

HYPOTHESIS

Molecular evidence for the evolution of the eukaryotic mitochondrial arginyl-tRNA synthetase from the prokaryotic suborder Cystobacterineae

Gabor L. Igloi 

Institute of Biology, University of Freiburg, Germany

Correspondence

G. L. Igloi, Institute of Biology, University of Freiburg, Schänzlestr 1 D-79104, Freiburg, Germany
 Tel.: +49 7664 6870
 E-mail: igloi@biologie.uni-freiburg.de

(Received 26 September 2019, revised 25 October 2019, accepted 28 October 2019, available online 20 November 2019)

doi:10.1002/1873-3468.13665

Edited by Michael Ibba

The evolutionary origin of the family of eukaryotic aminoacyl-tRNA synthetases that are essential to all living organisms is a matter of debate. In order to shed molecular light on the ancient source of arginyl-tRNA synthetase, a total of 1347 eukaryotic arginyl-tRNA synthetase sequences were mined from databases and analyzed. Their multiple sequence alignment reveals a signature sequence that is characteristic of the nuclear-encoded enzyme, which is imported into mitochondria. Using this molecular beacon, the origins of this gene can be traced to modern prokaryotes. In this way, a previous phylogenetic analysis linking Myxococcus to the emergence of the eukaryotic mitochondrial arginyl-tRNA synthetase is supported by the unique existence of the molecular signature within the suborder Cystobacterineae that includes Myxococcus.

Keywords: bacterial endosymbiont; eukaryotic arginyl-tRNA synthetase; mitochondrial gene; molecular evolution; multiple sequence alignment; sequence conservation

Aminoacyl-tRNA synthetases are ancient enzymes that are essential to all forms of life. Arginyl-tRNA synthetase is present in all organisms [1] and is responsible for joining L-Arg to its cognate tRNA in the first step of protein synthesis [2], as well as, paradoxically, in the first step of the N-end rule proteasomal degradation [3]. It is a member of the class I aminoacyl-tRNA synthetase family that possess a Rossmann fold (typified by HIGH and KMSK motifs). These motifs constitute a portion of the catalytic site in three dimensions and play an important role in catalysis. Uniquely, among all class I enzymes, the majority of arginyl-tRNA synthetase species lack the typical KMSK sequence. Thus, the mechanism by which the catalytic reaction is achieved differs from other class I enzymes [4]. The crystal structure of the yeast arginyl-tRNA synthetase complexed with its cognate tRNA

[5] revealed details of the significance of the characteristic motifs in their interaction with the tRNA in this organism. According to Woese, 'In its evolutionary profile, ArgRS is arguably the most complex of all the aminoacyl-tRNA synthetases...' [6]. To assist in shedding some light on this complexity within the eukaryotic kingdom, a comprehensive collection of 1347 arginyl-tRNA synthetase amino acid sequences has been compiled, while being aware that this constitutes a snap-shot of the eukaryotic data sets available up to the first half of 2018. The evolutionary history of prokaryotic aminoacyl-tRNA synthetases has been explored recently [7]. Evolutionary aspects covering a single member of the eukaryotic aminoacyl-tRNA synthetase family have not been described.

One aspect of the evolutionary complexity is the fact that with the appearance of mitochondria, two distinct

Abbreviation

Cyto, cytoplasmic; Mito, mitochondrial; TSA, transcriptome shotgun assembly.

forms of arginyl-tRNA synthetase have emerged. One is the nuclear-encoded gene that is responsible for the cytoplasmic (Cyto) translation, whereas the nuclear-encoded, so-called mitochondrial (Mito) gene that migrated from the ancient symbiont produces an enzyme which is transported to the mitochondria to participate in mitochondrial protein synthesis. The functionality of the enzyme has been retained through evolution and one might expect that domains essential for its activity have been strongly conserved.

With the possibility of extracting and comparing numerous eukaryotic arginyl-tRNA synthetase sequences from the database, it was of interest to investigate whether traces of the ancient Mito gene product, as identified by a distinct sequence motif, could be detected in the modern prokaryotic superkingdoms. This notion appeared likely, particularly in the case of metazoans whose Mito arginyl-tRNA synthetase have had to coevolve with the structurally atypical tRNAs encoded by the Mito genome [8].

Materials and methods

Arginyl-tRNA synthetase sequences were obtained by online TBLASTN searches of the NCBI transcriptome (TSA), EST, and WGS databases of the appropriate class of organisms using the default search parameters. Additional sequences could be recovered from specialized resources such as <http://medicinalplantgenomics.msu.edu/> and <https://eupathdb.org/eupathdb/>. TSA and EST hits were translated directly [9]. WGS hits that included intronic regions were transferred to FGENESH+ [10] and scanned for protein similarity using the corresponding or closely related organism-specific gene-finding parameters. In some instances, after inspection of the prediction, pre-existing gene annotations were ignored to obtain more obviously acceptable alignments. Multiple protein alignments were performed online using MUSCLE [11] using default parameters, and the CLUSTALW outputs were transferred for visualization and to obtain consensus sequences to GENEDOCv2.7 [12], and stored as FASTA alignments [13].

Results and Discussion

Cytoplasmic versus Mitochondrial Arginyl-tRNA Synthetase

In arginyl-tRNA synthetase, the so-called KMSK region is an essential component of the catalytic site and is involved in the stabilization of the transition state of the amino acid activation reaction. In the alignment of the functionally conserved KMSK domain [4], it was noted that this region in

Saccharomyces cerevisiae and *S. pombe* diverged from the 15 other species (both prokaryotes and eukaryotes) in having a pentapeptide deletion. The gene for the ancient Mito form has, in all eukaryotes, migrated to the nucleus where it has either replaced the Cyto form or it coexists with its Cyto version. A nuclear gene replacement and duplication of the Mito form of arginyl-tRNA synthetase had previously been shown to have taken place in *S. cerevisiae* [14]. It was, therefore, of interest to investigate whether the pentapeptide deletion in yeasts could provide a Mito footprint that had been retained through evolution and that could be traced to a prokaryotic precursor. A previous compilation of 138 sequences (last updated in 2003) (http://rose.man.poznan.pl/aars/seq_main.html) was of little help as it contained data for just four higher eukaryotes.

Eukaryotic genomic and TSA databases were screened using TBLASTN and a probe that had been established as being either a Cyto or a Mito arginyl-tRNA synthetase sequence. This generated a total of 1347 derived protein sequences corresponding to 855 Cyto and 492 Mito [13] arginyl-tRNA synthetases. Eukaryotes were divided into Embryophyta, Metazoa, Non-Metazoa, and Fungi and to reduce the complexity of the sequence alignment a further subdivision of Metazoa into Chordata, Arthropoda, and Other Metazoa was undertaken.

Despite the evolutionary diversity of eukaryotes ranging from Chordata over Protists to Fungi, an alignment of hundreds of arginyl-tRNA synthetase sequences [13] and comparison of the resulting consensus sequences has revealed the existence of a robust sequence element [M(k/s)TR] (Fig. 1) in the KMSK domain of the protein that, together with a preceding pentapeptide deletion (referred to as 5 Δ MSTR), is characteristically and uniquely present in all Mito forms and was then used to follow its evolutionary appearance.

Similarly, but not quite as obviously conserved in the N-terminal region is a domain associated mainly with the enzyme of Cyto origin (Fig. 2). This is located N-terminal to the classic HIGH motif and, within the frequently highly dissimilar N-terminal region of the protein, marks the first significant patch of conserved amino acids.

This motif, which extends to chordate Mito sequences, has been described previously [7] but its functionality remains obscure, in particular since the GDYQ region is not present in the yeast enzyme and has evaded crystallization in the only higher eukaryote structure known to date [15]. The motif QCNNAM may be involved in RNA interaction as it appears in

```

Embryophyta (149 Sequences)      : HvGFGLVLGxDGKRFRtRxxevVrL
Chordata Cyto (205 Sequences)    : HagFGVVLGEDKKKFKTRsGxTVRL
Arthropoda Cyto (158 Sequences)  : hxxFGvVLGeDxKKFKTRSGdTvXL
Other Metazoa Cyto (152 Sequences): hxxFGxVLGeDxKKfKtRsGxtxxL
Non-Metazoa Cyto (191 Sequences): HxxFGxVxgxxgKxxxTRxGxLxxL
Chordata Mito (128 Sequences)    : HvxFGxVqG----MKTRxGxVxFL
Arthropoda Mito (202 Sequences)  : HxxFGrxxG----MstRxGxxxFL
Other Metazoa Mito (72 Sequences): HxxfGrxxg----xstRxGxxxxx
Non-Metazoa Mito (27 Sequences) : HxxxGxxxg----mstRxGxvxFL
Saccharomyces cerevisiae Cyto   : HVNFGMVQG----MSTRKGTVVFL
Saccharomyces cerevisiae Mito   : HINFGMVQG----MSTRKGNVFL
Fungi (63 Sequences)            : HxxfGxvxg----MSTRKGTVVFL

```

399 418

Fig. 1. Alignment of consensus sequences covering the KMSK region of arginyl-tRNA synthetases from different groups of organisms and their comparison to *Saccharomyces cerevisiae* and fungi, in general. Upper case letters indicate 100% conservation within the original group of organisms. Lower case letters correspond to > 80% conservation within the original group of organisms. x denotes < 80% conservation. Gaps introduced for optimal alignment are shown by a hyphen. Red letters show 100% identity within this alignment, green marks 80–99% identity within this alignment, blue and black signify 50–80% and < 50% identity, respectively. Cyto, cytoplasmic; Mito, mitochondrial. The numbering within the yeast cytoplasmic sequence is given. Consensus sequences were obtained from multiple sequence alignment [13].

```

Embryophyta (149 Sequences)      : xfgDYQCNNAMxL
Chordata Cyto (205 Sequences)    : KFGDYQCNSAMxi
Arthropoda Cyto (158 Sequences)  : kxxDYQcNxmxx
Other Metazoa Cyto (152 Sequences): xxxDYQcnxaxxx
Non-Metazoa Cyto (191 Sequences): xxxdFYqxxxaxxx
Chordata Mito (128 Sequences)    : x-ADFqlsxx-sx
Arthropoda Mito (202 Sequences)  : xxxxxxxxxxxxxx
Other Metazoa Mito (72 Sequences): xxxxxxxxxxxxxx
Non-Metazoa Mito (27 Sequences) : xxxxxxxxxxxxxx
Fungi (63 Sequences)            : xxxxxxxxxxxxxx

```

Fig. 2. Alignment of consensus sequences covering the GDYQ region of arginyl-tRNA synthetases from different groups of organisms and their comparison to fungi. Upper case letters indicate 100% conservation within the original group of organisms. Lower case letters correspond to 80–99% conservation within the original group of organisms. x denotes < 80% conservation. Blue marks > 60% identity within this alignment (using F/Y equivalence), with italics highlighting the weak cytoplasmic footprint. Cyto, cytoplasmic; Mito, mitochondrial. Consensus sequences were obtained from multiple sequence alignment [13].

numerous RNases (e.g., XP_002931285, XP_007537145, XP_006731454, XP_008561618, etc.).

Although this feature is only a rough guide to a Cyto characteristic, the pentapeptide insert and the KFKTR-like motif that makes up the KMSK domain within the C-terminal region is conserved in the Cyto enzymes. Using a combination of these footprints, it was possible to assign each of the sequences in the collection to either a Mito or to a Cyto origin irrespective of the existence of a database annotation.

In the case of fungi, alignment of the derived protein sequences [13] revealed that the majority of species within this kingdom only possess Mito-like arginyl-tRNA synthetases (as in *S. cerevisiae*). Despite the loss of mitochondria in Microsporidia, the ancient Mito gene has replaced the Cyto version, as was observed for amitochondrial *Entamoeba* [13]. Remarkably, some exceptions, within different phyla, were detected.

These, apparently atypical occurrences, were subjected to a closer inspection. For example, the sequence from *Termitomyces* sp. (Basidiomycota) that exists in symbiosis with termites, on BLAST analysis gives hits with 85% identity with the Cyto arginyl-tRNA synthetase of the termites *Cryptotermes secundus* (XM_023852624) and *Zootermopsis nevadensis* (XM_022069742). This may, therefore, be considered to be a result of sample contamination [16] or, intriguingly, of horizontal gene transfer. Similar arguments may be applied to the apparent Cyto-like sequences extracted for *Schizophyllum commune* from rotting wood, the fungus *Craterellus lutescens* that exists as a symbiont with trees, and the obligate plant pathogen *Puccinia arachidis*. On the other hand, *Microbotryum* is a parasite of the Caryophyllaceae plant family but the *Microbotryum* (Cyto-like) sequence does not share a high similarity (31%) with the typical host *Silene latifolia*

```

S. cerevisiae_cyto :GDFLEKVEANGP-FIQFFFNQPFLAKL
Bacteria (86 sequences) :xxxxxxExA-GP-FINxxLxxxxLxxx
MetazoaCyto (500 sequences) :xExxxxxIA-GPGFINxxLxKxVvxxx
MetazoaMito (400 sequences) :xxxxxxxxLxxxx-----xxxxxxxxx

```

106

111

Fig. 3. Alignment of consensus sequences covering the AGPGFIN region of arginyl-tRNA synthetases from various groups of organisms and their comparison to the cytoplasmic *S. cerevisiae* protein. Capital letters in the consensus sequences denote > 60% identity within the original multiple sequence alignments [13]. Green letters denote 75% identity within this alignment. Gaps (-) have been introduced to optimize the alignment. Cyto, cytoplasmic; Mito, mitochondrial. Numbering is that of the yeast enzyme.

(FMHP01018389). Although Zoopagomycota appear to encode two types of Mito arginyl-tRNA synthetase, the sequence *Entomophthora muscae2* has Cyto characteristics. The sequence originated from a TSA database and there is no corresponding genomic sequence. This species is parasitic on flies and shares an 88% identity with the Cyto sequence of the flesh fly, *Sarcophaga peregrine* (GGEP01011577). It is, then, likely that in the fungal kingdom the ancient Cyto arginyl-tRNA synthetase gene has been replaced by the Mito counterpart, either as a duplicate or to encode a dual-targeted protein. Recovery of the Cyto gene by horizontal transfer into symbiotic or parasitic fungal species cannot be excluded on the basis of the available data.

In certain other instances, the loss or retention of the Cyto enzyme can be related to the structure of the tRNA that is to be recognized by the arginyl-tRNA synthetase. Cyto and Mito aminoacyl-tRNA synthetases frequently recognize distinct identity elements within their cognate tRNAs. This prevents hetero-organellar aminoacylation [17]. In the case of arginyl-tRNA synthetase, the enzyme from *Escherichia coli* requires nucleotide A20 as a major identity element in tRNA and is unable to arginylate the corresponding yeast tRNA that possesses C20 [18], although the yeast enzyme recognizes the bacterial tRNA. Extrapolating this observation to the Cyto / Mito system, it is well-established that the Cyto arginyl-tRNA synthetase in mammals and plants relies on the presence of A20 [19,20]. Conversely, Coleopteran Mito tRNA^{Arg} lacks a canonical A20 base and is not aminoacylated by the (Cyto-like) *E. coli* enzyme [21].

Matching identity element evolution to recognition by the aminoacyl-tRNA synthetase requires a coevolution of the macromolecular partners [22–25]. In the case of nonmetazoans, if one excludes questionable sequences, Mito enzymes are found in just 6 phyla/classes. Of these, only Amoebozoa are non-Opisthokonts and possess both Cyto and Mito forms of the enzyme. An examination of the encoded arginyl-tRNA synthetase shows that Dictyostelium possesses a yeast-

like Mito arginyl-tRNA synthetase with a 5 Δ MSTR deletion. In this organism, Mito tRNA_{UCA} has U20 which would not be recognized by a Cyto enzyme. In contrast, the Cyto tRNA_{UCU}^{Arg} and tRNA_{ACG}^{Arg} have A20, the essential identity element for the Cyto form of the enzyme.

A correlation between the nature of the MSTR residues and N20 recognition is unlikely. At a structural level, C20 in yeast tRNA does not interact with the MSTR residues of the yeast enzyme (of Mito origin) but with N₁₀₆F₁₀₉Q₁₁₁ [5], although this interaction is unimportant for catalysis [26]. Indeed, the ‘AGPFIN’ motif that contains this tRNA binding domain is, in general, well-conserved in bacterial and metazoan Cyto arginyl-tRNA synthetase (Fig. 3). A notable metazoan exception is seen in the two available Echinodermata species in which the apparently unrelated VE^{K/R}GVLY sequence is found [13]. On the other hand, in the consensus sequence generated from the aligned metazoan Mito sequences no such tRNA binding motif is recognizable in a highly heterogeneous region of the protein. This is consistent with the reduction of metazoan tRNA^{Arg} structure to a noncanonical form [8] with frequent truncation or loss of the nucleotide 20-containing D stem-loop.

Monosiga brevicollis and *Salpingoeca rosetta* are the only Choanozoans (family Salpingoecidae), to date, for which both Cyto and Mito arginyl-tRNA synthetases have been detected. The *Monosiga* Cyto sequence is included in the compilation [13], whereas an incomplete *Salpingoeca* Cyto sequence with a clear-cut GDYQ-like motif can be recovered from the WGS database (ACSY01002895). The Mito sequences are characterized by the missing GDYQ-like domain and the typical 5 Δ MSTR-like feature. It is unlikely to be coincidental that the tRNA_{UCU}^{Arg} and tRNA_{UCG}^{Arg} encoded in the Mito genome possess a yeast (Mito)-like U20, to which the Mito enzyme is insensitive [26], whereas the nuclear-encoded tRNA_{ACG} has an A20 that is an essential identity element for recognition by the Cyto arginyl-tRNA synthetase [27]. It is notable that *Monosiga* and *Salpingoeca* are the only ones in a collection

Fig. 4. Alignment of consensus sequences covering the KMSK (A) and GDYQ regions (B) of arginyl-tRNA synthetases from different classes of prokaryotes (of those that provided more than three sequences). Upper case letters indicate 100% conservation within the original group of organisms. Lower case letters correspond to > 80% conservation within the original group of organisms. x denotes < 80% conservation. Green marks 80–95% identity within this alignment. Blue and black signify 50–80% and < 50% identity, respectively. Gaps (-) have been introduced to optimize the alignment. The 5 Δ MSTR domain in (A) is given in bold. Alphaproteobacteria and Deltaproteobacteria with their suborders are bracketed.

A	
Alphaproteobacteria (16 Sequences);	xxxxxxxxxxxxxxxxkxxxxxxxx-xxxxxxLxxx
Alphaproteobacteria without Rickettsiales (12 Sequences);	HxxfxGTxxGxDGKPFkTRx-Ggx-xxLxDL
Rickettsiales (4 Sequences);	hxxVNFxeNGxpvK-MSKRx-GxF-ltxxDV
Betaproteobacteria (13 Sequences);	HvgFG-xMmGxDgKPFKTRs-GgTVK-LxxL
Chlamydia (8 Sequences);	HvpFGLvxLxxxGkKfxTRs-Gxxxx-Lxxl
Cyanobacteria (16 Sequences);	HVPFG-xVxGEDGKkxKTRs-GxTvr-LdLL
Deltaproteobacteria (12 Sequences);	HxxFG-xxxgxxxxxxxTRxxxgxvxxLxxx
Deltaproteobacteria without Cystobacterineae (7 Sequences);	Hxxfg-xvxxxxxxkxxxTRxxxGxxxxlxxL
Cystobacterineae (5 Sequences);	HvxFG-RxHG----- MSTRK -GxVVxLxdVL
Gammaproteobacteria (8 Sequences);	HxxFGxxlxtxxGkklS TRK -Gxxxx-Lxxx
B	
Alphaproteobacteria (16 Sequences);	xxxxxxNxaxxx
Alphaproteobacteria without Rickettsiales (12 Sequences);	xxxxQxNGaxxx
Rickettsiales (4 Sequences);	hGDxytNxAXxx
Betaproteobacteria (13 Sequences);	FGDYQaNGVMxA
Chlamydia (8 Sequences);	fGhyqxnxAmkL
Cyanobacteria (16 Sequences);	FGDYQxNxAXxL
Deltaproteobacteria (12 Sequences);	xxDxxxxxxxxxl
Deltaproteobacteria without Cystobacterineae (7 Sequences);	xxDxxxxxaxxx
Cystobacterineae (5 Sequences);	HGDxSFxTFPLA
Gammaproteobacter (8 Sequences);	xGDxafpaxfxl

of 12 Choanozoans that also have two distinct genes for the essential tRNA-modifying ‘CCA-enzyme’ (CCA tRNA nucleotidyltransferase) [28].

Cryptomycota, of which there are only two representatives in this compilation, possess exclusively the Mito enzyme and are, therefore, closer to other members of fungi where the Cyto gene has been replaced by the ancient Mito gene (see above). Paramicrosporidium Mito tRNA^{Arg}_{UCU} and tRNA^{Arg}_{ACG} (MTSL01000001) have U20 which can be recognized by the Mito enzyme. Rozella has no mitochondrion-encoded tRNA^{Arg} and would need to import the Mito-like enzyme and its cognate tRNA from the cytoplasm.

The existence of the Mito tRNA^{Arg} gene, therefore, demands the evolutionary retention of an arginyl-tRNA synthetase that is not dependent on the A20 identity element. On the other hand, replacement of the Cyto by the Mito arginyl-tRNA synthetase permits a relaxation of the evolutionary pressure to maintain the crucial A20 element.

Molecular evidence and the evolutionary history of the eukaryotic mitochondrial arginyl-tRNA synthetase

Compilation and phylogenetic analysis of prokaryotic aminoacyl-tRNA synthetases is well-established [6,7,29,30]. However, to complete the comparison and to approach further the origin of the Mito gene, 83

arginyl-tRNA synthetase sequences from a selection of bacteria (11 phyla or classes) were assembled [13] following a BLAST search with the Mito *Homo sapiens* sequence as a probe.

Alignment of the selected derived protein sequences [13] reveals that several of the Deltaproteobacteria (including multiple *Myxococcus* species) possess the characteristic Mito-like pentapeptide deletion followed by MSTR. A more detailed examination dividing the class into suborders shows that species from both major families within the suborder Cystobacterineae and from all five genera within the Myxococcaceae family have the Mito feature (Fig. 4). The characteristic Mito 5 Δ MSTR segment is lacking in all other prokaryotic classes.

However, two species within Cystobacterineae appeared to form an exception. Despite BLAST hits on more than 20 different *Coralloccoccus* isolates that confirm the Mito signature, the N-terminally incomplete sequence from *Coralloccoccus* sp. *CAG:1435* (from the human gut metagenome) (CBAX010000090) possesses the Cyto pentapeptide insert. Similarly, the genome from five different *Myxococcus* strains reveal the presence of the Mito-like arginyl-tRNA synthetase, whereas the genome of *Myxococcaceae bacterium isolate GW715* (SCUS01000160), belonging to the unclassified Myxococcaceae, encodes the Cyto-like protein with a recognizable ‘GDYQ’ N-terminal signature. These two species are also the only examples for which

a Cyto-like protein was found within the Cystobacterineae when using the *H. sapiens* Cyto sequence as a BLAST probe. The other available organisms within the suborders of the Myxococcales order possess the Cyto-like sequence elements.

Within the endosymbiont order Rickettsiales, which, on the one hand, is considered to be related to the ancestral proto Mito symbiont [31] but, on the other hand, has been classified as being a special case because the gene is of archaeal origin [32] there is considerable divergence from the poorly defined Cyto-like GDYQ region found in other alphaproteobacteria (Fig 3). The Cyto feature is recognizable in Rickettsiales.

The evolutionary bacterial origin of the Mito arginyl-tRNA synthetase and its appearance in the nuclear genome has been discussed previously [33,34]. It has been suggested that transient endosymbioses, distinct from the Mito one, could also have been the source of such a gene [32,35]. More specifically a phylogenetic analysis of all aminoacyl-tRNA synthetases [36] leads to the proposal of a three-step acquisition of the two types of arginyl-tRNA synthetase. Firstly, in this model, eukaryotes acquired the Cyto arginyl-tRNA synthetase of Chlamydiae. Secondly, the common ancestor of Fungi, Amoebozoa, and Metazoa acquired arginyl-tRNA synthetase from Myxococcus as the Mito arginyl-tRNA synthetase, and thirdly, Cyto arginyl-tRNA synthetase of Fungi and Amoebozoa was replaced by Mito ones. According to the analysis presented here, although the Cyto form has been replaced by the Mito form in the amitochondrial Entamoeba [13], other Amoebozoans possess both forms of the enzyme. Consequently, a complete substitution of the ancient Cyto enzyme by the Mito enzyme is, so far, only evident in fungi. Furthermore, the molecular evidence provided by the multiple alignments given above does not necessarily confirm Chlamydiae as being the source of the Cyto enzyme. The Cyto GDYQ- and KFKTR- Cyto signatures are clearly recognizable in Cyanobacteria or Betaproteobacteria, as well as in Chlamydia whose proximity to the Cyto arginyl-tRNA synthetase has been noted previously [37]. However, the acquisition by Fungi, Amoebozoa, and Metazoa of the gene from Myxococcus (or a member of its suborder) to act as the Mito counterpart is well-supported at molecular level by the tell-tale 5 Δ MSTR footprint.

In summary, the phylogenetic analysis linking Myxococcus to the emergence of the eukaryotic Mito arginyl-tRNA synthetase [36] has been confirmed here directly by the molecular signature within the suborder Cystobacterineae that distinguishes the eukaryotic Cyto and Mito forms of the enzyme.

Acknowledgements

Open access funding enabled and organized by Projekt DEAL.

References

- 1 Bullwinkle TJ and Ibba M (2013) Emergence and Evolution. In Topics in current chemistry, pp. 43–87.
- 2 Ibba M and Söll D (2000) Aminoacyl-tRNA synthesis. *Annu Rev Biochem* **69**, 617–650.
- 3 Varshavsky A (2011) The N-end rule pathway and regulation by proteolysis. *Protein Sci* **20**, 1298–1345.
- 4 Sekine SI, Shimada A, Nureki O, Cavarelli J, Moras D, Vassilyev DG and Yokoyama S (2001) Crucial role of the high-loop lysine for the catalytic activity of arginyl-tRNA synthetase. *J Biol Chem* **276**, 3723–3726.
- 5 Delagoutte B, Moras D and Cavarelli J (2000) tRNA aminoacylation by arginyl-tRNA synthetase: induced conformations during substrates binding. *EMBO J* **19**, 5599–5610.
- 6 Woese CR, Olsen GJ, Ibba M and Söll D (2000) Aminoacyl-tRNA synthetases, the genetic code, and the evolutionary process. *Microbiol Mol Biol Rev* **64**, 202–236.
- 7 Chalot A, Vlastaridis P, Mossialos D, Ibba M, Becker HD, Stathopoulos C and Amoutzias GD (2017) The complex evolutionary history of aminoacyl-tRNA synthetases. *Nucleic Acids Res* **45**, 1059–1068.
- 8 Watanabe K (2010) Unique features of animal mitochondrial translation systems. The non-universal genetic code, unusual features of the translational apparatus and their relevance to human mitochondrial diseases. *Proc Jpn Acad Ser B Phys Biol Sci* **86**, 11–39.
- 9 Perry WL (2002) JavaScript DNA Translator: DNA-Aligned Protein Translations. *Biotechniques* **33**, 1318–1320.
- 10 Solovyyev V (2008) Statistical Approaches in Eukaryotic Gene Prediction. Handbook of Statistical Genetics, (Balding DJ, Bishop M and Cannings C, eds), pp. 97–159. John Wiley & Sons Ltd, Chichester.
- 11 Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**, 1792–1797.
- 12 Nicholas KB and Nicholas HBJ (1997) GeneDoc: a tool for editing and annotating multiple sequence alignments. 1997. *Distrib. by author.*
- 13 [Dataset] Igloi GL (2019) Compilation and Alignment of Eukaryotic Arginyl-tRNA Synthetases 2018, Mendeley Data, v1. <https://doi.org/10.17632/ts4jbw9nft.1>
- 14 Karlberg O, Canbäck B, Kurland CG and Andersson SGE (2000) The dual origin of the yeast mitochondrial proteome. *Yeast* **1**, 170–187.
- 15 Kim HS, Cha SY, Jo CH, Han A and Hwang KY (2014) The crystal structure of arginyl-tRNA synthetase from *Homo sapiens*. *FEBS Lett* **588**, 2328–2334.

- 16 Igloi GL (2019) Evidence for fungal sequence contamination in plant transcriptome databases. *Plant Syst Evol* **305**, 563–568.
- 17 Kumazawa Y, Himeno H, Miura K and Watanabe K (1991) Unilateral aminoacylation specificity between bovine mitochondria and eubacteria. *J Biochem* **109**, 421–427.
- 18 Liu W, Huang Y, Eriani G, Gangloff J, Wang E and Wang Y (1999) A single base substitution in the variable pocket of yeast tRNA(Arg) eliminates species-specific aminoacylation. *BBA* **1473**, 356–362.
- 19 Guigou L and Mirande M (2005) Determinants in tRNA for activation of arginyl-tRNA synthetase: evidence that tRNA flexibility is required for the induced-fit mechanism. *Biochemistry* **44**, 16540–16548.
- 20 Aldinger CA, Leisinger AK and Igloi GL (2012) The influence of identity elements on the aminoacylation of tRNA Arg by plant and *Escherichia coli* arginyl-tRNA synthetases. *FEBS J* **279**, 3622–3638.
- 21 Leisinger AK, Janzen DH, Hallwachs W and Igloi GL (2013) Amino acid discrimination by the nuclear encoded mitochondrial arginyl-tRNA synthetase of the larva of a bruchid beetle (*Caryedes brasiliensis*) from northwestern Costa Rica. *Insect Biochem Mol Biol* **43**, 1172–1180.
- 22 Salazar JC, Ahel I, Orellana O, Tumbula-Hansen D, Krieger R, Daniels L and Soll D (2003) Coevolution of an aminoacyl-tRNA synthetase with its tRNA substrates. *Proc Natl Acad Sci* **100**, 13863–13868.
- 23 Ardell DH and Andersson SGE (2006) TFAM detects co-evolution of tRNA identity rules with lateral transfer of histidyl-tRNA synthetase. *Nucleic Acids Res* **34**, 893–904.
- 24 Pett W and Lavrov DV (2015) Cytonuclear interactions in the evolution of animal mitochondrial tRNA metabolism. *Genome Biol Evol* **7**, 2089–2101.
- 25 Ho JM, Bakkalbasi E, Söll D and Miller CA (2018) Drugging tRNA aminoacylation. *RNA Biol* **15**, 667–677.
- 26 Geslain R, Martin F, Camasses A and Eriani G (2003) A yeast knockout strain to discriminate between active and inactive tRNA molecules. *Nucl Acids Res* **31**, 4729–4737.
- 27 Giegé R, Sissler M, Florentz C, Giegé R, Sissler M and Florentz C (1998) Universal rules and idiosyncratic features in tRNA identity. *Nucl Acids Res* **26**, 5017–5035.
- 28 Betat H, Mede T, Tretbar S, Steiner L, Stadler PF, Mörl M and Prohaska SJ (2015) The ancestor of modern Holozoa acquired the CCA-adding enzyme from alphaproteobacteria by horizontal gene transfer. *Nucleic Acids Res* **43**, 6739–6746.
- 29 O'Donoghue P and Luthey-Schulten Z (2003) On the Evolution of Structure in Aminoacyl-tRNA Synthetases. *Microbiol Mol Biol Rev* **67**, 550–573.
- 30 Luque I, Riera-Alberola ML, Andújar A and Ochoa De Alda JAG (2008) Intrapylum diversity and complex evolution of cyanobacterial aminoacyl-tRNA synthetases. *Mol Biol Evol* **25**, 2369–2389.
- 31 Ferla MP, Thrash JC, Giovannoni SJ and Patrick WM (2013) New rRNA gene-based phylogenies of the Alphaproteobacteria provide perspective on major groups, mitochondrial ancestry and phylogenetic instability. *PLoS ONE* **8**, e83383.
- 32 Philippe H, Germot A and Moreira D (2000) The new phylogeny of eukaryotes. *Curr Opin Genet Dev* **10**, 596–601.
- 33 Brown JR and Doolittle WF (1995) Root of the universal tree of life based on ancient aminoacyl-tRNA synthetase gene duplications. *Proc Natl Acad Sci USA* **92**, 2441–2445.
- 34 Hashimoto T, Sánchez LB, Shirakura T, Müller M and Hasegawa M (1998) Secondary absence of mitochondria in *Giardia lamblia* and *Trichomonas vaginalis* revealed by valyl-tRNA synthetase phylogeny. *Proc Natl Acad Sci USA* **95**, 6860–6865.
- 35 Sogin ML (1997) Organelle origins: energy-producing symbionts in early eukaryotes? *Curr Biol* **7**, R315–R317.
- 36 Furukawa R, Nakagawa M, Kuroyanagi T, Yokobori SI and Yamagishi A (2017) Quest for ancestors of eukaryal cells based on phylogenetic analyses of aminoacyl-tRNA synthetases. *J Mol Evol* **84**, 51–66.
- 37 Brindefalk B, Viklund J, Larsson D, Thollesson M and Andersson SGE (2007) Origin and evolution of the mitochondrial aminoacyl-tRNA synthetases. *Mol Biol Evol* **24**, 743–756.