Journal of Entomology and Zoology Studies

Available online at www.entomoljournal.com

J E Z S Journal of Entomology and Zoology Studies

E-ISSN: 2320-7078
P-ISSN: 2349-6800
JEZS 2019; 7(3): 304-315
© 2019 JEZS
Received: 09-03-2019
Accepted: 13-04-2019

Eric Faure
Aix Marseille University, CNRS, Centrale Marseille, I2M, UMR 7373, 13453 Marseille, France

Roxane-Marie Barthélémy
Aix Marseille University, CNRS, Centrale Marseille, I2M, UMR 7373, 13453 Marseille, France

Correspondence
Eric Faure
Aix Marseille University, CNRS, Centrale Marseille, I2M, UMR 7373, 13453 Marseille, France

Specific mitochondrial ss-tRNAs in phylum Chaetognatha

Eric Faure and Roxane-Marie Barthélémy

Abstract

Chaetognaths are marine invertebrate organisms that constitute a small phylum of very abundant animals. Chaetognaths exhibit several morphological, physiological, anatomical and molecular peculiarities. In all known chaetognath mitochondrial (mt) genomes, the number of genes encoding tRNAs is always small. In addition, the primary sequences and secondary structures of chaetognath tRNAs are not conventional. Several metazoan mt-*trn* genes (encoding tRNAs) exhibit nucleotide triplets corresponding to stop codons (TAG/TAA) and/or start codons (ATG/ATA) at specific conserved positions. The products of genes that bear one or both types of these codons are known as ss-tRNAs (ss for stop/start). Indirect analyses strongly suggest that in chaetognath mt genomes, some of these start and stop codons could be functional. Moreover, taking these codons into account in the algorithms predicting tRNAs makes it possible to identify and correct several potential annotation errors. Furthermore, a type of ss-*trn* gene appears to have emerged by duplication in order Aphragmophora.

Keywords: Chaetognatha, mitochondrion, ss-tRNA, start codon, stop codon, overlapping genes

Introduction

Chaetognaths are a small phylum of marine invertebrates with a size range of 2-120 mm. This phylum comprises approximately 130-140 reported species and is subdivided into two orders, Phragmophora and Aphragmophora [1], based on the addition of the transverse musculature (phragma) in the body. Chaetognaths live in various habitats, being found in the open sea, on or near the bottom, from polar to tropical regions, at all depths, in dark submarine caves, in the interstitial milieu and even around hydrothermal vent sites [2]. Chaetognaths were long considered strictly carnivorous but different aspects of their feeding biology have led to the conclusion that they feed primarily on dissolved and fine particulate matter and not on prey [3]. Despite their soft bodies, early Cambrian (~540–520 million years ago) chaetognath fossils with morphologies almost identical to recent forms have been discovered in China, suggesting a Precambrian origin [4]. Their exact phylogenetic position remains controversial [5-7]. Besides, the two longest species of giant viruses identified to date have been found in this taxon [8]. They also exhibit numerous other peculiarities including certain characteristics of their nuclear and mitochondrial (mt) genomes (*e.g.*, [9-13]).

Until 2016, it was thought that chaetognath mitogenomes bore no [14] or only one *trn* gene (specifying the tRNA-Met) [15-17], rather than the 22 such genes found in most of the other invertebrates that employ the same mt genetic code. However, four studies invalidated the view that all chaetognath mitogenomes include zero or a single *trn* gene [12, 18-20]. Nevertheless, in chaetognaths, most of the mt-*trn* sequences are non-canonical, and in a given chaetognath mt genome, only approximately half a dozen *trn*-like sequences have been found [12]. As 22 *trn* genes are required, the missing tRNAs may be nuclear encoded and imported from the cytosol, or they may be highly post-transcriptionally modified or "bizarre" *trn* genes that have escaped our detection [21].

In 2004, searching for chaetognath mt-*trn* genes [16], we incidentally observed that tRNAs bear nucleotide (nt) triplets corresponding to stop or start codons at precise conserved positions. The products of these genes that bear one or two types of these codons are referred to as ss-tRNAs (ss for stop/start) [22]. As most of the studies on this topic focus on DNA sequences, these codons are usually annotated TAR or ATR instead of UAR or AUR (R for purine). While numerous *trn* genes exhibiting nt triplets corresponding to stop codons (TAG/TAA) at specific conserved positions have been found in all taxa and genomic systems examined to date, relatively high frequencies of start codons (ATG/ATA) occur principally in fungal/metazoan mt-*trn* genes [22]. The last nucleotide of these triplets is the first nucleotide involved in the 5'-D- or 5'-T-stem, so they are referred to as TAR10 and ATR49, respectively. In the chaetognath mt genomes, which are "very constrained", the length of the intergenic regions between two protein-encoding genes is quite often compatible with the average distance found between the TAR10 codon and the ATR49 codon; thus, the main focus of the present study is to analyse the ss-*trn* sequences found in chaetognath mtDNAs for the first time and to propose plausible scenarios concerning the origin of these genes.

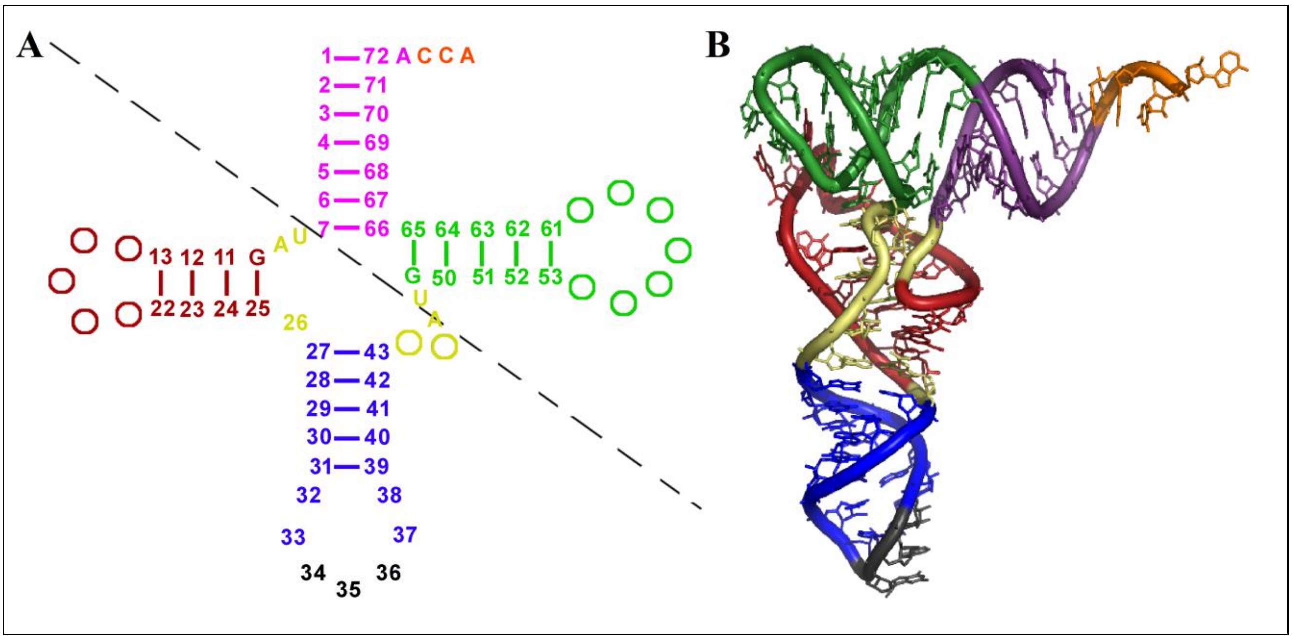


Fig 1: Typical cloverleaf secondary structure of a metazoan mt-ss-tRNA (**A**) with a 3D image of an L-shaped tRNA (**B**). In the 2D structure, standard numbering was applied. The first two nucleotides of the variable region and those of the D-loops and T-loops are represented by circles. The diagonal dashed line indicates the approximate separation between the "top half" and the "cherry-bob"/"bottom half". Nucleotide types are given for UAG10 and AUG49 triplets, the discriminator base (preferentially an A), and the CCA tail at the 3′-end. Short lines connect nucleotides undergoing pairing within stems. Colouring: acceptor-stem, purple; D-arm, red; anticodon-arm, blue with the anticodon in black; T-arm, green; and CCA tail, orange. The yellow segments in descending order of size, represent the variable region (connector 2), connector 1 and nt 26. Figure adapted from Faure and Barthélémy [22] and 3D image reproduced with the kind permission of Prof. N.R. Voss (Roosevelt University, Ill., USA)

2. Materials and Methods

2.1 Mitogenomic sequences

This study analysed complete or almost complete chaetognath mtDNAs available in GenBank (Table 1). For the mt genomes of *Sagitta ferox* and *Pterosagitta draco,* the GenBank staff noted that they were "unable to verify sequence and/or annotation provided by the submitter"; however, these mtDNAs were still included in the analyses even if reannotation was carried out, and possible sequence problems have been discussed.

Table 1: List of the chaetognath mt genomes used in this study.

Orders	Species	Accession numbers	Abbre-viations	Complete genomes	Refe-rences
Aphrag-mophora	*Decipisagitta decipiens*	AP011546	D.d.	yes	[17]
	Sagitta elegans.1	KP899780	S.el.1	no	[13]
	Sagitta elegans.2 (lineage G)	KP899787	S.el.2	no	[13]
	Sagitta enflata	AP011547	S.en.	yes	[17]
	Sagitta ferox	KT818830	S.f.	yes	[18]
	Sagitta setosa	KP899756	S.s.	no	[13]
	Pterosagitta draco	KU507531	P.d.	yes	[20]
	Zonosagitta nagae.1	AP011545	Z.n.1	yes	[17]
	Zonosagitta nagae.2	KF051939	Z.n.2	yes	[19]
Phrag-mophora	*Paraspadella gotoi*	AY619710	P.g.	yes	[15]
	Spadella cephaloptera	AY545549	S.c.	yes	[14]

2.2 Bio-informatic analyses

Some of the *trn* genes were detected by the authors reporting the sequences but were not identified as ss-*trn* genes. Otherwise, previously unidentified tRNAs were detected using tRNAscan-Se 1.21 [23] or via alignments of intergenic regions with known *trn*-sequences or BLASTn searches. Multiple sequence alignments were performed with ClustalW software [24]; the alignments were manually refined at positions where clear misalignments were produced by the algorithm (*i.e.*, they were verified for the conservation of important features and sites, principally in 2D alignments). Analyses of possible sequence homology with non-chaetognath *trn* genes were performed using two databases that include primary sequences and graphical representations of tRNA 2D structures: tRNAdb and mitotRNAdb [25].

3. Results

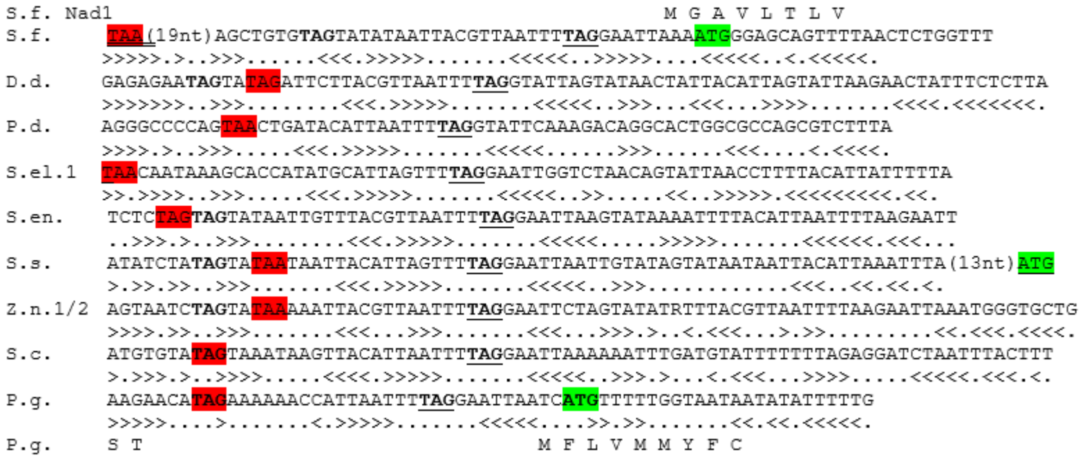
3.1 *trnM* genes

```
A
        >>>>>..>>>>....<<<<.>>>>>.......<<<<<....>>>....<<<..>>>>>....<<<<<.....<<<<<.
Z.n.1/2 ATTTTTGTAAGCTATAAAAGCTATTGAGTTCATACCTCAAAAAAGACGAAAGTCCAAAAATTTTTGTTTTGTTTATAAAAG
        >>>>>>>..>>>>....<<<<.>>>>>......<<<<<...>>>...<<<<<<<<<<.

            >>>>...>>>>......<<<<.>>>>>......<<<<<...>>.....<<..>>>>......<<<<....<<<<.
D.d. TAAGTACTTATATAAGCTAAACATAAGCTAATGAGTTCATACCTCAAAAATGGTAAATCCTATAAGTTTCTGTCTTGGACAATAAA
          >>>>>>>..>>>>......<<<<..>>>>......<<<<.....>>.....<<<<<<<<<.

      >>>>>>>..>>>>.....<<<<.>>>>>......<<<<<...>>>....<<<..>>>>>....<<<<<.....<<<<<<<.
S.en. ATTTTTGTAAGCTATGTTAAGCTATTGAGTTCATACCTCAAAAAAGATTTTTATCCAAAAATTTTTGTTTTGACGTCAAAAATT
      >>>>>>>..>>>>.....<<<<.>>>>>......<<<<<...>>>....<<<<<<<<<<.

       >>>>>.>..>>>>....<<<<.>>>>>......<<<<<...>>>>...<<<<>>.>......<.<<<<.<<<<<.
S.s. TAATTTTAGTAAGCTATATAAGCTATTGAGTTCATACCTCAAAAAAGGGGGAACCCCGTAAATACGTTTCTACCAAAAATC
         >>>>...>>>>....<<<<.>>>>>......<<<<<.....>>>>..<<<<.<<<<.

S.el.1   TAACTCTTGTAAGCTATATAAGCTATTAAGTTCATACCTTAAATAAGGACCAGTCCCGTGAGTT
                >>>>.>>..>>>>....<<<<.>>>>>......<<<<<...>>>>..<<<<<<<.<<<<.
S.el.2 TAA(71nt)ACTCTTGTAAGCTATTTAAGCTATTAAGTTCATACCTTAAAAAAGGATACGTCCCGTGAGTT
                >>>>.>>..>>>>....<<<<.>>>>>......<<<<<...>>>>..<<<<<<.<<<<.
S.f.            ATTTTTGTAAGCTATATAAGCTATTGAGTTCATACCTCAAAAAAGGCCTCGCCCGGAAATATTTGTTTTAA
                >>>>>>>..>>>>....<<<<.>>>>>......<<<<<...>>...<<<<<<<<<.
S.c.
TTTTTAATAAGCTAAAAAAAGCTATCGAGTTCATACCTCGGAAATGAGAAATCTTGTTTTTTTGTTTCGGAAA(111nt TAA
                >>>>.>>..>>>>.....<<<<.>>>>>......<<<<<...>>>>>>>...<<<<<<<...<<.<<<<.
P.g.           TTTTTTAATAAGCTAAAAAAGCTATTGAGTTCATACCTCAAAAATGAGAAATCTTAATTTTTTGTTTTAGAAC
                >>>>>>..>>>>....<<<<.>>>>>......<<<<<...>>>>>>>....<<<<<<<..<<<<<<<.

B
Z.n.1/2 ATTTTTGTAAGCTATAA--
AAGCTATTGAGTTCATACCTCAAAAAAGACGAAAGTCCAAAAATTTTTGTTTTGTTTATAAAAG--
D.d. -CTTATATAAGCTAAACATAAGCTAATGAGTTCATACCTCAAAAATGG-TAAATCCTATAAGTTTCTGTCTTGGACAATAAA---
S.en.   ATTTTTGTAAGCTATGTT-
AAGCTATTGAGTTCATACCTCAAAAAAGATTTTTATCCAAAAATTTTTGTTTTGACGTCAAAAATT
S.s. ATTTTAGTAAGCTATAT--AAGCTATTGAGTTCATACCTCAAAAAAGGGGGAACCCCGTAAATACGTTTCTACCAAAAATC
S.el.1 ACTCTTGTAAGCTATAT--AAGCTATTAAGTTCATACCTTAAATAAGGACCAGTCCCGTGAGTT--------------------
-
S.el.2 ACTCTTGTAAGCTATTT--AAGCTATTAAGTTCATACCTTAAAAAAGGATACGTCCCGTGAGTT--------------------
-
S.f. ATTTTTGTAAGCTATAT--AAGCTATTGAGTTCATACCTCAAAAAAGGCCTCG-CCCGGAAATA---------------------
S.c. TTTTTAATAAGCTAAAAA-AAGCTATCGAGTTCATACCTCGGAAATGA-GAAATCTTGTT--TTTTTGTTTCGGAAA--------
P.g. -TTTTAATAAGCTAAAAA--AGCTATTGAGTTCATACCTCAAAAATGA-GAAATCTTAAT--TTTTTGTTTTAGAAC--------
          *    *******      *****   ***********   * * *               *
```

Fig 2: Alignments of chaetognath mt-*trnM*-like sequences with their secondary (**A**) and primary structures (**B**). For abbreviations see Table 1. All the *trnM* genes are found within the *cox3-rrnS* gene block except for lineage G of *S. elegans* (*S.el.2*), which is located within the *nd3-rrnS* gene block. The positions of the various genes in their respective mtDNAs are as follows: *Z. nagae* (for long sequences, 3643-3720 and 5203-5281 for Z.n.1 and Z.n.2, respectively; for short sequences, 3640-3701 and 5201-5262 for Z.n.1 and Z.n.2, respectively), *D. decipiens* (2892-2970 and 2890-2952 for long and short sequences, respectively), *S. enflata* (3647-3730 and 3647-3709 for long and short sequences, respectively), *S. setosa* (4756-4834 and 2379-2437 for long and short sequences, respectively), *S. elegans.1* (3395-3456), *S. elegans.2* (826-887), *S. ferox* (6182-6242), *S. cephaloptera* (11182-11254) and *P. gotoi* (5157-5216). The TAA10 and ATG49 triplets, and the anticodons (CAT) are indicated in boldface. The pseudo-ATG49 of the *D. decipiens* sequence and all the nucleotide flanking regions are underlined. The complete and incomplete stop codons given by the authors who reported the sequences are highlighted in red. Moreover, when the authors noted an incomplete stop codon, the first in-frame complete stop codon is also highlighted in red. In (**A**), the lines with the ">" and "<" symbols correspond to the secondary structure predictions; in addition, for some sequences, two possible lengths for the same *trn* type are proposed and the upper and lower lines correspond to the 2D structures of the long and short *trn* sequences, respectively. In (**B**), asterisks (*) indicate conserved nucleotides. NB, the presence of *trnM* genes was not detected by the authors who reported the sequences of *S. cephaloptera*, *S. elegans* and *S. ferox*; similarly, the possible short *trnM* sequences of *S. enflata* and *D. decipiens* are proposed by our group.

The first *trn* genes found in chaetognaths were those specifying methionyl mt-tRNA [15-17]. The sizes of the *trnM* genes may differ slightly from those proposed by the authors who reported the sequences because nucleotide alignments and prediction of secondary structure allow the pairings to be optimized, especially at the 5'- and 3'-termini of these genes (Fig 2). To date, only the *trnM* genes of order Phragmophora sequenced to date exhibit a canonical structure [15, 16], in contrast to those of order Aphragmophora. In the latter order, in the same region of *Z. nagae* mtDNA (Fig 2A), the authors who reported one sequence assumed that the *trnM* gene size would be long with an extra loop at the level of the variable region (*Z. nagae.1* [17]), whereas according to the authors who reported another complete mtDNA sequence of the same species, the *trnM* gene size is shorter (*Z. nagae.2* [19]). The *cox3-rrnS* gene block has been found in the mtDNAs of all chaetognath species except for *S. elegans* lineage "G" named by Marletaz *et al.* [13] (the *S. elegans.2* mtDNA belongs to this lineage), and *P. draco* [20]. In all the chaetognath mtDNAs exhibiting the *cox3-rrnS* gene block, a putative *trnM* sequence has been detected between these two genes. The authors who reported the annotation of the *S. ferox* mt-DNA did not mention the two *rrn* genes [18]; however, a rapid analysis suggested that the *rrnS* and *rrnL* genes are located between *cox3* and *nad6* genes, and *cytb* and *cox1* genes, respectively. The *trnM* gene of *S. ferox*, which was previously incorrectly identified as a *trnH* gene by the authors who reported the sequence, is located within the *cox3-rrnS* block, as expected. In the sequence of *S. elegans* belonging to lineage G, the *trnM* gene is found within the *nad3-rrnS* gene block (see S.el2 in Fig 2) and the stop codon of the *nad3* gene is located considerably upstream of the *trnM* sequence (Fig 2A). It must be noted that the TAA49 triplets of *trnM* genes are never used as stop codons. The TAA49 triplets are in-frame for three sequences (*D. decipiens*, *S. elegans.1*, *S. setosa*) but are preceded by one or two (for *S. setosa*) in-frame stop codon(s). For three sequences (*S. ferox*, *P. gotoi* and *S. cephaloptera*), the first in-frame complete stop codons are found at the 3'-termini of the *trnM* gene or downstream of this gene. However, Blast analysis of chaetognath Cox3 proteins strongly suggested that *trnM* sequences are not included in the *cox3* sequences (data not shown). So, at least for *S. cephaloptera*, *S. ferox* and *P. gotoi* mt DNAs, there is an incomplete stop codon in the corresponding mRNA upstream the *trnM* gene (or at the very beginning of this sequence); this had been proposed by authors who reported the sequences, either based on those of the *trnM* gene [15] or intuitively [14]. Moreover, in three *trnM* genes there is an ATG49 triplet, but because the downstream genes encode 12S rRNA, this triplet cannot be used as a start codon. In addition, no *trnM*-like gene has been found in the *P. draco* mt genome.

Blast analyses using the "cherry-bob" sequence of *S. cephaloptera* (5'-TAAGCTAAAAAAAGCTATCGAGTTCATACCTCGGAAATG-3') as the query against all nucleotide sequences in GenBank revealed that at least all the first 300 sequences specify mt-tRNA-Met (data not shown). The "cherry-bob" structure begins at TAR10 and ends at the ATR49 triplet; this sequence could produce a conformation exhibiting two loops linked by a forked-stem structure, roughly resembling a pair of cherries, and could have played a significant role in tRNA evolution [22]. Moreover, analysis of the 166 sequences (chaetognath sequences not included) that exhibited a percentage of identity higher than 70% showed that all the mt-*trnM* genes belong to bilaterian animal species (2.4% Deuterostomia, 97.6% Protostomia including 83% insects; however, this result is biased because the last taxon is overrepresented in the database). In addition, analyses using the tRNAdb and mitotRNAdb databases showed that the sequence TAAGCT (TAA10 + 5'-D-stem)/-/AGCT(3'-D-stem)/-/TTCATAC (Ac-loop)/-/ATR49 has been found only in mt-*trnM* genes, and these genes come from only bilaterian animals. These genes represent 8.9% of the mt-*trnM* genes of bilaterian animals present in the mitotRNAdb database, and 8.5% of these genes belong to Protostomia (all are insects). Moreover, most of the sequences found in all of these analyses specify initiator methionyl-tRNAs.

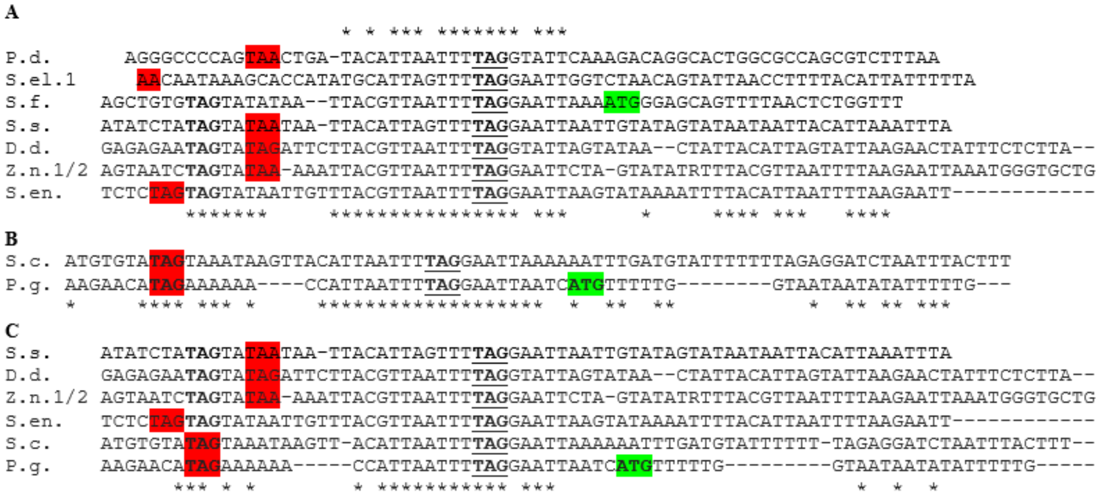


```
S.f. Nad1                                                        M G A V L T L V
S.f.     TAA(19nt)AGCTGTGTAGTATATAATTACGTTAATTTTAGGAATTAAAATGGGAGCAGTTTTAACTCTGGTTT
         >>>>>.>..>>>......<<<.>>>>>......<<<<<..>>>>>....<<<<<..<.<<<<<.
D.d.     GAGAGAATAGTATAGATTCTTACGTTAATTTTAGGTATTAGTATAACTATTACATTAGTATTAAGAACTATTTCTCTTA
         >>>>>>>..>>>........<<<.>>>>>.......<<<<<..>>>...<<<...>>>>.......<<<<.<<<<<<<.
P.d.     AGGGCCCCAGTAACTGATACATTAATTTTAGGTATTCAAAGACAGGCACTGGCGCCAGCGTCTTTA
         >>>>.>..>>>.....<<<.>>>>>.......<<<<<......>>>......<<<....<.<<<<.
S.el.1   TAACAATAAAGCACCATATGCATTAGTTTTAGGAATTGGTCTAACAGTATTAACCTTTTACATTATTTTTA
         >>.>>>>..>>>.....<<<.>>>>>.......<<<<<..>>.....<<..>>>>>.<<<<<<<<<.<<.
S.en.     TCTCTAGTAGTATAATTGTTTACGTTAATTTTAGGAATTAAGTATAAAATTTTACATTAATTTTAAGAATT
          ..>>>.>..>>>.........<<<.>>>>>.......<<<<<.....>>>>>.......<<<<<<.<<<...
S.s.      ATATCTATAGTATAATAATTACATTAGTTTTAGGAATTAATTGTATAGTATAATAATTACATTAAATTTA(13nt)ATG
          >.>>.>>..>>>........<<<.>>>>>.......<<<<<..>>>.............<<<..<<.<<.<.
Z.n.1/2   AGTAATCTAGTATAAAAATTACGTTAATTTTAGGAATTCTAGTATATRTTTACGTTAATTTTAAGAATTAAATGGGTGCTG
          >>>>.>>..>>>........<<<...>>>.......<<<...>>>.>..<.<<<...>.>>..........<<.<<<.<<<<.
S.c.      ATGTGTATAGTAAATAAGTTACATTAATTTTAGGAATTAAAAAATTTGATGTATTTTTTTAGAGGATCTAATTTACTTT
          >.>>>.>..>>>>.....<<<<.>>>>>.......<<<<<..>>>.>...<.<<<...>>>>.....<<<<<.<<<.<.
P.g.      AAGAACATAGAAAAAACCATTAATTTTAGGAATTAATCATGTTTTTGGTAATAATATATTTTTG
          >>>>>....>.......<.>>>>>.......<<<<<....>>.>>.......<<.<<.<<<<<.
P.g.      S T                                    M F L V M M Y F C
```

Fig 3: *trnL1*-like sequences of chaetognaths with their 2D structures. In mtDNAs, these sequences are found within those of the *cox2-nd1* gene block in *S. ferox* (3201-3265), *D. decipiens* (11035-11113), *P. draco* (3331-3396), *S. elegans.1* (352-422), *S. enflata* (654-724), *S. setosa* (11015-11084), and *Z. nagae.1* and *Z. nagae.2* (640-720 and 2201-2081, respectively), and within the *nd1-nd3* gene block in *P. gotoi* (3149-3212) and *S. cephaloptera* (9070-9148). TAR10, ATG49 and the anticodon are presented in boldface, and the anticodon is underlined. The first in-frame stop codon is highlighted in red, whereas the assumed start codon is highlighted in green. For the *Z. nagae.1/2* sequence, the letter R represents a purine nucleotide. For *P. gotoi*, the amino acids of the COOH terminus and of the NH_2 terminus of the Nad1 and Nad3 sequences, respectively, are noted. For *S. ferox*, the NH_2 terminus of Nad1 sequence is also noted. Amino acids are shown in single-letter code.

3.2 *trnL1* genes

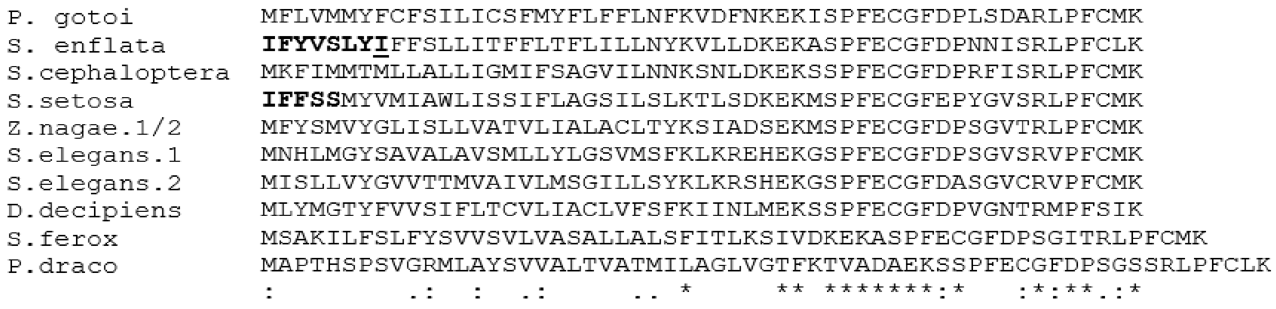
Within chaetognath mtDNAs, seven putative leucinyl(1) ss-*trn* genes have been found as well as two "classical" *trn*-like sequences (Fig 3). All the mtDNAs of the species of order Aphragmophora that harbour the *cox2-nd1* gene block have this *trn* gene (however, this excludes lineage G of *S. enflata*). The two other ss-*trnL1* genes belong to the *S. cephaloptera* and *P. gotoi* mt genomes (order Phragmophora) and are found between the *nd1* and *nd3* genes. Three ss-*trnL1* genes exhibit an extra loop between the Ac- and T-stem-loop-stem (*D. decipiens*, *Z. nagae* and *S. cephaloptera*). Only the sequences of *P. draco* and *S. enflata1* do not contain the TAG10 triplet, but this codon is a possible stop codon only for the two species belonging to order Phragmophora. Concerning the sequences of order Aphragmophora, in five cases, the TAG10 triplet is out of frame, and for *S. enflata*, there are two stop codons in tandem, with TAG10 in second position. The intergenic region between the *cox2* and *nd1* genes in species of order Aphragmophora is generally too long to also exhibit an ATR49 start codon, which is the case for *S. cephaloptera* mtDNA as well. In contrast, a complete ss-tRNA containing both the putative stop codon TAR10 and the start codon ATR49 has been found in *P. gotoi* mt genome. This tRNA does not exhibit an extra loop, and only a complementary base pair has been found in the D-arm; however, changes could occur post-transcriptionally.



```
A
                            * * *** ******* ***
P.d.       AGGGCCCAGTAACTGA-TACATTAATTTTAGGTATTCAAAGACAGGCACTGGCGCCAGCGTCTTTAA
S.el.1     AACAATAAAGCACCATATGCATTAGTTTTAGGAATTGGTCTAACAGTATTAACCTTTTACATTATTTTTA
S.f.    AGCTGTGTAGTATATAA--TTACGTTAATTTTAGGAATTAAAATGGGAGCAGTTTTAACTCTGGTTT
S.s.    ATATCTATAGTATAATAA-TTACATTAGTTTTAGGAATTAATTGTATAGTATAATAATTACATTAAATTTA
D.d.    GAGAGAATAGTATAGATTCTTACGTTAATTTTAGGTATTAGTATAA--CTATTACATTAGTATTAAGAACTATTTCTCTTA--
Z.n.1/2 AGTAATCTAGTATAA-AAATTACGTTAATTTTAGGAATTCTA-GTATATRTTTACGTTAATTTTAAGAATTAAATGGGTGCTG
S.en.   TCTCTAGTAGTATAATTGTTTACGTTAATTTTAGGAATTAAGTATAAAATTTTACATTAATTTTAAGAATT------------
               *******     **************** ***      *     **** ***   ****
B
S.c. ATGTGTATAGTAAATAAGTTACATTAATTTTAGGAATTAAAAAATTTGATGTATTTTTTTAGAGGATCTAATTTACTTT
P.g. AAGAACATAGAAAAAA----CCATTAATTTTAGGAATTAATCATGTTTTTG--------GTAATAATATATTTTTG---
     *     **** *** *     *******************  *  **  **           *  ** ** ***
C
S.s.    ATATCTATAGTATAATAA-TTACATTAGTTTTAGGAATTAATTGTATAGTATAATAATTACATTAAATTTA
D.d.    GAGAGAATAGTATAGATTCTTACGTTAATTTTAGGTATTAGTATAA--CTATTACATTAGTATTAAGAACTATTTCTCTTA--
Z.n.1/2 AGTAATCTAGTATAA-AAATTACGTTAATTTTAGGAATTCTA-GTATATRTTTACGTTAATTTTAAGAATTAAATGGGTGCTG
S.en.   TCTCTAGTAGTATAATTGTTTACGTTAATTTTAGGAATTAAGTATAAAATTTTACATTAATTTTAAGAATT------------
S.c.    ATGTGTATAGTAAATAAGTT-ACATTAATTTTAGGAATTAAAAAATTTGATGTATTTTTT-TAGAGGATCTAATTTACTTT--
P.g.    AAGAACATAGAAAAAA-----CCATTAATTTTAGGAATTAATCATGTTTTTG---------GTAATAATATATTTTTG-----
               *** * *        * *********** ***                         *  *  *
```

Fig 4: Alignments of the primary sequences of genes specifying tRNA-Leu1. The alignments of the sequences of (**A**), (**B**) and (**C**) correspond to sequences from species of order Aphragmophora, order Phragmophora, and both orders Aphragmophora and Phragmophora, respectively. In (**A**), asterisks indicate conserved nucleotides in all the sequences on the upper line or only those of *D. decipiens*, *S enflata*, *S. setosa*, *Z. nagae.1* and *Z. nagae.2* on the lower line, respectively. For other data, see Fig 3.

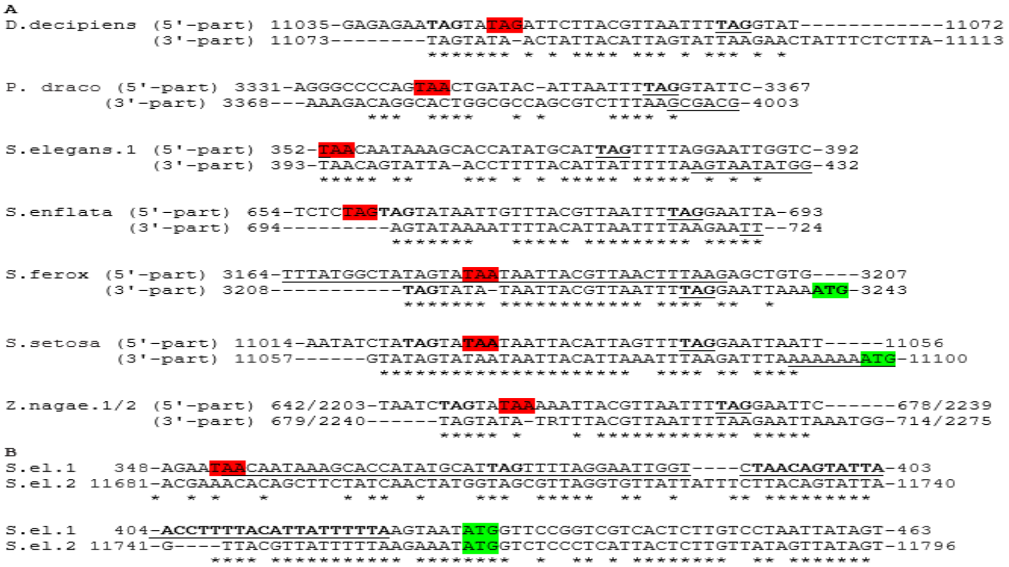
Alignments of the primary sequences of genes specifying tRNA-Leu1 strongly suggest that at least the sequences of *D. decipiens*, *S enflata*, *S. setosa*, *Z. nagae.1* and *Z. nagae.2*, which belong to order Aphragmophora, come from a common ancestral sequence (Fig 4A). The two sequences of order Phragmophora also exhibit a relatively high level of homology which each other (Fig 4B). However, in these last two alignments, the homology after the Ac-arm could be an artefact due to the high AT content. Alignments of *trnL1* sequences from the two orders exhibit homology only at the level of TAG10 and the Ac-5'-stem-loop (Fig 4C); however, the high AT content prompts us to remain cautious. In the mtDNAs of two species (*P. draco* [20] and *S. elegans.1*), *trnL1*-like sequences have been found, but there is no TAR10 triplet and sequence alignment with the *trnL1* genes of other Aphragmophora reveals a low level of nucleotide conservation. Additionally, analyses using the mitotRNAdb database show that the sequence TAGTAT(TAG10 + 5'-D-stem)/-/TTTAGGN(Ac-loop) has only been found in some mt-*trnL1* genes (19 of 1374 total mt-tRNA-Leu1 sequences) and these 19 tRNAs belong only to bilaterian animals; however, the database exhibits a strong bias for Metazoa. A total of four, one, nine and five tRNAs have been found among Cephalochordata, Xenacoelomorpha, Nematoda and Mollusca mtDNAs, respectively.



```
P. gotoi        MFLVMMYFCFSILICSFMYFLFFLNFKVDFNKEKISPFECGFDPLSDARLPFCMK
S. enflata      IFYVSLYIFFSLLITFFLTFLILLNYKVLLDKEKASPFECGFDPNNISRLPFCLK
S.cephaloptera  MKFIMMTMLLALLIGMIFSAGVILNNKSNLDKEKSSPFECGFDPRFISRLPFCMK
S.setosa        IFFSSMYVMIAWLISSIFLAGSILSLKTLSDKEKMSPFECGFEPYGVSRLPFCMK
Z.nagae.1/2     MFYSMVYGLISLLVATVLIALACLTYKSIADSEKMSPFECGFDPSGVTRLPFCMK
S.elegans.1     MNHLMGYSAVALAVSMLLYLGSVMSFKLKREHEKGSPFECGFDPSGVSRVPFCMK
S.elegans.2     MISLLVYGVVTTMVAIVLMSGILLSYKLKRSHEKGSPFECGFDASGVCRVPFCMK
D.decipiens     MLYMGTYFVVSIFLTCVLIACLVFSFKIINLMEKSSPFECGFDPVGNTRMPFSIK
S.ferox         MSAKILFSLFYSVVSVLVASALLALSFITLKSIVDKEKASPFECGFDPSGITRLPFCMK
P.draco         MAPTHSPSVGRMLAYSVVALTVATMILAGLVGTFKTVADAEKSSPFECGFDPSGSSRLPFCLK
                :        .:  :  .:     .. *     ** *******:*   :*:**.:*
```

Fig 5: Alignments of the NH_2 terminus of Nad3 amino acid sequences of several chaetognath species. For *S. enflata* and *S. setosa*, the residues in boldface have been added to the sequences. In the *S. enflata* sequence, the isoleucine residue, which is underlined, corresponds to a methionine in the original sequence. Asterisks indicate conserved residues, double dots indicate similar residues, and single dots indicate like residues.

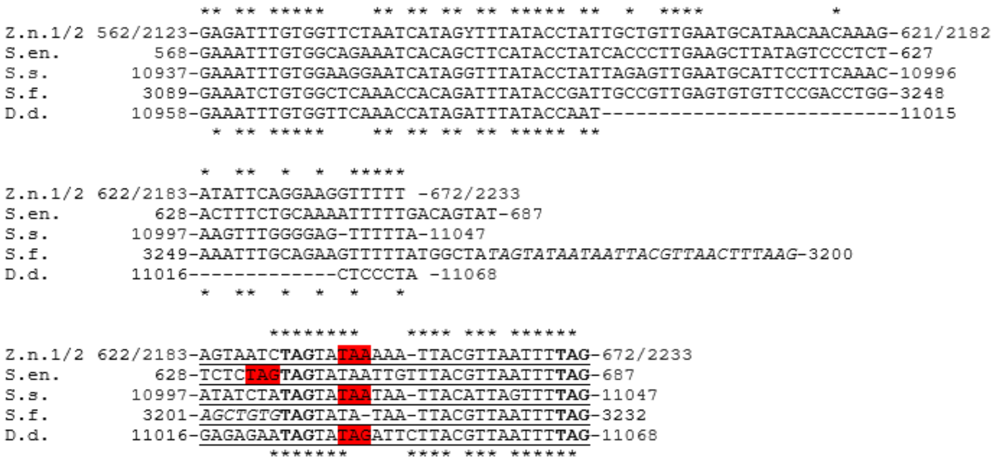
The *trnL1* gene of *P. gotoi* appears to be a complete ss-*trn* gene with both a functional TAG10 and ATG49 codons; indeed, alignment of the NH_2 termini of Nad3 proteins strongly suggests that the ATG49 triplet of the *P. gotoi trnL1* gene is the physiological start codon as the position of the initial methionine is very conservative (Fig 5). Moreover, in *S. ferox* and *P. draco*, the amino acid residues isoleucine (in the fifth position) and valine (in the ninth position) respectively, could be alternative start codons. This alignment also demonstrates the very low level of homology between the NH_2 termini of chaetognath Nad3 proteins. Moreover, the authors who reported the *S. enflata* and *S. setosa* sequences considered the start codon to be downstream of the other genes. However, it can be assumed that an upstream alternate ATT start codon could be used. In addition, if the *S. enflata* and *S. setosa* sequences are removed, the alignment reveals only two new conserved residues including the initial methionine. Additionally, in the *S. setosa* mt genomes [13] there are possible annotation errors; indeed, the COOH terminus of the NAD1 proteins is actually part of the sequence of the NAD3 proteins; *e.g.*, in the sequence with accession number KP899755, the first and the last nts of the NAD1 and NAD3 genes should be 7688/8320 and 8321/8629, respectively.



```
A
D.decipiens (5'-part) 11035-GAGAGAATAGTATAGATTCTTACGTTAATTTTAGGTAT------------11072
            (3'-part) 11073--------TAGTATA-ACTATTACATTAGTATTAAGAACTATTTCTCTTA-11113
                                   ******* * * **** *** * *** * *

P. draco (5'-part) 3331-AGGGCCCCAGTAACTGATAC-ATTAATTTTAGGTATTC-3367
        (3'-part) 3368---AAAGACAGGCACTGGCGCCAGCGTCTTTAAGCGACG-4003
                              ***  ****   * *     ****  *

S.elegans.1 (5'-part) 352-TAACAATAAAGCACCATATGCATTAGTTTTAGGAATTGGTC-392
            (3'-part) 393-TAACAGTATTA-ACCTTTTACATTATTTTTAAGTAATATGG-432
                          ***** **    *** * * ***** ***** * * *

S.enflata (5'-part) 654-TCTCTAGTAGTATAATTGTTTACGTTAATTTTAGGAATTA-693
          (3'-part) 694---------AGTATAAAATTTTACATTAATTTTAAGAATT--724
                                *******   ***** ********* *****

S.ferox (5'-part) 3164-TTTATGGCTATAGTATAATAATTACGTTAACTTTAAGAGCTGTG----3207
        (3'-part) 3208-----------TAGTATA-TAATTACGTTAATTTTAGGAATTAAAATG-3243
                                 ******* ************ **** **   *

S.setosa (5'-part) 11014-AATATCTATAGTATAATAATTACATTAGTTTTAGGAATTAATT-----11056
         (3'-part) 11057------GTATAGTATAATAATTACATTAAATTTAAGATTTAAAAAAAATG-11100
                               *********************  **** ** ****

Z.nagae.1/2 (5'-part) 642/2203-TAATCTAGTATAAAAATTACGTTAATTTTAGGAATTC------678/2239
            (3'-part) 679/2240------TAGTATA-TRTTTACGTTAATTTTAAGAATTAAATGG-714/2275
                                    ***** *    * ************ *****
B
S.el.1   348-AGAATAACAATAAAGCACCATATGCATTAGTTTTAGGAATTGGT----CTAACAGTATTA-403
S.el.2 11681-ACGAAACACAGCTTCTATCAACTATGGTAGCGTTAGGTGTTATTATTTCTTACAGTATTA-11740
             *  * *   *      * **  *    ***  *****  **  *    ** *********

S.el.1   404-ACCTTTTACATTATTTTTAAGTAATATGGTTCCGGTCGTCACTCTTGTCCTAATTATAGT-463
S.el.2 11741-G----TTACGTTATTTTTAAGAAATATGGTCTCCCTCATTACTCTTGTTATAGTTATAGT-11796
                  **** *********** ********  *  ** * ********  ** *******
```

Fig 6: Duplications found in *trnL1*-like sequences found within the *cox2-nd1* gene block of order Aphragmophora. The TAG10 triplets and the anticodons are presented in boldface, and the latter are underlined. The first in-frame stop codons are highlighted in red. The start codons of the *nd1* genes are highlighted in green. For other data, see Fig 3. (**A**) Alignments of duplicated regions. The nucleotides that do not belong to the *trn* sequence are underlined. (**B**) Alignment of the ss-*trnL1* gene of *S. elegans.1* and its flanking sequences with the homologous region found in the *cox1-nad1* gene block of *S. elegans.2*; the ss-*trnL1* sequence is underlined, and the 3' region of the duplicated region is in boldface.

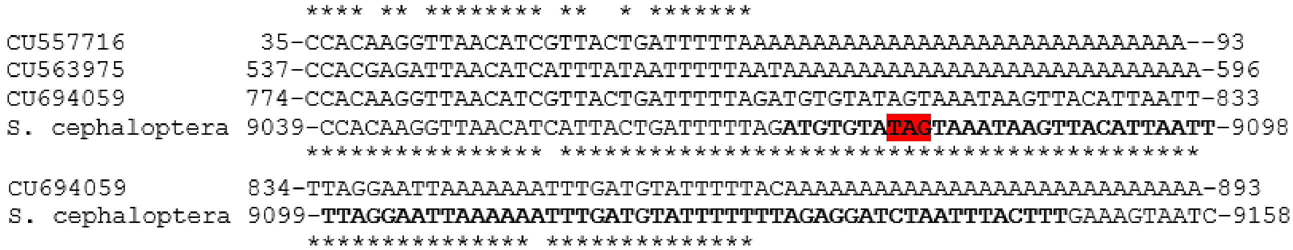
The ss-*trnL1*-like sequences of mt genomes belonging to order Aphragmophora resulted from a duplication event of a region stretching from approximatively TAG10 to the first part of the 3'-Ac-stem (Fig 6A). This duplication involves nearly the entire "cherry-bob" structure. However, in *S. ferox*, the possibility cannot be excluded that the duplication principally involved a region upstream of the *trn* gene. Concerning the two non ss-*trn* genes, only those belonging to *S. elegans.1* exhibit a high level of nucleotide conservation. It must be noted that the "G" nucleotide of the anticodons is never found in the duplicated regions. Analysis of Fig 6A suggests that the duplication occurred in the mt genome of the ancestor of suborder Ctenodontina. In the *S. elegans.2* mtDNA, the *cox2-nd1* gene block is not present, and no *trnL1*-like sequence has been found, but one mt-region exhibits sequence homology to the 3'-part of the *trnL1* gene of *S. elegans.1* (Fig 6B). This homology is found within the *cox1-nad1* gene block of *S. elegans.2*. The alignment suggests that during gene order rearrangement in lineage G, the 5' region of the *trnL1* gene was lost.



```
                 ** ** *****    ** ** ** ** ***** **  *  ****           *
Z.n.1/2 562/2123-GAGATTTGTGGTTCTAATCATAGYTTTATACCTATTGCTGTTGAATGCATAACAACAAAG-621/2182
S.en.        568-GAAATTTGTGGCAGAAATCACAGCTTCATACCTATCACCCTTGAAGCTTATAGTCCCTCT-627
S.s.       10937-GAAATTTGTGGAAGGAATCATAGGTTTATACCTATTAGAGTTGAATGCATTCCTTCAAAC-10996
S.f.        3089-GAAATCTGTGGCTCAAACCACAGATTTATACCGATTGCCGTTGAGTGTGTTCCGACCTGG-3248
D.d.       10958-GAAATTTGTGGTTCAAACCATAGATTTATACCAAT--------------------------11015
                  * ** *****    ** ** ** ** ***** **

                 *  **  *  *  *****
Z.n.1/2 622/2183-ATATTCAGGAAGGTTTTT -672/2233
S.en.        628-ACTTTCTGCAAAATTTTTGACAGTAT-687
S.s.       10997-AAGTTTGGGGAG-TTTTTA-11047
S.f.        3249-AAATTTGCAGAAGTTTTTATGGCTATAGTATAATAATTACGTTAACTTTAAG-3200
D.d.       11016-------------CTCCCTA -11068
                 *  **  *  *  *   *

                       ********    **** *** ******
Z.n.1/2 622/2183-AGTAATCTAGTATAAAAA-TTACGTTAATTTTAG-672/2233
S.en.        628-TCTCTAGTAGTATAATTGTTTACGTTAATTTTAG-687
S.s.       10997-ATATCTATAGTATAATAA-TTACATTAGTTTTAG-11047
S.f.        3201-AGCTGTGTAGTATA-TAA-TTACGTTAATTTTAG-3232
D.d.       11016-GAGAGAATAGTATAGATTCTTACGTTAATTTTAG-11068
                       *******     **** *** ******
```

Fig 7: Alignments of the 5' regions of the ss-*trn-L1* genes of order Aphragmophora with their 5' flanking sequences. The 5' sequence of the *trn*-genes is underlined. In *S. ferox*, the 5' region of the duplicated region are in italics. For the *Z. nagae.1/2* sequence, the letter Y represents a pyrimidine nucleotide. Asterisks indicate conserved nucleotides for all the sequences and for all the sequences except for those of *D. decipiens* on the upper and the lower lines, respectively. For other data, see Fig 3.

The 5' flanking regions of the ss-*trn-L1* genes of order Aphragmophora were analysed (Fig 7). For all the sequences, the approximately forty-base pair regions upstream of the TAG10 triplet exhibit a low homology rate with each other, especially compared with the 5' termini of the *trn* genes (Fig 7). Indels have also been found. The regions upstream of the TAG10 triplet encode the COOH termini of Cox1 proteins (see Fig 3); however, as noted by [12], in chaetognath mt genomes, the intergenic regions between protein-encoding genes may exhibit a higher level of homology than those found in protein-encoding regions. This strongly suggests that, at least in this case, the selection pressure exerted on the *trn* genes is much higher than that found in regions encoding proteins.

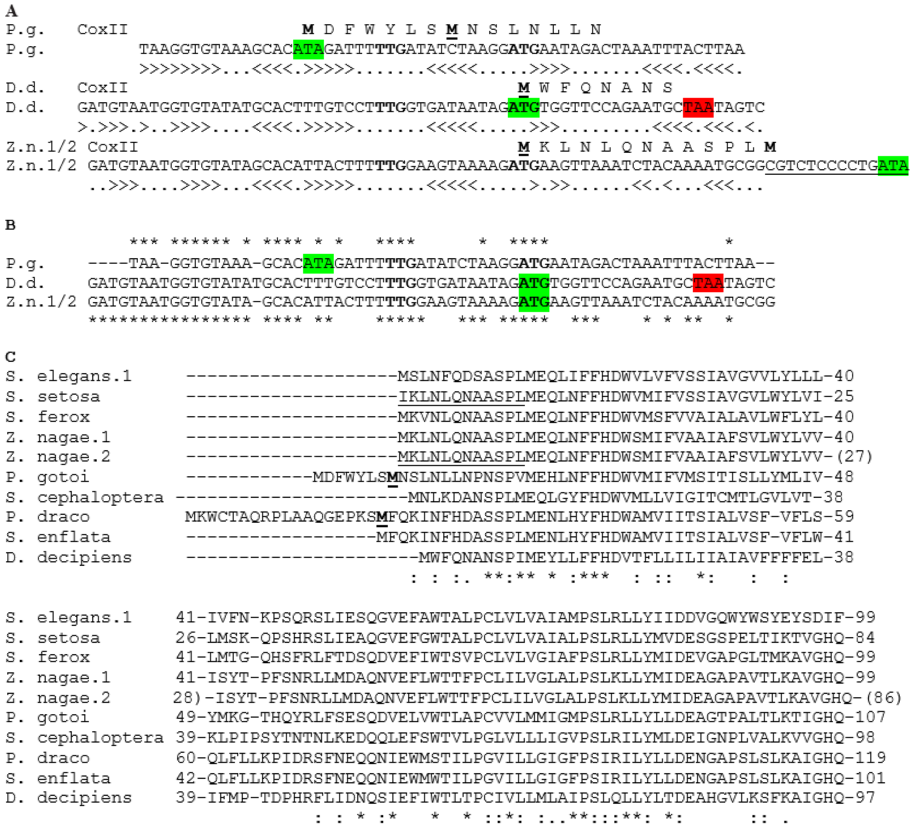


```
                      **** ** ******** **  * *******
CU557716           35-CCACAAGGTTAACATCGTTACTGATTTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA--93
CU563975          537-CCACGAGATTAACATCATTTATAATTTTTAATAAAAAAAAAAAAAAAAAAAAAAAAAAAA-596
CU694059          774-CCACAAGGTTAACATCGTTACTGATTTTTAGATGTGTATAGTAAATAAGTTACATTAATT-833
S. cephaloptera  9039-CCACAAGGTTAACATCATTACTGATTTTTAGATGTGTATAGTAAATAAGTTACATTAATT-9098
                      **************** ********************************************

CU694059          834-TTAGGAATTAAAAAAATTTGATGTATTTTTACAAAAAAAAAAAAAAAAAAAAAAAAAAAA-893
S. cephaloptera  9099-TTAGGAATTAAAAAATTTGATGTATTTTTTTAGAGGATCTAATTTACTTTGAAAGTAATC-9158
                      **************** ***************
```

Fig 8: Alignments of ESTs from *S. cephaloptera* with the region containing the *trnL1* gene. The sequence of the gene is in boldface and the TAG10 triplet, which was considered to be the stop codon of the *nad1* gene by the authors who reported the sequence, is highlighted in red. Asterisks indicate nucleotides conserved with the ESTs CU557716 and CU563975 on the upper line and for the three ESTs on the lower line, respectively.

For one chaetognath species (*S. cephaloptera*) a large collection of expressed sequence tags (ESTs) has been assembled by Marlétaz *et al.* [26] and submitted to the EMBL website. Blast analyses using all the *trn* genes analysed in this study as the query against the *S. cephaloptera* EST database only revealed homology with the *trnL1* gene or its flanking regions (Fig 8). Two ESTs contain the 3’ portion of the *nad1* mRNAs and polyadenylation begins one or two nucleotide(s) upstream of the *trnL1* sequence (CU557716) or after the first nt of this *trn* gene (CU563975). In the latter case, the "T" nt at position 568 could be an artefact, as suggested by analyses of the beginning portions of other polyadenylated sequences. Even though the analysis only involves two EST sequences, this suggests that the *nad1* mRNAs could exhibit an incomplete stop codon immediately before the 5’ terminus of the *trnL1* gene. The third EST (CU694059) is bicistronic and contains a large part of the *nad1* gene and the *trnL1* sequence up to the beginning of the 5’-T-stem. As the 3’-T-stem and the 3’-Acc-stem are missing, the putative truncated tRNA would require post-transcriptional maturation to become functional.

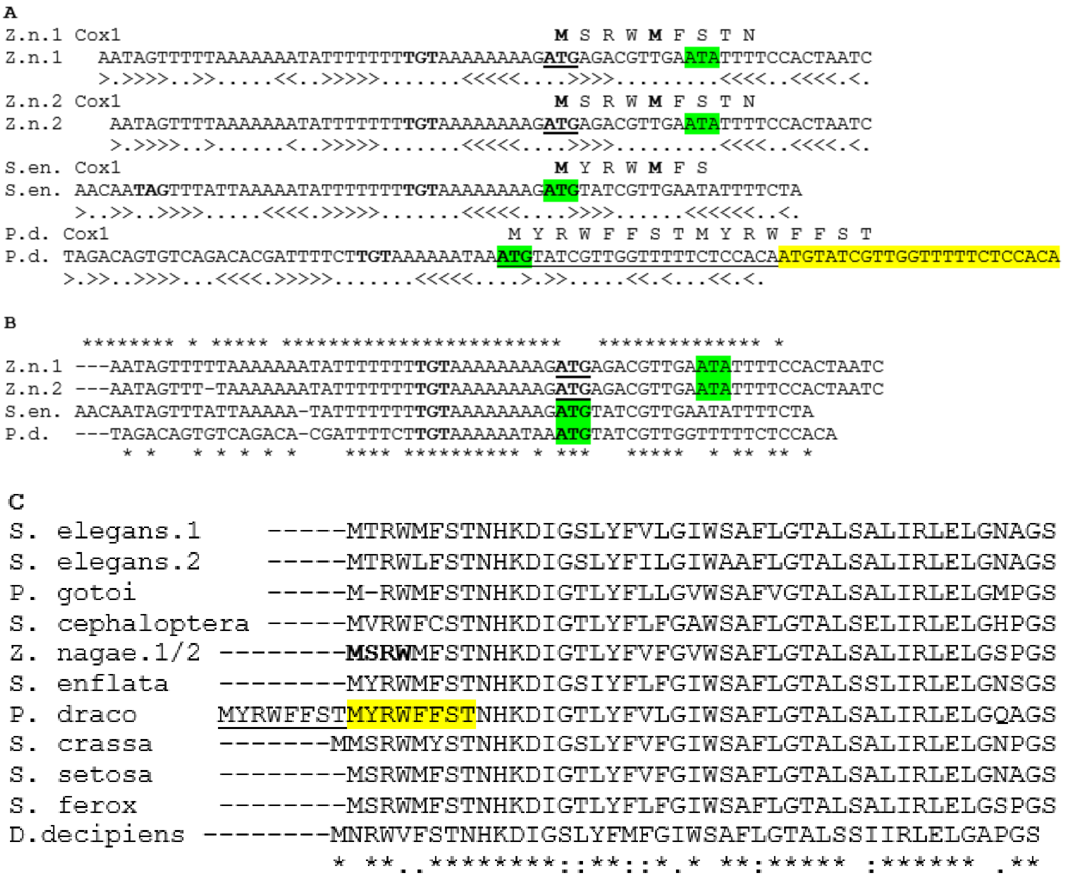
3.3. *trnG* genes



```
A
P.g.   CoxII                 M D F W Y L S M N S L N L L N
P.g.         TAAGGTGTAAAGCACATAGATTTTTGATATCTAAGGATGAATAGACTAAATTTACTTAA
             >>>>>>>>...<<<<.>>>>>.......<<<<<.....>>>>.......<<<<.<<<<.
D.d.   CoxII                                      M W F Q N A N S
D.d.   GATGTAATGGTGTATATGCACTTTGTCCTTTGGTGATAATAGATGTGGTTCCAGAATGCTAATAGTC
       >.>>>.>..>>>>....<<<<.>>>>>.......<<<<<.....>>>.........<<<<.<<<.<.
Z.n.1/2 CoxII                                     M K L N L Q N A A S P L M
Z.n.1/2 GATGTAATGGTGTATAGCACATTACTTTTTGGAAGTAAAAGATGAAGTTAAATCTACAAAATGCGGCGTCTCCCCTGATA
        ..>>>....>>>>...<<<<.>>>>>.......<<<<<.....>.>>......<<.<...<<<...

B
            *** ****** * **** * *   ****      *  ****                 *
P.g.    ----TAA-GGTGTAAA-GCACATAGATTTTTGATATCTAAGGATGAATAGACTAAATTTACTTAA--
D.d.    GATGTAATGGTGTATATGCACTTTGTCCTTTGGTGATAATAGATGTGGTTCCAGAATGCTAATAGTC
Z.n.1/2 GATGTAATGGTGTATA-GCACATTACTTTTTGGAAGTAAAAGATGAAGTTAAATCTACAAAATGCGG
        **************** **** **    *****   *** *****  ***    * * **  *

C
S. elegans.1     --------------------MSLNFQDSASPLMEQLIFFHDWVLVFVSSIAVGVVLYLLL-40
S. setosa        --------------------IKLNLQNAASPLMEQLNFFHDWVMIFVSSIAVGVLWYLVI-25
S. ferox         --------------------MKVNLQNAASPLMEQLNFFHDWVMSFVVAIALAVLWFLYL-40
Z. nagae.1       --------------------MKLNLQNAASPLMEQLNFFHDWSMIFVAAIAFSVLWYLVV-40
Z. nagae.2       --------------------MKLNLQNAASPLMEQLNFFHDWSMIFVAAIAFSVLWYLVV-(27)
P. gotoi         ------------MDFWYLSMNSLNLLNPNSPVMEHLNFFHDWVMIFVMSITISLLYMLIV-48
S. cephaloptera ----------------------MNLKDANSPLMEQLGYFHDWVMLLVIGITCMTLGVLVT-38
P. draco         MKWCTAQRPLAAQGEPKSMFQKINFHDASSPLMENLHYFHDWAMVIITSIALVSF-VFLS-59
S. enflata       ------------------MFQKINFHDASSPLMENLHYFHDWAMVIITSIALVSF-VFLW-41
D. decipiens     ----------------------MWFQNANSPIMEYLLFFHDVTFLLILIIAIAVFFFFEL-38
                                      : : :. **:** * :***  : ::  *:   :  :

S. elegans.1    41-IVFN-KPSQRSLIESQGVEFAWTALPCLVLVAIAMPSLRLLYIIDDVGQWYWSYEYSDIF-99
S. setosa       26-LMSK-QPSHRSLIEAQGVEFGWTALPCLVLVAIALPSLRLLYMVDESGSPELTIKTVGHQ-84
S. ferox        41-LMTG-QHSFRLFTDSQDVEFIWTSVPCLVLVGIAFPSLRLLYMIDEVGAPGLTMKAVGHQ-99
Z. nagae.1      41-ISYT-PFSNRLLMDAQNVEFLWTTFPCLILVGLALPSLKLLYMIDEAGAPAVTLKAVGHQ-99
Z. nagae.2     28)-ISYT-PFSNRLLMDAQNVEFLWTTFPCLILVGLALPSLKLLYMIDEAGAPAVTLKAVGHQ-(86)
P. gotoi        49-YMKG-THQYRLFSESQDVELVWTLAPCVVLMMIGMPSLRLLYLLDEAGTPALTLKTIGHQ-107
S. cephaloptera 39-KLPIPSYTNTNLKEDQQLEFSWTVLPGLVLLLIGVPSLRILYMLDEIGNPLVALKVVGHQ-98
P. draco        60-QLFLLKPIDRSFNEQQNIEWMSTILPGVILLGIGFPSIRILYLLDENGAPSLSLKAIGHQ-119
S. enflata      42-QLFLLKPIDRSFNEQQNIEWMWTILPGVILLGIGFPSIRILYLLDENGAPSLSLKAIGHQ-101
D. decipiens    39-IFMP-TDPHRFLIDNQSIEFIWTLTPCIVLLMLAIPSLQLLYLTDEAHGVLKSFKAIGHQ-97
                             : : * :*   *  * ::*: :..**:::**: *:      :: .
```

Fig 9: Analyses of chaetognath *trnG* genes. (**A**), In the mtDNAs of *D. decipiens* (10402-10468), *P. gotoi* (1535-1593), and *Z. nagae.1* and *Z. nagae.2* (1521-1600 and 11419-11459/1-35 respectively); these sequences are at the level of the regions flanking the *cox1* and *cox2* genes. ATG49 and the anticodon (TTG) are indicated in bold. The putative start codons indicated by the authors who reported the sequences are highlighted in green. For other data, see Fig 3. In the *P. gotoi* sequence, ATG49 is in frame with the putative start codon ATA. Moreover, the first complete stop codon of the *cox1* gene of *D. decipiens*, a TAA triplet, is highlighted in red. Alignments are presented using secondary structure predictions, and the Cox2 amino acid sequences are indicated. (**B**), Alignment of primary sequences. (**C**), Alignment of the NH_2 termini of Cox2 proteins. For *Z. nagae.2* and *S. setosa*, the residues in boldface have been added to the sequences proposed by the authors. For the *P. gotoi* protein, the methionine encoded by the ATR49 triplet is in bold and underlined. Similarly, the putative initiation methionine of the *P. draco* protein is in bold and underlined.

Three glycinyl ss-*trn* genes have been found in chaetognath mt genomes (Fig 9), but a high level of nucleotide conservation has been found only between the two sequences of order Aphragmophora (Fig 9B). Analyses of the ss-*trn* genes of *D. decipiens* and *P. gotoi* suggest that the ATG49 triplets present in these sequences might be the physiological start codons of the *cox2* genes (Fig 9). For *D. decipiens*, the authors who reported the sequences assumed that the start codon was ATG49, whereas for *P. gotoi*, they hypothesized that a triplet ATA, located upstream of the anticodon, was the translation initiation codon [15] (Fig 9A); for *Z. nagae trnG* genes, according to the authors who reported the sequences, the start codon is downstream from the *trn* gene, but ATG49 is in frame with the *cox2* gene. Blast analyses strongly suggest that the three ATG49 triplets are the physiological start codon for the *cox2* gene (Fig 9C). Indeed, the alignment of the regions around the first amino acid residues of chaetognath Cox2 proteins shows that, except for the *P. gotoi* and *P. draco* proteins, all the other sequences subjected to Blast searches are shorter by at least six residues (Fig 9C). Additionally, whereas for the *cox1* gene of *P. gotoi*, the stop codon (TAA) is located upstream of the *trnG* gene, for *D. decipiens,* the first complete stop codon of the *cox1* gene is the TAA triplet located at the 5' terminus of the 3'-Ac-stem, suggesting overlap between the two protein-encoding sequences. However, in this last hypothesis, Blast analyses revealed that the COOH-terminal portion of the *D. decipiens* protein, which is abnormally long, does not exhibit any sequence that homologous (or analogous) to other Cox1 protein sequences (data not shown). An incomplete stop codon located upstream of the *trnG* gene is presumably completed by polyadenylation. In addition, relatively strong sequence alignment is found between the two *trnG* genes of order Aphragmophora (Fig 9B); whereas, the level of nucleotide sequence identity between the *P. gotoi trnG* gene and the two sequences of species of order Aphragmophora is approximately 42%. Moreover, the level of sequence identity is lower in the overlapping region (*trnG/cox2*) than in the non-overlapping area, 17% versus 58%, highlighting against a high degree of sequence conservation between the first residues of the various Cox2 proteins (see Fig 9C). Besides, there could be annotation errors at the beginning of the *S. elegans* Cox2 proteins, *e.g.*, for *S. elegans.1*, the start codon begins at nt 13364 and not at nt 1, and the CAG codon is probably not an alternative initiation codon. The Cox2 sequence of *S. elegans.2* has not been included in the analysis because its NH_2 terminus differs strongly from those of the other sequences.



```
A
Z.n.1 Cox1                                       M S R W M F S T N
Z.n.1   AATAGTTTTTAAAAAAATATTTTTTTTGTAAAAAAAAGATGAGACGTTGAATATTTTCCACTAATC
        >.>>>>..>>.....<<..>>>>>.......<<<<<....>>>>.........<<<<..<<<<.<.
Z.n.2 Cox1                                       M S R W M F S T N
Z.n.2    AATAGTTTTAAAAAAATATTTTTTTTGTAAAAAAAAGATGAGACGTTGAATATTTTCCACTAATC
         >.>>>>..>......<..>>>>>.......<<<<<....>>>>.........<<<<..<<<<.<.
S.en. Cox1                                       M Y R W M F S
S.en. AACAATAGTTTATTAAAAATATTTTTTTTGTAAAAAAAAGATGTATCGTTGAATATTTTCTA
      >..>>..>>>>.....<<<<.>>>>>.......<<<<<....>>>>......<<<<<<..<.
P.d. Cox1                                    M Y R W F F S T M Y R W F F S T
P.d. TAGACAGTGTCAGACACGATTTTCTTGTAAAAAATAAATGTATCGTTGGTTTTTCTCCACAATGTATCGTTGGTTTTTCTCCACA
     >.>>..>>>>...<<<<.>>>>>.......<<<<<....>.>>.....<<.<...<<.<.

B
       ******** * ***** ************************   ************** *
Z.n.1 ---AATAGTTTTTAAAAAAATATTTTTTTTGTAAAAAAAAGATGAGACGTTGAATATTTTCCACTAATC
Z.n.2 ---AATAGTTT-TAAAAAAATATTTTTTTTGTAAAAAAAAGATGAGACGTTGAATATTTTCCACTAATC
S.en. AACAATAGTTTATTAAAAA-TATTTTTTTTGTAAAAAAAAGATGTATCGTTGAATATTTTCTA
P.d.  ---TAGACAGTGTCAGACA-CGATTTTCTTGTAAAAAATAAATGTATCGTTGGTTTTTCTCCACA
          * *   * * * * *    **** ********** * ***   *****  * ** ** *

C
S. elegans.1    -----MTRWMFSTNHKDIGSLYFVLGIWSAFLGTALSALIRLELGNAGS
S. elegans.2    -----MTRWLFSTNHKDIGSLYFILGIWAAFLGTALSALIRLELGNAGS
P. gotoi        -----M-RWMFSTNHKDIGTLYFLLGVWSAFVGTALSALIRLELGMPGS
S. cephaloptera -----MVRWFCSTNHKDIGTLYFLFGAWSAFLGTALSELIRLELGHPGS
Z. nagae.1/2 --------MSRWMFSTNHKDIGTLYFVFGVWSAFLGTALSALIRLELGSPGS
S. enflata   --------MYRWMFSTNHKDIGSIYFLFGIWSAFLGTALSSLIRLELGNSGS
P. draco     MYRWFFSTMYRWFFSTNHKDIGTLYFVLGIWSAFLGTALSALIRLELGQAGS
S. crassa    -------MMSRWMYSTNHKDIGSLYFVFGIWSAFLGTALSALIRLELGNPGS
S. setosa    --------MSRWMFSTNHKDIGTLYFVFGIWSAFLGTALSALIRLELGNAGS
S. ferox     --------MSRWMFSTNHKDIGTLYFLFGIWSAFLGTALSALIRLELGSPGS
D.decipiens --------MNRWVFSTNHKDIGSLYFMFGIWSAFLGTALSSIIRLELGAPGS
                    * **..********::**::*.* **:***** :****** .**
```

Fig 10: Alignments of *trnT* and Cox1 sequences of chaetognaths. In the *Z. nagae1* (9849-9914), *Z. nagae2* (11410-11458/1-16) *S. enflata* (11020-11081) and *P. draco* (937-1021) mtDNAs, the *trnT*-like genes are at the level of the regions flanking the *16S rrn* genes and *cox1* genes. The ATG49 triplets, the TAG10 triplet of *S. enflata* and the anticodons are in bold. The start codons given by the authors who reported the sequences are highlighted in green. In the *Z. nagae* sequences, the ATG49 triplet, which is the first possible start codon, is underlined. (**A**) Alignment of *trnT*-like genes with their 2D structures. For the *P. draco* sequence, the two parts of the duplicated regions are underlined or highlighted in yellow. The NH_2 regions of the Cox1 proteins are also reported. (**B**) Alignment of the primary sequences of the *trnT*-like genes. Asterisks indicate conserved nucleotides for *Z. nagae* and *S. enflata* sequences on the upper line and for the four sequences on the lower line, respectively. (**C**) Alignment of the NH_2 regions of the Cox1 proteins of Aphragmophora specimens. All the sequences correspond to those found in GenBank, except for the four first residues (in bold) of the *Z. nagae* sequences.

In chaetognath mt genomes, four threoninyl ss-*trn* genes have been found (Fig 10A). All of these sequences have been found among specimens of order Aphragmophora and contain the ATG49 triplet. According to the authors who reported the sequences, this triplet serves as the start codon for the *S. enflata* and *P. draco cox1* genes, whereas for the *Z. nagae* sequences, the initiation codon is an in-frame ATA codon located downstream from the ATG49 triplet [17]. According to Wei *et al.* [20], the *trnT*-like sequence of *P. draco* starts at nt-934 and ends at nt-1021, but our reanalysis shows that the 3' terminus of the *trn*-gene of *P. draco* has been duplicated and that this *trn* gene is shorter and does not include the duplicated region (Fig 10A). Alignment of the primary sequences of the *trnT*-like genes revealed that the sequences of *Z. nagae* and *S. enflata* which are closely related species exhibit a high level of homology (> 89%) (Fig 10B). Alignment of the four sequences shows that the 5'-Acc-stem and the D-loop of the *P. draco* sequence are very different from those of the two other species. Additionally, alignment of the NH_2 regions of the Cox1 proteins of chaetognath specimens strongly suggests that the physiological start codon would be the ATG49 triplet; however, four amino acid residues downstream of the methionine encoded by ATG49 triplets, there is often a methionine residue (for 7 of 12 sequences) suggesting that there may be an alternative initiation codon (Fig 10C).

4. Discussion

4.1 Chaetognath *trnM* genes

In some species, only one or two *trn* genes have been found among their mtDNAs. When only one gene is present, it specifies tRNA-Met (to date, this situation is only observed in some Anthozoa, a class of Cnidaria), whereas when only two are found, one specifies tRNA-Met and the other tRNA-Trp (this occurs in cnidarians and a lineage of demosponges: Keratosa) [27]. This observation supports the inference of special roles of these two tRNAs in animal mitochondria, in translation initiation and the recognition of the standard stop codon UGA as tryptophan. Indeed, in most animal mt genetic codes, the two major deviations from the standard genetic code are the reassignment of the AUA codon from isoleucine to both initial and internal methionine residues (*trnM*(CAU)) and reassignment of the UGA codon from the termination codon to tryptophan (*trnW*(UCA)) (see *e.g.*, [16, 27]). This situation suggests that these reassignments would have appeared very early in the evolution of, at least, metazoans and would have resulted from the optimization of the mt genome [22]. A previous analysis showed that a *trnW* gene exists in the mtDNAs of almost all examined chaetognaths [12], and the present study showed that only the *P. draco* mt genome does not appear to exhibit the *trnM* gene. Interestingly, the chaetognath *trnM* genes seem to correspond to the initiating tRNA (tRNA-fMet) which may be charged with formyl-methionine to be used for protein initiation. As found in numerous taxa, including chaetognaths, the AUA and AUG codons are used in both initiator and internal positions; if only one tRNA-Met is specified mitochondrially, it can either play the two translational roles, or it might be charged only with formyl-methionine, whereas another tRNA-Met from the nucleus only functions in elongation. All the chaetognath *trnM* genes identified to date are partial or complete ss-*trn* genes, but the TAA10 and ATG49 triplets are never used as stop or start codons respectively. Concerning the ATG49 triplets found in the *D. decipiens*, *P gotoi* and *S. cephaloptera* mt genomes, as the downstream genes specifying 12S rRNA, they cannot play a role in translation. In addition, in chaetognath *trnM* genes, TAA10 triplets are never used as stop codons; when they are in-frame, there is always an upstream complete stop codon. Analyses of chaetognath *trnM* genes suggest that the use of TAR10 and ATR10 as stop or start codons, respectively, in other ss-*trn* genes could partially represent an exaptation. This concept suggests that traits that evolved for one purpose have been co-opted for their current use [28]. However, in the ss-*trn* genes these two types of triplets continue to perform their primitive function, which is essential for maintaining the spatial conformation of tRNAs [22]; nevertheless, in some cases, they can also play a crucial role in translation. The presence of ss-*trn* genes in bacterial and nuclear genomes that are relatively distant from neighbouring protein-encoding genes also argues in favour of this hypothesis.

Furthermore, the *S. elegans* mt genomes have been divided into eight lineages [13]. In all the lineages except G, the *trnM* genes are within the *cox3-rrnS* gene block, whereas in the last lineage, the *trn* gene is within the *nd3-rrnS* gene block (see S.el2 in Fig 2 and Table 2), and the stop codon of the *nad3* gene is located considerably upstream of the *trnM* sequence. Despite the difference in the upstream protein-encoding genes, the sequences of the *trn* genes are very similar; *e.g.*, the two analysed *trnM* genes of *S. elegans* exhibit a high level of nucleotide identity (92%), and the percentage is always 100% at the level of the stems (Fig 2). This finding suggests that, these ss-*trn* genes at least appeared independent of any interactions with protein-encoding genes; indeed, when these interactions existed, they would have been selected later. The high degree of conservation of the *trnM* genes of different lineages of *S. elegans* suggests that strong selection pressure is exerted on this region suggesting that the specified tRNA would be functional.

Despite the fact that in animal mtDNAs, the level of similarity between *trn* genes, even those of the same type, is relatively low, a relatively large number of *trnM* genes exhibit sequence similarity at the level of their cherry-bob motif. As these similarities have been found principally in protostomian mt genomes (mostly chaetognaths and insects), this situation could represent evolutionary convergence, or the motif may already have been present in the common ancestor of chaetognaths and insects, considering the assumed age of phylum Chaetognatha, this motif would extend back to the mitochondria of the first protostomians. The latter hypothesis suggests that this motif would have disappeared from the mt genomes of the majority of protostomial taxa. According to the current knowledge, it is difficult to reject one of these two assumptions rather than the other. Additionally, for most of the chaetognath *trnM* genes, 2D structure predictions show non-canonical spatial arrangements after the Ac-5'-stem suggesting possible post-transcriptional modifications.

Table 2: MtDNAs of species bearing a specific ss-*trn* gene within a given gene block. In this study, the notion of a gene block only refers to protein-encoding genes and those specifying rRNAs. The names of the species belonging to the order Phragmophora are underlined. For lineage G of *S. elegans* cf. *S. elegans.2*.

Gene block	Species bearing a mt-DNA having this gene block	Type of ss-*trn* gene found in the gene block	Species bearing a mt-DNA with this ss-*trn* gene ou a *trn*-like gene of the same type
cox3-rrnS	All except *P. draco* and the lineage G of *S. elegans*	*trnM*	All except *P. draco* and the lineage G of *S elegans*
nd3-rrnS	The lineage G of *S. elegans*	*trnM*	The lineage G of *S. elegans*
cox2-nd1	All except the lineage G of *S. elegans*	*trnL1*	All except *P. gotoi, S. cephaloptera* and the lineage G of *S elegans*
nd1-nd3	All except the lineage G of *S. elegans*	*trnL1*	*P. gotoi, S. cephaloptera*
cox1-cox2	All except the lineage G of *S. elegans*	*trnG*	*D. decipiens, Z. nagae.1/2, P. gotoi*
rrnL-cox1	All except *S. cephaloptera*	*trnT*	*P. draco, S. enflata, Z. nagae.1/2*

4.2 Chaetognath ss-*trn* genes not specifying tRNA-Met

Within chaetognath mtDNAs, seven putative leucinyl(1)-, four glycinyl- and four threoninyl-ss-*trn* genes have been found. The *trnT* genes have only been detected in some species of order Aphragmophora. Analyses of the COOH and NH_2 termini of proteins encoded by overlapping genes strongly suggest that most of these ss-tRNAs exhibit TAG10 or ATG49 triplets, which are probably used as physiological stop or start codons respectively. Moreover, ss-tRNA-Leu1 of *P. gotoi* could be a complete true ss-tRNA in which the two types of triplet play a role at the translation level. The predicted 2D structures suggest that these tRNAs do not adopt canonical structures but improvements could occur during post-transcriptional maturation, which can also include the removal of extra loops. Only *trnM* and *trnL1* genes exhibit a relatively high level of nucleotide identity with *trn* genes of the same type belonging to non-chaetognath species but for the *trnL1* genes, this could be due in part to a high AT content [12]. In contrast to the *trnL1* genes, the *trnG* genes, which have also been found in species belonging to the two chaetognath orders, show low inter order-level nucleotide identity. It seems that the *trn* genes present among the current chaetognath mtDNAs arose via a type of DIY mechanism ("molecular tinkering"), partially independent of phylogenetic relationships.

Generally, in nucleotidic regions where two or more genes overlap, selection pressure is such that the sequences are highly conserved, and the reading frame is said to be locked. Paradoxically, in the ss-*trnL1*, ss-*trnG* and ss-*trnT* genes, the level of nucleotide identity is low in overlapping regions. This finding suggests that ss-*trn* genes, at least in chaetognaths, play a role in the precision of the initiation or termination of transcription (or in transcript maturation) and especially in the translation of mt-mRNAs but do not act at the level of sequence conservation of protein-encoding genes. Moreover, as suggested previously [22], the genomic organization of ss-*trn* genes could allow regulation of the synthesis of products of the mt genomes of chaetognaths.

In chaetognath mitogenomes, the *rrnL* gene could constitute a type of *trn* gene nursery [12], and the annotation of *S. ferox* mtDNA reinforces this hypothesis [18]; however, in chaetognath *rrnL* genes no sequence exhibiting strong homology with the four types of ss-*trn* genes analysed in this study has been found within the sequences specifying rRNAs.

Among the *S. cephaloptera* EST collection, only three ESTs exhibit sequence homology to one of the *trn* genes (*trnL1*) analysed in this article or to its upstream region. Two ESTs suggest that there is probably an incomplete stop codon just before the beginning of the *trn* gene. This situation could argue in favour of functionality of this *trn* gene. No EST that ends after the 3' terminus of the *trn* sequence has been found, and the polyadenylated region of an EST starts at the beginning of the putative 5'-T-stem. The three following hypotheses can be proposed: 1/ this site of polyadenylation is artefactual; 2/ the 3' portion of the *trnL1* sequence is shorter; or 3/ as has already been shown, incomplete cloverleaf structures may also be repaired post-transcriptionally [29] and, in some cases, mt-*trn* genes that have incomplete 3'-ends are completed post-transcriptionally by polyadenylation [30]. However, the polyadenylation of mt RNAs could be ambivalent; indeed, this process can stabilize them or may be a signal of the degradation of the RNAs [31]; thus, further experiments are required to reach a conclusion.

Several models have been proposed to explain the origin of tRNA molecules (see reviews [32-34]), which is beyond the scope of this work, but the two major hypotheses can be briefly proposed. Many authors have assumed that the current tRNA cloverleaf structure arose through direct duplication of an ancestral RNA hairpin (*e.g.,*) [32], whereas according to the "two halves" hypothesis [35], tRNAs are composed of two independent structural and functional domains: the "top half", containing the acceptor-stem and the T-arm, and the "bottom half", equivalent to the cherry-bob structure, which contains the D- and anticodon-arms (Fig 1). Several analyses have strongly suggested that the "bottom half", which would have appeared much later, was integrated into the pre-existing top half structure [36]. Analyses of *trnL1*-like sequences of species of order Aphragmophora revealed that they are a result of duplication of the cherry-bob region, which constitutes a sort of syncretism between the hypotheses of Di Giulio [32] and of Maizels and Weiner [35].

Additionally, in the living world, mechanisms exist for avoiding the possibility that in biological systems, a molecule that is similar but not strictly identical (at the sequence and therefore structure levels) to another that is functional could interfere with and make the latter molecule non functional in a given system. Under a hypothesis that we refer to as the "double funnel" hypothesis, the first funnel, in which the larger hole is at the top, symbolizes the fact that within multigene families, all members retain almost identical sequences over very long periods of time. This suggests the presence of active mechanisms causing sequences that begin to diverge to become almost all identical again. Gene conversion is one of the mechanisms in this process of "concerted evolution" [37]. In contrast, to avoid the fact that the equivalence principle does not apply between molecules that are relatively closely related at the sequence level, the opposite mechanism corresponding to the second funnel for which the pipe is at the top quickly accentuates differences. The natural mutation rate allows the induction of variations between sequences but, it is probably too slow when the situation is critical, *e.g.*, after allopolyploidization events (see, *e.g.*,) [9-11]. Several examples of the "double funnel hypothesis" can also be found in the tRNA world, which could imply the occurrence of duplication of *trn* genes followed by substitutions (*e.g.*,) [38, 39] and the existence of ss-*trn* genes that can combine several specificities related to the type of tRNA and to the overlap of protein-encoding genes in their 5'- and 3'-regions, which confers upon them singular characteristics.

What might have been the event(s) exerting such selection pressure that all the original mt-*trn* genes would have disappeared? The numerus rearrangements that have occurred in chaetognath mtDNAs would have been detrimental to *trn* genes because selection pressure would have been exerted preferentially at the level of protein-encoding genes. However, some observations strongly suggest that codon reassignments might be adaptive. Mt-codon reassignments might prevent viral infections as most viruses follow the standard genetic code [40, 41]. Approximately half of known alternative genetic codes are mt genetic codes and three of them have been found in fungi [42]. Viruses infecting mitochondria are relatively rare, but members of genus *Mitovirus*, family *Narnaviridae* (which are all capsidless), are ubiquitously detected in filamentous fungi; these viruses exhibit a positive single-stranded RNA genome encoding only an RNA-dependent RNA polymerase [43] and share their mitochondrial host's codon reassignment [44]. Moreover, endogenized mitovirus elements are widespread in land plant genomes [45]; thus, chaetognath nuclear genomes could still contain traces of past infections. Although very hypothetical at present, the ss-*trn* genes could also play a protective role in viral mt-infections by a completely unknown mechanism.

5. Conclusion

In addition to the loss of many *trn* genes and those encoding ATP6 and ATP8, the chaetognath mt genomes are characterized by an accelerated mt-mutation rate, which could be related to the extreme size reduction of these genomes (from c.11,000 to 14,000 bp for current sizes) and their propensity for structural rearrangements [17]. We assumed that in chaetognaths, all of the original mt-*trn* genes were lost during evolution, and that tRNAs allowing the translation of mt-mRNAs have been superseded by nuclear-specified tRNAs. Actually, since mt-tRNAs are perhaps more effective than their cytosolic equivalents, we are witnessing the neosynthesis of new specific mt-*trn* genes via what seems, at least partly, to consist of a large-scale DIY mechanism. The fact that the pool of *trn* genes differs between phylogenetically closely related species and that tRNA-like sequences seem to be *in statu nascendi* favours this hypothesis. The observation that some of the new *trn* genes are ss-*trn* sequences strongly suggests the use of UAR10 or AUR49 as stop or start codons, respectively, as an exaptation. DNA alignments corresponding to COOH- or NH_2-terminal protein regions, in which TAR10 and ATR49 are not always the physiological stop and start codons, reinforce this last hypothesis. The high level of nucleotide conservation in non-overlapping regions of ss-*trn* genes strongly suggests that selection pressure occurs preferentially in these areas, but this suggestion contrasts with known data, which deserves special attention in subsequent works. In the future, research on specific mt-amino-acyl tRNA synthetases could be very informative; direct sequencing of tRNAs and *in vitro* translation experiments are also needed.

6. References

1. Tokioka T. Supplementary notes on the Systematics of Chaetognatha. Publ. Seto Mar. Biol. Lab. 1965; 13(3):231-242.
2. Casanova JP. Chaetognatha. In *South Atlantic zooplankton*. Boltovskoy D, Ed.; Backhuys, Leiden, Germany, 1999, 1353-1374.
3. Casanova JP, Barthélémy R, Duvert M, Faure E. Chaetognaths feed primarily on dissolved and fine particulate organic matter, not on prey: implications for marine food webs. Hypotheses Life Sci. 2012; 2:20-29.
4. Vannier J, Steiner M, Renvoisé E, Hu SX, Casanova JP. Early Cambrian origin of modern food webs: evidence from predator arrow worms. Proc. Biol. Sci. 2007; 274(1610):627-633.
5. Barthélémy RM, Casanova JP. The glandular canals present in some chaetognath species are coelomoducts: Phylogenetical implication. Int. J. Fauna Biol. Studies. 2018; 5(6B):95-95.
6. Barthélémy RM, Casanova JP. Progress in the knowledge of the phylogeny of the Chaetognatha needs both molecular biology and zoology. Int. J. Fauna Biol. Studies. 2019; 6(2A):21-26.
7. Marlétaz F, Peijnenburg KTCA, Goto T, Satoh N, Rokhsar DS. A new spiralian phylogeny places the enigmatic arrow worms among gnathiferans. Curr. Biol. 2019; 29(2):312-318.
8. Barthélémy RM, Faure E, Goto T. Serendipitous discovery in a marine invertebrate (Phylum Chaetognatha) of the longest giant viruses reported till date. Virol. Curr. Res. 2019; 3:110.
9. Barthélémy RM, Casanova JP, Grino M, Faure E. Selective expression of two types of 28S rRNA paralogous genes in the chaetognath *Spadella cephaloptera*. Cell. Mol. Biol. (Noisy-le-grand). 2007; 53:S989-S993.
10. Barthélémy RM, Chenui A, Brancart S, Casanova JP, Faure E. Translational machinery of the chaetognath *Spadella cephaloptera*: A transcriptomic approach to the analysis of cytosolic ribosomal protein genes and their expression. BMC Evol. Biol. 2007; 7:146.
11. Barthélémy RM, Grino M, Pontarotti P, Casanova JP, Faure E. The differential expression of ribosomal 18S RNA paralog genes from the chaetognath *Spadella cephaloptera*. Cell. Mol. Biol. Lett. 2007; 12(4):573-583.
12. Barthélémy RM, Seligmann H. Cryptic tRNAs in chaetognath mitochondrial genomes. Comput. Biol. Chem. 2016; 62:119-132.
13. Marlétaz F, Le Parco Y, Liu S, Peijnenburg KTCA. Extreme mitogenomic variation in natural populations of chaetognaths. Genome Biol. Evol. 2017; 9(6):1374-1384.
14. Papillon D, Perez Y, Caubit X, Le Parco Y. Identification of chaetognaths as protostomes is supported by the analyses of their mitochondrial genome. Mol. Biol. Evol. 2004; 21(11):2122-2129.
15. Helfenbein KG, Fourcade HM, Vanjani RG, Boore JL. The mitochondrial genome of *Paraspadella gotoi* is highly reduced and reveals that chaetognaths are a sister group to protostomes. Proc. Natl. Acad. Sci. USA. 2004; 101(29):10639-10643.
16. Faure E, Casanova JP. Comparison of chaetognath mitochondrial genomes and phylogenetical implications. Mitochondrion. 2006; 6(5):258-262.
17. Miyamoto H, Machida RJ, Nishida S. Complete

mitochondrial genome sequences of the three pelagic chaetognaths *Sagitta nagae, Sagitta decipiens* and *Sagitta enflata*. Comp. Biochem. Physiol. Part D Genomics Proteomics. 2010; 5(1):65-72.

18. Li P, Yang M, Ni S, Zhou L, Wang Z, Wei S *et al*. Complete mitochondrial genome sequence of the pelagic chaetognath, *Sagitta ferox*. Mitochondrial DNA A, DNA Mapp. Seq. Anal. 2016; 27(6):4699-4700.
19. Shen X, Sun S, Zhao FQ, Zhang GT, Tian M, Tsang LM *et al*. Phylomitogenomic analyses strongly support the sister relationship of the Chaetognatha and Protostomia. Zool. Scripta. 2016; 45(2):187-199.
20. Wei S, Li P, Yang M, Zhou L, Yu Y, Ni S *et al*. The mitochondrial genome of the pelagic chaetognath, *Pterosagitta draco*. Mitochondrial DNA Part B. 2016; 1(1):515-516.
21. Doublet V, Ubrig E, Alioua A, Bouchon D, Marcadé I, Maréchal-Drouard L. Large gene overlaps and tRNA processing in the compact mitochondrial genome of the crustacean *Armadillidium vulgare*. RNA Biol. 2015; 12(10):1159-68.
22. Faure E, Barthélémy R. True mitochondrial tRNA punctuation and initiation using overlapping stop and start codons at specific and conserved positions. In *Mitochondrial DNA*; Seligmann H, IntechOpen, Croatia, 2018, 3-29.
23. Lowe TM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. 1997; 25(5):955-964.
24. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 1994; 22(22):4673-4680.
25. Jühling F, Mörl M, Hartmann RK, Sprinzl M, Stadler PF, Pütz J. tRNAdb 2009: compilation of tRNA sequences and tRNA genes. Nucleic Acids Res. 2009; 37:D159-162.
26. Marlétaz F, Martin E, Perez Y, Papillon D, Caubit X, Lowe CJ *et al*. Chaetognath phylogenomics: a protostome with deuterostome-like development. Curr. Biol. 2006; 16(15):R577-578.
27. Lavrov DV, Pett W. Animal mitochondrial DNA as we do not know it: mt-genome organization and evolution in nonbilaterian lineages. Genome Biol. Evol. 2016; 8(9):2896-2913.
28. Gould SJ, Vrba ES. Exaptation: a missing term in the science of form. Paleobiology. 1982; 8(1):4-15.
29. Doublet V, Helleu Q, Raimond R, Souty-Grosset C, Marcadé I. Inverted repeats and genome architecture conversions of terrestrial isopods mitochondrial DNA. J. Mol. Evol. 2013; 77(3):107-118.
30. Yokobori SI, Paabo S. tRNA editing in metazoans. Nature. 1995; 377(6549):490.
31. Levy S, Schuster G. Polyadenylation and degradation of RNA in the mitochondria. Biochem. Soc. Trans. 2016; 44(5):1475-1482.
32. Di Giulio M. A comparison among the models proposed to explain the origin of the tRNA molecule: A synthesis. J. Mol. Evol. 2009; 69(1):1-9.
33. Di Giulio M. The origin of the tRNA molecule: independent data favor a specific model of its evolution. Biochimie. 2012; 94(7):1464-1466.
34. Fujishima K, Kanai A. tRNA gene diversity in the three domains of life. Front. Genet. 2014; 5:142.
35. Maizels N, Weiner AM. Phylogeny from function: evidence from the molecular fossil record that tRNA originated in replication, not translation. Proc. Natl. Acad. Sci. USA. 1994; 91(15):6729-6734.
36. 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(7):e2799.
37. Duret L, Galtier N. Biased gene conversion and the evolution of mammalian genomic landscapes. Annu. Rev. Genomics Hum. Genet. 2009; 10:285-311.
38. Higgs PG, Jameson D, Jow H, Rattray M. The evolution of tRNA-Leu genes in animal mitochondrial genomes. J. Mol. Evol. 2003; 57(4):435-445.
39. Cornuet JM, Garnery L, Solignac M. Putative origin and function of the intergenic region between COI and COII of *Apis mellifera* L. mitochondrial DNA. Genetics. 1991; 128(2):393-403.
40. Shackelton LA, Holmes EC. The role of alternative genetic codes in viral evolution and emergence. J. Theor. Biol. 2008; 254(1):128-34.
41. Seligmann H. Alignment-based and alignment-free methods converge with experimental data on amino acids coded by stop codons at split between nuclear and mitochondrial genetic codes. Biosystems. 2018; 167:33-46.
42. The Genetic Codes compiled at National Center for Biotechnology Information (NCBI). https://www.ncbi.nlm.nih.gov/Taxonomy/Utils/wprintgc.cgi. 17 April 2019.
43. Shahi S, Eusebio-Cope A, Kondo H, Hillman BI, Suzuki N. Investigation of host range of and host defense against a mitochondrially replicating mitovirus. J. Virol. 2019; 93(6):e01503-01518.
44. Nibert ML. Mitovirus UGA (Trp) codon usage parallels that of host mitochondria. Virology. 2017; 507:96-100.
45. Nibert ML, Vong M, Fugate KK, Debat HJ. Evidence for contemporary plant mitoviruses. Virology. 2018; 518:14-24.