Oswatitsch suggested that in the RNA world to the RNA/protein world transition, ancestral tRNAs were mRNAs [85]. Assuming that the first mRNAs had been recruited from proto-tRNAs, it follows that TLSs were inside viral and cellular mRNAs [86]. Self-recognition between tRNA-like mRNAs and canonical cloverleaf tRNAs could stabilize these molecules and produce proto-proteins [87]. The first proteins potentially emerged from junctions of ancestral tRNAs, and among the modern proteins, the only polymerase which matched with tRNAs translated like a mRNA was the RNA-dependent RNA polymerase [87]. Otherwise, eubacterial rRNAs could also encode several active sites of key proteins involved in the translation machinery [82]. Then, analyzes of sequences and secondary structures of ribosomes suggested that these derived from tRNAs also functioned as a protogenome [82]. The very parsimonious syncretic model “tRNA core hypothesis” assumes that some proto-tRNAs were classical tRNAs and also functioned as rRNAs and mRNAs, a self-recognition between these molecules allowed to obtain proto-proteins [88].

Assuming that the ATR49 triplets are a primitive character lost during the first genome expansions and that they could already act as an initiation codon seems too speculative, but RNA structures having characteristics of ss-tRNAs could have accumulated many advantages in

the RNA/protein world. Structures with both start and stop codons partially in a stem-loop (as ss-tRNA), constituting basic signals for translation, could be a missing link of the RNA world hypothesis. Furthermore, in these proto-tRNAs, 3D structures could act as initiation and termination signals before the emergence of standard codons. Moreover, mRNAs in the form of ss-tRNA or a combination of several of these molecules would have been relatively stable. The cloverleaf structure could facilitate its entry into the PTC, and then interactions with other factors could allow a short region to be in linear form and thus could be read. Upstream and downstream of the linear region, the arrangement in hairpins protected the proto-mRNA from degradation during its reading, and as soon as a long enough region was read, it could take again its original 3D structure. Otherwise, circular proto-mRNAs derived from ss-tRNA-like molecules could not be excluded, although the hypothesis of circular tRNA-like ancestor ("proto-tRNA") was first proposed by Ohnishi in 1990 [89]. Furthermore, nuclear-encoded mt-tRNAs of Kinetoplastid protists are imported into the mitochondrion, and circularized mature tRNA molecules are produced probably by mt-endogenous RNA ligase activity (*in vivo* or during mt-isolation) [90]. Moreover, in red and green algae and possibly in one Archaea, the maturation of permuted *trn* genes, in which the sequences encoding the 5'-half and 3'-half of the specific tRNA are separated and inverted on the genome, needs the formation of a characteristic circular RNA intermediate which after cleavage at the acceptor-stem generates the typical cloverleaf structure with functional termini [91]. If in a ss-tRNA with a T-loop of 7 nts, the nt72 is ligated to the nt1; this creates a small ORF starting with a start codon (AUR49), which potentially codes for a peptide of 12 amino acids if UAR10 is used as stop codon. However, the circularization could be done elsewhere than at levels of nts 72 and 1. Thus, UAR10 would not be in frame, and therefore, this could allow the synthesis of smaller or longer peptides. To date, the formation of this type of structure and its translation remains hypothetical; however, experimental data shown that circular RNAs can be translated in prokaryotic and eukaryotic systems in the absence of any particular element for internal ribosome entry as SD sequence, poly-A tail, or cap structure [92]. Therefore, the evolutionary advantage of a circular proto-mRNA is also posited to be the simplicity of its replication mechanism and not be able to be degraded by the extremities that do not have one.

Besides, the fusion of tRNA-like mRNA and a classical tRNA could be at the origin of the ancestors of tmRNAs, and it can be mentioned just for guidance that the size of the tag peptide encoded by bacteria is of the same order of magnitude as those corresponding to putative translation of a ss-tRNA from the ATR49 triplet. Moreover, evolution of self-charging proto-tRNAs may also be selected [93], it has even been proposed that the activity of the juxtaposed 2'/3'-OHs of the tRNA A76 ribose qualifies tRNA as a ribozyme [94] and some RNAs (the early tRNA adaptor) must have had the ability to undergo 3'-aminoacylation. It has also been previously shown that many hairpin-structured RNAs bear ribozyme activity. These catalyze self-cleavage and ligation reactions [95]. In addition, it remains possible that circular ss-tRNAs with amino acid-anchored structure could be at the origins of tmRNAs. Indeed, two-piece bacterial tmRNAs (e.g., in  $\alpha$ -proteobacteria) are encoded by a circularly permuted gene sequence implying that pre-tmRNA is processed, and that the two pieces are held together by noncovalent interactions. Moreover, in line with an  $\alpha$ -proteobacterial origin of mitochondria, probable mt-encoded circular permuted *tmRNA* genes have been found in the oomycete (water mold) *Phytophthora sojae* and in the jakobid *Reclinomonas americana* [96]. A proto-*trnA* gene could be

at the origin of modern tmRNAs [41]. Metazoan mt-*trnA* genes combine the highest levels of TAR10 and ATR49 triplets (>95% for each), but in the prokaryotic world, if the rate of TAG10 is always higher than 91%, only one ATR49 occurs in Eubacteria and none in Archaea.

#### 4. Conclusions

Studies strongly suggest that the tRNA cloverleaf structure unfolded prior to the appearance of a fully functional ribosomal core, making it one of the most ancient RNAs of the RNA world [70, 97] or even the oldest [98]. Though the “RNA-world” hypothesis is well accepted, the successive events leading to the emergence of different partners playing a role in translation and the involvement of tRNAs in this evolution are highly controversial coveted field [99]. However, some hypotheses as the “tRNA core” [88] strongly suggest that tRNAs would be at the origin of the primitive genetic material and gave rise to mRNA and rRNA, as well as the conformational structure of the first proto-ribozymes. The base module being a pleiofunctional RNA that can adopt the cloverleaf structure is found today in various sequences without direct link with translation. One may conclude that “one should not change a winning secondary structure.” In a precellular context, a molecule with ss-tRNA characteristics (small ORF associated with cloverleaf structure) would be advantageous. Putatively, ss-tRNA-like molecules cumulating both tRNA and mRNA functions would have been the first molecules on Earth to support nonrandom protein synthesis.

The antiquity of ss-tRNAs can be discussed, and it is very likely that the TAR10 (and especially TAG) triplets played very early a critical role in the tertiary folding of some tRNAs. Their implication in translation termination would be an exaptation where firstly, they were part of a structural signal. Origin of ATR49 triplets is less clear perhaps tracing to the first endosymbiosis. Hence it would be apomorphic (derived character). Analyzes by taxa and tRNA species suggest a nonhomogeneous evolution. At the beginning of the RNA/protein world, it has quickly become essential to start peptide synthesis at particular codons and one cannot exclude that ATR49 was an ancestral state which would have not been retained as intergenetic spaces increased. Analyzes of known tRNAs of  $\alpha$ -proteobacteria and cyanobacteria could suggest that in organelles, ATR49 triplets would have been selected with genome reduction. Organelle genomes may be under increased pressure for size reduction with resulting overlaps (see, [100]). However, several features strongly suggest that overlapping genes are not a direct mechanism to substantially reduce genome size. Gene overlaps allow mtDNA genome compaction while avoiding the loss of tRNA genes [53]. Nevertheless, overlaps may allow a more efficient control in the regulation of gene expression, the regulatory pathways are simplified, and the number of proteins (and genes) required decreases [100]. Among others, short antiparallel overlaps may be involved in antisense regulatory mechanisms. Consequently, genomes with compact sizes enable putatively less flexible but more efficient physiologies.

The selection of tRNAs had to be done mainly on two seemingly opposite criteria, stability and plasticity, making it a kind of Swiss army knife of the RNA world. This explains that beyond their central role in protein synthesis, tRNAs have many other crucial functions. To date, it can be hypothesized that ss-tRNAs might regulate gene expression, stress responses, and metabolic processes. Indeed, *in silico* analyzes allowed to speculate that several overlapping sequences may code simultaneously for mRNAs and tRNAs in most of the metazoan

mt-genomes. These overlaps can have a variable (sometimes large) number of nts; however, when annotating their genomes, several authors voluntarily underestimated the number and the size of overlaps, speculating that there would be upstream abbreviated stop codons or downstream alternative start codons but most often without any direct demonstration so far. However, the high number of possible overlaps on the same strand in which the first in-frame complete stop codon or standard start codon are located at specific positions in the sequences of *trn* genes (TAR10 and ATR49, respectively) strongly suggest an exclusive relationship between obtaining tRNAs and translation of mRNAs and/or the development of repair system to keep the two genes functional due in some cases to co-evolution during several hundred MY. We can therefore speculate that *ss-trn* genes could allow true tRNA punctuation and initiation. Noted that *ss-tRNAs* seem to be hybrid molecules which would contain three essential coding or decoding informations in the form of nt triplets (i.e., anticodon and stop/start codons) which are all at least in part integrated into stem or loop; moreover, after the ATR49, nt triplets play the role of internal sense codons. To date, it is unclear what biochemical mechanism would allow to choose between different alternate cleavage sites, leading to the complete tRNA rather than to the mRNA or *vice versa*, but reduced/expanded proteins can be functional, and various processes including editing suggest this also for incomplete tRNAs. Hence, despite lacking experimental evidence, TAR10 and ATR49 triplets have probable roles, including regulation. Future analyzes of the processed bicistronic transcripts (tRNA/protein-encoding or the contrary) are required. Moreover, even if mt-*trn* genes are most often expressed at very low levels [53], only direct sequencing of tRNAs can validate transcription, epitranscriptomic maturation and can pinpoint nt modifications including post-transcriptionally edited positions. Purified native, or even synthetic, tRNAs should also be tested for their *in vitro* activity to confirm the functionality of aberrant transcripts. Similar experiments must be made on the flanking mRNAs and their products. If as we think, *ss-tRNAs* could play regulatory roles, initially experiments should compare stress and nonstress conditions.

Here, the bias for metazoan mtDNA does not allow for a complete picture of variation in the entire eukaryotic world, and protist mt-genomes should also be considered. Special attention should also be paid to noncanonical base pairings potentially formed by UAR10 and AUR49 nts, in perspective with tRNA structure and V-R length. Accounting for TAR10 and ATR49 triplet presences in the algorithms predicting tRNAs could improve mt-genome annotations, reducing numbers of false positives and negatives, and more accurately determine tRNA termini while accounting tRNA species, taxa, and genomic systems.

MtDNA plays a central role in apoptosis, aging, and cancer [13]. Moreover, mt-diseases are among the most common inherited metabolic and neurological disorders [101]. In addition, as new functions and new mechanisms of action of tRNAs are continuously discovered [1] and as *ss-trn* genes could affect the cellular dynamic during normal and stress conditions leading to pathologies, potential subtleties of action and regulation of these genes and products should be more thoroughly investigated.

## Conflict of interest

The authors declare no potential commercial or financial conflicts of interest.

## Author details

Eric Faure\* and Roxane Barthélémy

\*Address all correspondence to: eric.faure@univ-amu.fr

Aix Marseille University, CNRS, Centrale Marseille, I2M, Marseille, France

## References

- [1] Barciszewska MZ, Perrigue PM, Barciszewski J. tRNA-the golden standard in molecular biology. *Molecular BioSystems*. 2016;**12**:12-17. DOI: 10.1039/c5mb00557d
- [2] Li S, Mason CE. The pivotal regulatory landscape of RNA modifications. *Annual Review of Genomics and Human Genetics*. 2014;**15**:127-150. DOI: 10.1146/annurev-genom-090413-025405
- [3] Wende S, Platzer EG, Jühling F, Pütz J, Florentz C, Stadler PF, Mörl M. Biological evidence for the world's smallest tRNAs. *Biochimie*. 2014;**100**:151-158. DOI: 10.1016/j.biochi.2013.07.034
- [4] Bernt M, Braband A, Schierwater B, Stadler PF. Genetic aspects of mitochondrial genome evolution. *Molecular Phylogenetics and Evolution*. 2013;**69**:328-338. DOI: 10.1016/j.ympev.2012.10.020
- [5] Lang BF, Lavrov D, Beck N, Steinberg SV. Mitochondrial tRNA structure, identity and evolution of the genetic code. In: Bullerwell CE, editor. *Organelle Genetics*. Berlin: Heidelberg; 2012. pp. 431-474. DOI: 10.1007/978-3-642-22380-8\_17
- [6] Francklyn C, Perona JJ, Puetz J, Hou YM. Aminoacyl-tRNA synthetases: Versatile players in the changing theater of translation. *RNA*. 2002;**8**:1363-1372. DOI: 10.1017/S1355838202021180
- [7] Schimmel P, Giege R, Moras D, Yokoyama S. An operational RNA code for amino acids and possible relation to genetic code. *Proceedings of the National Academy of Sciences of the United States of America*. 1993;**90**:8763-8768. DOI: 10.1073/pnas.90.19.8763
- [8] Agris PF, Vendeix FA, Graham WD. tRNA's wobble decoding of the genome: 40 years of modification. *Journal of Molecular Biology*. 2007;**366**:1-13. DOI: 10.1016/j.jmb.2006.11.046
- [9] Rogalski M, Karcher D, Bock R. Superwobbling facilitates translation with reduced tRNA sets. 1. *Nature Structural & Molecular Biology*. 2008;**15**:192-198. DOI: 10.1038/nsmb.1370
- [10] Huot JL, Enkler L, Megel C, Karim L, Laporte D, Becker HD, Duchêne AM, Sissler M, Maréchal-Drouard L. Idiosyncrasies in decoding mitochondrial genomes. *Biochimie*. 2014;**100**:95-106. DOI: 10.1016/j.biochi.2014.01.004

- [11] Haen K, Pett W, Lavrov D. Parallel loss of nuclear-encoded mitochondrial aminoacyl-tRNA synthetases and mtDNA-encoded tRNAs in Cnidaria. *Molecular Biology and Evolution*. 2010;**27**:2216-2219. DOI: 10.1093/molbev/msq112
- [12] Barthélémy RM, Seligmann H. Cryptic tRNAs in chaetognath mitochondrial genomes. *Computational Biology and Chemistry*. 2016;**62**:119-132. DOI: 10.1016/j.compbiolchem.2016.04.007
- [13] Ladoukakis ED, Zouros E. Evolution and inheritance of animal mitochondrial DNA: Rules and exceptions. *Journal of Biological Research-Thessaloniki*. 2017;**24**:2. DOI: 10.1186/s40709-017-0060-4
- [14] Ojala D, Montoya J, Attardi G. tRNA punctuation model of RNA processing in human mitochondria. *Nature*. 1981;**290**:470-474. DOI: 10.1038/290470a0
- [15] Levinger L, Mörl M, Florentz C. Mitochondrial tRNA 3' end metabolism and human disease. *Nucleic Acids Research*. 2004;**32**:5430-5441. DOI: 10.1093/nar/gkh884
- [16] Kyriakou E, Chatzoglou E, Zouros E, Rodakis GC. The rRNA and tRNA transcripts of maternally and paternally inherited mitochondrial DNAs of *Mytilus galloprovincialis* suggest presence of a "degradosome" in mussel mitochondria and necessitate the re-annotation of the l-rRNA/CR boundary. *Gene*. 2014;**540**:78-85. DOI: 10.1016/j.gene.2014.01.080
- [17] Fiedler M, Rossmanith W, Wahle E, Rammelt C. Mitochondrial poly(A) polymerase is involved in tRNA repair. *Nucleic Acids Research*. 2015;**43**:9937-9949. DOI: 10.1093/nar/gkv891
- [18] Wolstenholme DR. Animal mitochondrial DNA: Structure and evolution. *International Review of Cytology*. 1992;**141**:173-216. DOI: 10.1016/S0074-7696(08)62066-5
- [19] Morrison D. How and where to Look for tRNAs in Metazoan Mitochondrial Genomes, and What you might Find when you Get there [Internet]. 2010. DOI: 10.1101/001875 Available from: <http://arxiv.org/abs/1001.3813> [Accessed: 2018-01-25]
- [20] Reichert A, Rothbauer U, Mörl M. Processing and editing of overlapping tRNAs in human mitochondria. *The Journal of Biological Chemistry*. 1998;**273**:31977-31984. DOI: 10.1074/jbc.273.48.31977
- [21] Faure E, Casanova JP. Comparison of chaetognath mitochondrial genomes and phylogenetical implications. *Mitochondrion*. 2006;**6**:258-262. DOI: 10.1016/j.mito.2006.07.004
- [22] Jühling F, Mörl M, Hartmann RK, Sprinzl M, Stadler PF, Pütz J. tRNADB 2009: Compilation of tRNA sequences and tRNA genes. *Nucleic Acids Research*. 2009;**37**:D159-D162. DOI: 10.1093/nar/gkn772
- [23] Lowe TM, Eddy SR. tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Research*. 1997;**25**:955-964. DOI: 10.1093/nar/25.5.0955
- [24] Dodd MS, Papineau D, Grenne T, Slack JF, Rittner M, Pirajno F, O'Neil J, Little CT. Evidence for early life in Earth's oldest hydrothermal vent precipitates. *Nature*. 2017;**543**:60-64. DOI: 10.1038/nature21377

- [25] Roger AJ, Muñoz-Gómez SA, Kamikawa R. The origin and diversification of mitochondria. *Current Biology*. 2017;**27**:R1177-R1192. DOI: 10.1016/j.cub.2017.09.015
- [26] Bengtson S, Rasmussen B, Ivarsson M, Muhling J, Broman C, Marone F, Stampanoni M, Bekker A. Fungus-like mycelial fossils in 2.4-billion-year-old vesicular basalt. *Nature Ecology & Evolution*. 2017;**1**:141. DOI: 10.1038/s41559-017-0141
- [27] Margulis L, Dolan MF, Guerrero R. The chimeric eukaryote: Origin of the nucleus from the karyomastigont in amitochondriate protists. *Proceedings of the National Academy of Sciences of the United States of America*. 2000;**97**:6954-6959. DOI: 10.1073/pnas.97.13.6954
- [28] Lewis LA. Hold the salt: Freshwater origin of primary plastids. *Proceedings of the National Academy of Sciences of the United States of America*. 2017;**114**:9759-9760. DOI: 10.1073/pnas.1712956114
- [29] Lang BF, Gray MW, Burger G. Mitochondrial genome evolution and the origin of eukaryotes. *Annual Review of Genetics*. 1999;**33**:351-397. DOI: 10.1146/annurev.genet.33.1.351
- [30] Le TH, Blair D, McManus DP. Mitochondrial genomes of parasitic flatworms. *Trends in Parasitology*. 2002;**18**:206-213. DOI: 10.1016/S1471-4922(02)02252-3
- [31] Roelants K, Bossuyt F. Archaeobatrachian paraphyly and pangaea and diversification of crown-group frogs. *Systematic Biology*. 2005;**54**:111-126. DOI: 10.1080/10635150590905894
- [32] Lavrov DV, Pett W. Animal mitochondrial DNA as we do not know it: Mt-genome organization and evolution in Nonbilaterian lineages. *Genome Biology and Evolution*. 2016;**8**:2896-2913. DOI: 10.1093/gbe/evw195
- [33] Gualberto JM, Milesina D, Wallet C, Niazi AK, Weber-Lotfi F, Dietrich A. The plant mitochondrial genome: Dynamics and maintenance. *Biochimie*. 2014;**100**:107-120. DOI: 10.1016/j.biochi.2013.09.016
- [34] Sloan DB, Alverson AJ, Chuckalovcak JP, Wu M, McCauley DE, Palmer JD, Taylor DR. Rapid evolution of enormous, multichromosomal genomes in flowering plant mitochondria with exceptionally high mutation rates. *PLoS Biology*. 2012;**10**:e1001241. DOI: 10.1371/journal.pbio.1001241
- [35] Rodríguez-Moreno L, González VM, Benjak A, Martí MC, Puigdomènech P, Aranda MA, Garcia-Mas J. Determination of the melon chloroplast and mitochondrial genome sequences reveals that the largest reported mitochondrial genome in plants contains a significant amount of DNA having a nuclear origin. *BMC Genomics*. 2011;**12**:424. DOI: 10.1186/1471-2164-12-424
- [36] Cognat V, Morelle G, Megel C, Lalande S, Molinier J, Vincent T, Small I, Duchêne AM, Maréchal-Drouard L. The nuclear and organellar tRNA-derived RNA fragment population in *Arabidopsis thaliana* is highly dynamic. *Nucleic Acids Research*. 2017;**45**:3460-3472. DOI: 10.1093/nar/gkw1122
- [37] Guo H, Tse LV, Nieh AW, Czornyj E, Williams S, Oukil S, Liu VB, Miller JF. Target site recognition by a diversity-generating retroelement. *PLoS Genetics*. 2011;**7**:e1002414. DOI: 10.1371/journal.pgen.1002414

- [38] Nosek J, Tomaska L, Burger G, Lang BF. Programmed translational bypassing elements in mitochondria: Structure, mobility, and evolutionary origin. *Trends in Genetics*. 2015;**31**: 187-194. DOI: 10.1016/j.tig.2015.02.010
- [39] Loong Chan V, Louie H, Joe A. Expression of the flgFG operon of *Campylobacter jejuni* in *Escherichia coli* yields an extra fusion protein. *Gene*. 1998;**225**:131-141. DOI: 10.1016/S0378-1119(98)00516-2
- [40] Nudler E, Mironov AS. The riboswitch control of bacterial metabolism. *Trends in Biochemical Sciences*. 2004;**29**:11-17. DOI: 10.1016/j.tibs.2003.11.004
- [41] Macé K, Gillet R. Origins of tmRNA: The missing link in the birth of protein synthesis? *Nucleic Acids Research*. 2016;**44**:8041-8051. DOI: 10.1093/nar/gkw693
- [42] Nan F, Feng J, Lv J, Liu Q, Fang K, Gong C, Xie S. Origin and evolutionary history of freshwater Rhodophyta: Further insights based on phylogenomic evidence. *Scientific Reports*. 2017;**7**:2934. DOI: 10.1038/s41598-017-03235-5
- [43] Maizels N, Weiner AM. Phylogeny from function: Evidence from the molecular fossil record that tRNA originated in replication, not translation. *Proceedings of the National Academy of Sciences of the United States of America*. 1994;**91**:6729-6734. DOI: 10.1073/pnas.91.15.6729
- [44] Xu WL, Yang Y, Wang YD, Qu LU, Zheng LL. Computational approaches to tRNA-derived small RNAs. *Non-Coding RNA*. 2017;**3**:2. DOI: 10.3390/ncrna3010002
- [45] Hou YM, Yang X. Regulation of cell death by transfer RNA. *Antioxidants & Redox Signaling*. 2013;**19**:583-594. DOI: 10.1089/ars.2012.5171
- [46] 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. DOI: 10.1016/j.gene.2014.02.021
- [47] Wang X, He C. Dynamic RNA modifications in posttranscriptional regulation. *Molecular Cell*. 2014;**56**:5-12. DOI: 10.1016/j.molcel.2014.09.001
- [48] Schimmel P. The emerging complexity of the tRNA world: Mammalian tRNAs beyond protein synthesis. *Nature Reviews. Molecular Cell Biology*. 2017;**19**:45-58. DOI: 10.1038/nrm.2017.77
- [49] Dethoff EA, Chugh J, Mustoe AM, Al-Hashimi HM. Functional complexity and regulation through RNA dynamics. *Nature*. 2012;**482**:322-330. DOI: 10.1038/nature10885
- [50] Lenglez S, Hermand D, Decottignies A. Genome-wide mapping of nuclear mitochondrial DNA sequences links DNA replication origins to chromosomal double-strand break formation in *Schizosaccharomyces pombe*. *Genome Research*. 2010;**20**:1250-1261. DOI: 10.1101/gr.104513.109
- [51] Raina M, Ibba M. tRNAs as regulators of biological processes. *Frontiers in Genetics*. 2014;**5**:171. DOI: 10.3389/fgene.2014.00171
- [52] Yona AH, Bloom-Ackermann Z, Frumkin I, Hanson-Smith V, Charpak-Amikam Y, Feng Q, Boeke JD, Dahan O, Pilpel Y. tRNA genes rapidly change in evolution to meet novel translational demands. *eLife*. 2013;**2**:e01339. DOI: 10.7554/eLife.01339

- [53] Doublet V, Helleu Q, Raimond R, Souty-Grosset C, Marcadé I. Inverted repeats and genome architecture conversions of terrestrial isopods mitochondrial DNA. *Journal of Molecular Evolution*. 2013;**77**:107-118. DOI: 10.1007/s00239-013-9587-7
- [54] Schild C, Hahn D, Schaller A, Jackson CB, Rothen-Rutishauser B, Mirkovitch J, Nuoffer JM. Mitochondrial leucine tRNA level and PTC1D1 are regulated in response to leucine starvation. *Amino Acids*. 2014;**46**:1775-1783. DOI: 10.1007/s00726-014-1730-2
- [55] Segovia R, Pett W, Trewick S, Lavrov DV. Extensive and evolutionarily persistent mitochondrial tRNA editing in velvet worms (phylum Onychophora). *Molecular Biology and Evolution*. 2011;**28**:2873-2881. DOI: 10.1093/molbev/msr113
- [56] Camps M. Modulation of ColE1-like plasmid replication for recombinant gene expression. *Recent Patents on DNA & Gene Sequences*. 2010;**4**:58-73. DOI: 10.2174/187221510790410822
- [57] Honda S, Kawamura T, Loher P, Morichika K, Rigoutsos I, Kirino Y. The biogenesis pathway of tRNA-derived piRNAs in *Bombyx* germ cells. *Nucleic Acids Research*. 2017;**45**:9108-9120. DOI: 10.1093/nar/gkx537
- [58] Halfmann R, Jarosz DF, Jones SK, Chang A, Lancaster AK, Lindquist S. Prions are a common mechanism for phenotypic inheritance in wild yeasts. *Nature*. 2012;**482**:363-368. DOI: 10.1038/nature10875
- [59] 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. DOI: 10.1016/j.gene.2009.06.006
- [60] van der Wijst MG, van Tilburg AY, Ruiters MH, Rots MG. Experimental mitochondria-targeted DNA methylation identifies GpC methylation, not CpG methylation, as potential regulator of mitochondrial gene expression. *Scientific Reports*. 2017;**7**:177. DOI: 10.1038/s41598-017-00263-z
- [61] Lorenz C, Lünse CE, Mörl M. tRNA modifications: Impact on structure and thermal adaptation. *Biomolecules*. 2017;**7**:2. DOI: 10.3390/biom7020035
- [62] Seligmann H. Phylogeny of genetic codes and punctuation codes within genetic codes. *Bio Systems*. 2015;**129**:36-43. DOI: 10.1016/j.biosystems.2015.01.003
- [63] NCBI. The Genetic Codes [Internet]. 2016. Available from: <https://www.ncbi.nlm.nih.gov/Taxonomy/Utils/wprintgc.cgi#SG4> [Accessed: 2018-01-25]
- [64] Osawa S, Jukes TH. Codon reassignment (codon capture) in evolution. *Journal of Molecular Evolution*. 1989;**28**:271-278. DOI: 10.1007/BF02103422
- [65] Wallis M. On the frequency of arginine in proteins and its implications for molecular evolution. *Biochemical and Biophysical Research Communications*. 1974;**56**:711-716. DOI: 10.1016/0006-291X(74)90663-9
- [66] Osawa S, Ohama T, Jukes TH, Watanabe K. Evolution of the mitochondrial genetic code. I. Origin of AGR serine and stop codons in metazoan mitochondria. *Journal of Molecular Evolution*. 1989;**29**:202-207. DOI: 10.1007/BF02100203

- [67] Temperley R, Richter R, Dennerlein S, Lightowlers RN, Chrzanowska-Lightowlers ZM. Hungry codons promote frameshifting in human mitochondrial ribosomes. *Science*. 2010;**327**:301. DOI: 10.1126/science.1180674
- [68] Di Giulio M. A comparison among the models proposed to explain the origin of the tRNA molecule: A synthesis. *Journal of Molecular Evolution*. 2009;**69**:1-9. DOI: 10.1007/s00239-009-9248-z
- [69] Di Giulio M. The origin of the tRNA molecule: Independent data favor a specific model of its evolution. *Biochimie*. 2012;**94**:1464-1466. DOI: 10.1016/j.biochi.2012.01.014
- [70] Fujishima K, Kanai A. tRNA gene diversity in the three domains of life. *Frontiers in Genetics*. 2014;**5**:142. DOI: 10.3389/fgene.2014.00142
- [71] Maizels N, Weiner AM. The genomic tag hypothesis: What molecular fossils tell us about the evolution of tRNA. In: Gesteland RF, Cech TR, Atkins JF, editors. *The RNA World*. 2nd ed. Cold Spring Harbor Laboratory Press; 1999. pp. 79-112. DOI: 10.1101/087969589.37.79
- [72] Sun FJ, Caetano-Anollés G. Evolutionary patterns in the sequence and structure of transfer RNA: A window into early translation and the genetic code. *PLoS One*. 2008;**3**:e2799. DOI: 10.1371/journal.pone.0002799
- [73] Weiner AM, Maizels N. The genomic tag hypothesis: Modern viruses as molecular fossils of ancient strategies for genomic replication, and clues regarding the origin of protein synthesis. *The Biological Bulletin*. 1999;**196**:327-330. DOI: 10.2307/1542962
- [74] Caetano-Anollés G, Wang M, Caetano-Anollés D. Structural phylogenomics retrodicts the origin of the genetic code and uncovers the evolutionary impact of protein flexibility. *PLoS One*. 2013;**8**:e72225. DOI: 10.1371/journal.pone.0072225
- [75] Yoshihisa T. Handling tRNA introns, archaeal way and eukaryotic way. *Frontiers in Genetics*. 2014;**5**:213. DOI: 10.3389/fgene.2014.00213
- [76] Larkin DC, Williams AM, Martinis SA, Fox GE. Identification of essential domains for *Escherichia coli* tRNA(Leu) aminoacylation and amino acid editing using minimalist RNA molecules. *Nucleic Acids Research*. 2002;**30**:2103-2113. DOI: 10.1093/nar/30.10.2103
- [77] Helm M, Brulé H, Friede D, Giegé R, Pütz D, Florentz C. Search for characteristic structural features of mammalian mitochondrial tRNAs. *RNA*. 2000;**6**:1356-1379. DOI: 10.1017/S1355838200001047
- [78] El Houmami N, Seligmann H. Evolution of nucleotide punctuation marks: From structural to linear signals. *Frontiers in Genetics*. 2017;**8**:36. DOI: 10.3389/fgene.2017.00036
- [79] Farias ST, Dos Santos Junior AP, Rêgo TG, José MV. Origin and evolution of RNA-dependent RNA polymerase. *Frontiers in Genetics*. 2017;**8**:125. DOI: 10.3389/fgene.2017.00125
- [80] Farias ST, Rêgo TG, José MV. Origin and evolution of the Peptidyl Transferase Center from proto-tRNAs. *FEBS Open Bio*. 2014;**8**:175-178. DOI: 10.1016/j.fob.2014.01.010

- [81] Noller HF. Evolution of protein synthesis from an RNA world. *Cold Spring Harbor Perspectives in Biology*. 2012;**4**:a003681. DOI: 10.1101/cshperspect.a003681
- [82] Root-Bernstein MM, Root-Bernstein RS. The ribosome as a missing link in evolution. *Journal of Theoretical Biology*. 2015;**367**:130-158. DOI: 10.1016/j.jtbi.2016.02.030
- [83] Mathews DH, Disney MD, Childs JL, Schroeder SJ, Zuker M, Turner DH. Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. *Proceedings of the National Academy of Sciences of the United States of America*. 2004;**101**:7287-7292. DOI: 10.1073/pnas.0401799101
- [84] Bean A, Weiner A, Hughes A, Itskovits E, Friedman N, Rando O. Genome-wide histone modification patterns in *Kluyveromyces Lactis* reveal evolutionary adaptation of a heterochromatin-associated mark. *bioRxiv* 039776. DOI: 10.1101/039776
- [85] Eigen M, Winkler-Oswatitsch R. Transfer-RNA, an early gene? *Die Naturwissenschaften*. 1981;**68**:282-292. DOI: 10.1007/BF01047470
- [86] Ariza-Mateos A, Gómez J. Viral tRNA mimicry from a biocommunicative perspective. *Frontiers in Microbiology*. 2017;**8**:2395. DOI: 10.3389/fmicb.2017.02395
- [87] Farias ST, Rêgo TG, José MV. A proposal of the proteome before the last universal common ancestor. *International Journal of Astrobiology*. 2016;**15**:27-31. DOI: 10.1017/S1473550415000464
- [88] Farias ST, Rêgo TG, José MV. tRNA core hypothesis for the transition from the RNA world to the ribonucleoprotein world. *Life (Basel)*. 2016b;**6**:15. DOI: 10.3390/life6020015
- [89] Ohnishi K. Evolutionary meanings of the primary and secondary structures of the "UR-RNA", a primitive possibly self-replicating ribo-organism commonly ancestral to tRNAs, 5S-rRNA and virusoids. In: Gruber B, Yopp JH, editors. *Symmetries in Science IV*. Boston, MA: Springer; 1990. pp. 147-176. DOI: 10.1007/978-1-4613-0597-2\_7
- [90] Aphasizhev R, Karmarkar U, Simpson L. Are tRNAs imported into the mitochondria of kinetoplastid protozoa as 5'-extended precursors? *Molecular and Biochemical Parasitology*. 1998;**93**:73-80. DOI: 10.1016/S0166-6851(98)00022-X
- [91] Soma A. Circularly permuted tRNA genes: Their expression and implications for their physiological relevance and development. *Frontiers in Genetics*. 2014;**5**:63. DOI: 10.3389/fgene.2014.00063
- [92] Naoko ABE, Hiroshi ABE. Rolling circle translation of circular RNA. *Seibutsu Butsuri*. 2017;**57**:5-10. DOI: 10.2142/biophys.57.005
- [93] Wolf YI, Koonin EV. On the origin of the translation system and the genetic code in the RNA world by means of natural selection, exaptation, and subfunctionalization. *Biology Direct*. 2007;**2**:198-223. DOI: 10.1186/1745-6150-2-14
- [94] Safro M, Klipcan L. The mechanistic and evolutionary aspects of the 2'- and 3'-OH paradigm in biosynthetic machinery. *Biology Direct*. 2013;**8**:17. DOI: 10.1186/1745-6150-8-17
- [95] Gwiazda S, Salomon K, Appel B, Muller S. RNA self-ligation: From oligonucleotides to full length ribozymes. *Biochimie*. 2012;**94**:1457-1463. DOI: 10.1016/j.biochi.2012.03.015

- [96] Hafez M, Burger G, Steinberg SV, Lang BF. A second eukaryotic group with mitochondrion-encoded tmRNA: In silico identification and experimental confirmation. *RNA Biology*. 2013;**10**:1117-1124. DOI: 10.4161/rna.25376
- [97] Harish A, Caetano-Anollés G. Ribosomal history reveals origins of modern protein synthesis. *PLoS One*. 2012;**7**:e32776. DOI: 10.1371/journal.pone.0032776
- [98] Megel C, Morelle G, Lalande S, Duchêne AM, Small I, Maréchal-Drouard L. Surveillance and cleavage of eukaryotic tRNAs. *International Journal of Molecular Sciences*. 2015;**16**:1873-1893. DOI: 10.3390/ijms16011873
- [99] Bhattacharyya S, Varshney U. Evolution of initiator tRNAs and selection of methionine as the initiating amino acid. *RNA Biology*. 2016;**13**:810-819. DOI: 10.1080/15476286.2016.1195943
- [100] Johnson ZI, Chisholm SW. Properties of overlapping genes are conserved across microbial genomes. *Genome Research*. 2004;**14**:2268-2272. DOI: 10.1101/gr.2433104
- [101] Sun H, Wang X, editors. *Mitochondrial DNA and Diseases*. Singapore: Springer; 2017. 230 p. DOI: 10.1007/978-981-10-6674-0

