

Review

# “Superwobbling” and tRNA-34 wobble and tRNA-37 anticodon loop modifications in evolution and devolution of the genetic code

Lei Lei <sup>1</sup> and Zachary Frome Burton <sup>2,\*</sup>

<sup>1</sup> Lei Lei, School of Biological Sciences, University of New England, Biddeford, Maine, USA 04005, [llei@une.edu](mailto:llei@une.edu)

<sup>2</sup> Zachary Frome Burton, Department of Biochemistry and Molecular Biology, Michigan State University, E. Lansing, MI, USA 48824-1319, Phone: 1-517-881-2243, [burton@msu.edu](mailto:burton@msu.edu) ORCID 0000-0003-1065-5222

\* Correspondence: [burton@msu.edu](mailto:burton@msu.edu); Tel.: 1-517-881-2243

**Abstract:** The genetic code evolved around the reading of the tRNA anticodon on the primitive ribosome, and tRNA-34 wobble and tRNA-37 modifications coevolved with the code. We posit that EF-Tu, the closing mechanism of the 30S ribosomal subunit, methylation of wobble U34 at the 5-carbon and suppression of wobbling at the tRNA-36 position were partly redundant and overlapping functions that coevolved to establish the code. The genetic code devolved in evolution of mitochondria to reduce the size of the tRNAome (all of the tRNAs of an organism or organelle). “Superwobbling” or four-way wobbling describes a major mechanism for shrinking the mitochondrial tRNAome. In superwobbling, unmodified wobble tRNA-U34 can recognize all four codon wobble bases (A, G, C and U), allowing a single unmodified tRNA-U34 to read a 4-codon box. During code evolution, to suppress superwobbling in 2-codon sectors, U34 modification by methylation at the 5-carbon position appears essential. As expected, at the base of code evolution, tRNA-37 modifications mostly related to the identity of the adjacent tRNA-36 base. tRNA-37 modifications help maintain the translation frame during elongation.

**Keywords:** elongation factor Tu; evolution of the genetic code; four-way wobbling; genetic code degeneracy; inosine; mitochondria; queuosine; superwobbling; tRNA modification; tRNA wobble U modification

## 1. Introduction

This review was written to support an interpretation of a confluence of recent and older data. We attempt to bring some simplicity, order and concept to what may seem, at first, like overwhelming complexity and confusion. The genetic code evolved in columns around the structure of the tRNA anticodon. Genetic code columns represent the middle position of the anticodon (tRNA-35), which is and was the easiest anticodon position to read. Initially, tRNA-34 and tRNA-36 were wobble positions, but wobbling was suppressed at tRNA-36, in part, by tRNA-37 modifications. Appreciation of tRNA anticodon loop structure and reading helps to explain genetic code structure and the evolution of tRNA modifications that affect reading of the anticodon.

Notably, “superwobbling” or four-way wobbling in evolution of the mitochondria has been described and supported by detailed tRNA modification data [1-5]. Phylogenetics indicates pathways of evolution of Archaea, ancient Bacteria, derived Bacteria and Eukarya [5,6]. Evolution of the mitochondria from a derived bacterial endosymbiont is fundamental to understand evolution of Eukarya [7-10]. Superwobbling indicates the

importance of ancient wobble U34 methylation-based modifications at the 5-carbon position. In the mitochondrion, unmodified wobble U34 can potentially read wobble codons ending in A, G, C and U to translate an entire 4-codon sector of the code using a single tRNA species [1,2,5]. At the base of genetic code evolution, however, it appears that tRNA-U34 may have often or always been modified, in part, to suppress superwobbling and to allow evolution of 2-codon sectors [3,4,11,12]. Recent tRNA modification data support this idea. To our knowledge, the relationship of superwobbling to initial genetic code evolution has, for the most part, not been discussed (but see [13]). We posit that 5-carbon U34 methylation-based wobble modifications were essential for the initial evolution of the genetic code.

Similarly, tRNA wobble adenosine deamination to inosine (tRNA-A34→I34) modifications appear fundamental to the later evolution and enrichment of the code [14-18]. I34, generally, can read wobble codons A, C and U, and the I34 modification is associated with the suppression of synonymous G34 anticodons. G34 is favoured in Archaea and, for the most part, in Bacteria [14]. Put another way, when the I34 wobble modification occurs, the corresponding G34 tRNA anticodon is rarely if ever present. Also, introduction of tRNAs with unnatural G34 anticodons can be toxic in Eukaryotes [14,19]. In Bacteria, A34→I34 modification is mostly found for the Arg anticodon (ACG→ICG). By contrast, in Eukarya, the A34→I34 wobble modification is found for Leu (AAG→IAG), Ile (AAU→IAU), Val (AAC→IAC), Ser (AGA→IGA), Pro (AGG→IGG), Thr (AGU→IGU), Ala (AGC→IGC) and, as in Bacteria, Arg (ACG→ICG). Interestingly, in Eukarya, Gly occupies a 4-codon box but does not utilize the A34→I34 modification. We offer two possible explanations below. Because of wobble ambiguity, the A34→I34 modification can only occur in 3- or 4-codon sectors of the genetic code. Some Bacteria encode A34 in 4-codon sectors other than Arg, but, in most of these cases, A34 does not appear to be converted to inosine [15,16]. Because of superwobbling in 4-codon sectors, the A34→I34 modification is not utilized in mitochondria [5]. In response to oxidative and starvation stress, Eukaryotes utilize endonuclease V to cleave I34 tRNAs to stall translation [20].

Bacteria utilize G34→Q34 modifications (Q for queuosine) [5,21-25]. These modifications are found in Eukaryotes, mitochondria and Bacteria but not in Archaea. In Archaea, the queuosine-related modification archaeosine, which involves a homologous enzyme, is found at the G15-position of tRNAs. In humans, queuine is a necessary coenzyme supplied by diet and generated by symbiotic enteric bacteria. Q34 modifications cause more balanced reading of NAU and NAC codons, so the lack of queuosine modifications slows translation [21,22]. Queuosine modifications are only found in column 3 of the genetic code (GUN→QUN anticodons).

Modifications of the anticodon loop tRNA-37 position, just 3' to the anticodon, also appear to be of importance [14,18,25]. tRNA-37 modifications tend to be bulky next to an anticodon U36 or A36 and may help to stabilize intrinsically weaker anticodon-codon interactions. Modifications of tRNA-37 limit frameshifting during translation [26-28]. tRNA-35 and -36 are rarely modified and are generally read by Watson-Crick pairing to their mRNA codon. We posit that modifications of tRNA-37 help to delimit the anticodon, stabilize base pairing at position 36, stabilize the anticodon-codon interaction, suppress frameshifting [28] and perform other roles, for instance, recognition by aaRS enzymes to charge the cognate tRNA [29]. We find that, as expected, at the base of genetic code evolution, tRNA-37 modifications primarily depend on the adjacent tRNA-36 base, which corresponds to genetic code rows 1-4.

A new tRNA database helps to follow the current trends in the literature [30]. Older databases are also useful [31-34]. Updated modification data for tRNAs were essential to understand tRNA modification features of translation. Some tRNA modifications (i.e., cm<sup>5</sup>U34-based, t<sup>6</sup>A37 and m<sup>1</sup>G37) appear to be as old as the genetic code and, probably, were coevolved with the code and necessary for its initial establishment. Analysis of tRNA modifications at the tRNA-34 and -37 positions strongly supports the hypothesis that the genetic code evolved around the reading of the tRNA anticodon [35-37].

The archaeal genetic code is simplest and closest to the code that was present at LUCA (the last universal common (cellular) ancestor). We consider LUCA to be the first membrane-enclosed cells with intact DNA genomes. *Pyrococcus furiosus* is a reasonable reference organism for an ancient Archaeon and an approximation of LUCA [38,39]. The code is simpler in older bacterial species such as *Thermus thermophilus*, compared to more derived Bacteria, such as *Escherichia coli* and  $\alpha$ -Proteobacteria. It appears that the mitochondria were derived from an  $\alpha$ -Proteobacteria (*Rickettsiales*) [5-7,10,40,41]. The eukaryotic cytosolic code was derived from Archaea with contributions from an  $\alpha$ -proteobacterial endosymbiont. Thus, the genetic code can be mostly traced, along with relevant tRNA modification data through evolution of life on Earth [18]. Currently, there is missing tRNA modification data for ancient Bacteria, such as *Thermus thermophilus*. At the time of writing, sequences of only ~5 modified *T. thermophilus* tRNAs have been reported out of a total of about 47 tRNAs. At the time of writing, no *T. thermophilus* tRNA with a modified or unmodified U34 has yet been reported [30]. Combining these missing data with this paper would be a useful contribution.

Aminoacyl-tRNA synthetases (aaRS) attach cognate amino acids to the 3'-ends of tRNAs [29,35,42]. Evolution of aaRS enzymes has been described in detail. AaRS are of the two incompatible folding classes I and II with structural subclasses A→E. The class II aaRS GlyRS-IIA was refolded into a class I aaRS (probably a primitive ValRS-IA). In addition to their incompatible fold, class I aaRS have an in-phase N-terminal extension relative to class II aaRS. The class II aaRS mounts the enzyme active site on a surface of anti-parallel  $\beta$ -sheets. By contrast, the class I aaRS mounts the enzyme active site at the C-terminal ends of a set of parallel  $\beta$ -sheets. GlyRS-IIA (glycine aminoacyl-tRNA synthetase; class II; structural subclass A) is the root of all aaRS enzymes. In ancient Archaea, GlyRS-IIA is a sequence homolog of ValRS-IA and IleRS-IA. Tracing the evolution of aaRS enzymes describes the evolution of the genetic code. The genetic code evolved from Archaea to ancient Bacteria to more derived Bacteria. Eukarya are a fusion of multiple Archaea and multiple Bacteria probably involving a number of endosymbionts and/or other large horizontal gene transfers [6,10,43]. We find that a simple narrative for the evolution of life on Earth is obtained by comparing genetic codes, tRNA-34 and tRNA-37 modifications, aaRS and tRNAome data from a small number of reference organisms.

## 2. Evolution of the genetic code around the tRNA anticodon

In Figure 1, the *Saccharomyces cerevisiae* tRNA<sup>Phe</sup> anticodon loop is shown (PDB 1EHZ) [44]. In Figure 1A, the linear modified sequence is shown. In Figure 1B, the folded structure is indicated. Figures 1C-1E are three orientations of the anticodon loop structure including part of the anticodon stem. The genetic code evolved around the structure of the tRNA anticodon. The anticodon triplet is tRNA positions 34, 35 and 36. TRNA-34 is the wobble position at which diverse wobble contacts to mRNA codons are allowed, adjusted and tuned in evolution. TRNA-35 is the central position, which represents genetic code columns and is the easiest position to read. TRNA-36 represents genetic code rows 1-4. Generally, the tRNA-35 and -36 positions are read during translation as Watson-Crick base pairs versus the mRNA codon. As in *S. cerevisiae* tRNA<sup>Phe</sup>, tRNA-35 and -36 are generally unmodified.

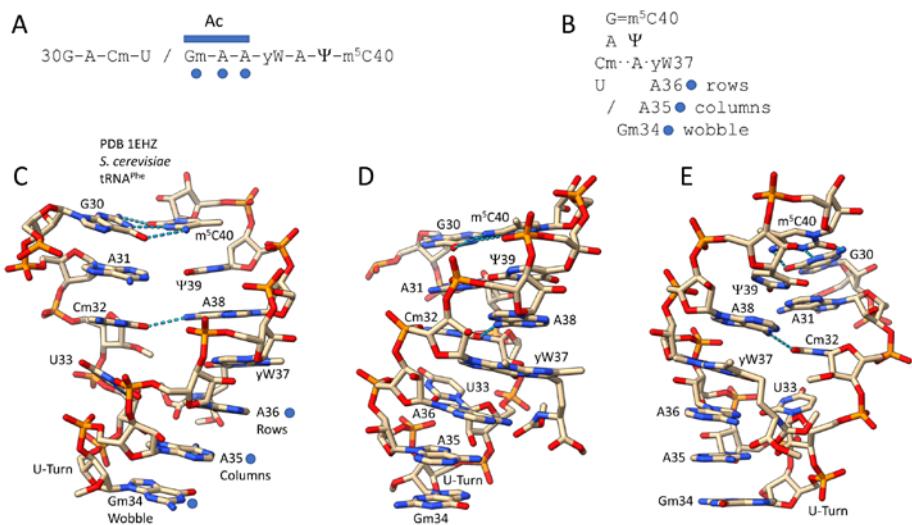


Figure 1. The *Saccharomyces cerevisiae* tRNA<sup>Phe</sup> anticodon loop (PDB 1EHZ). A) The linear sequence is shown. The anticodon (Ac) is indicated (3 blue dots). B) The folded loop structure is shown. / indicates a U-turn. C-E) Three views of the anticodon loop are shown. The anticodon is indicated in C (blue dots). Blue dashed lines indicate H-bonds. Colors: beige) carbon; blue) nitrogen; red) oxygen; orange) phosphorous. Abbreviations: Cm) 2'-O-methylcytidine; Gm) 2'-O-methylguanosine; yW) wybutosine;  $\Psi$ ) pseudouridine;  $m^5C$  5-methylcytidine.

A detailed and rational model for pre-LUCA evolution of the genetic code has been published [35-37]. The genetic code is highly structured and more simply structured in Archaea than in other organisms. Most evolution is in code columns, which represent the tRNA-35 base. For instance, in column 1 (tRNA-35A), related hydrophobic amino acids Val, Met, Ile and Leu are found, and these chemically similar amino acids are added to their cognate tRNAs by ValRS-IA, MetRS-IA, IleRS-IA and LeuRS-IA, which are closely-related aaRS class IA enzymes. Similarly, in column 2 (tRNA-35G), amino acids Thr, Pro and Ser are found. Thr and Ser are closely-related amino acids, and ThrRS-IIA, ProRS-IIA and SerRS-IIA are closely-related aaRS class IIA enzymes. The code is proposed to have evolved through stages. Initially, both tRNA anticodon positions 34 and 36 were wobble positions, at which only 2-assignments (purine versus pyrimidine) were possible. Wobbling was suppressed at position 36 by evolution of EF-Tu, the 16S rRNA "latch" (i.e., G530~A1492 and A1493; *Thermus thermophilus* numbering) [45,46] and modifications of anticodon loop position 37. Suppression of wobbling at position 36 allowed the code to expand from 8-amino acids (complexity  $2 \times 4$ ) to a maximum complexity of 32-assignments (complexity  $2 \times 4 \times 4$ ). Because of fidelity mechanisms, the standard genetic code froze at 20-amino acids + stops.

The primordial sequence of the 7-nt anticodon loop was close to 32-CU/BNNA-38 (/ indicates a U-turn; B=G, C or U (not A); N=A, G, C or U). In Figure 1, four bases (30, 31, 39 and 40) that are normally part of the anticodon stem are also shown. The G30= $m^5C$ 40 base pair is evident. The expected A30 $\rightarrow$  $\Psi$ 39 base pair was disrupted by the pseudouridine rearrangement, perhaps to adjust the conformation and dynamics of the loop. Typically, the loop includes a U-turn after U33. A U-turn is a U-shaped turn in the RNA backbone [47]. The U-turn loop conformation is important to present the 3-nt anticodon (tRNA-34, -35 and -36). The Cm32~A38 H-bond can be characterized as a weak reverse Hoogsteen pair Cm32 (O2) $\rightarrow$ A38 (N6). This interaction is thought to regulate the U-turn geometry and dynamics of the anticodon loop [18,25]. The yW37 (wybutosine) modification of G is a bulky modification that is thought to stabilize interactions of the A36 anticodon base with its cognate codon and also to suppress frameshifting during translation.

### 3. Evolution of life on Earth

A simple narrative for evolution of life on Earth is proposed in which LUCA evolved to Archaea [39,48]. As a reference organism that is close to LUCA, we propose *Pyrococcus furiosis*. *P. furiosis* has a tRNAome that is very similar in sequence to tRNA<sup>Pri</sup> (a primordial tRNA) [38]. We propose that Archaea evolved to ancient Bacteria, such as *Thermus thermophilus*. We selected *T. thermophilus* because it has a simple but intact tRNAome. Unfortunately, the reported tRNA modification data for *T. thermophilus* is not complete at the time of writing. As a model organism for more derived Bacteria, we relied mostly on *Escherichia coli*. If data were available, we would incorporate the closest bacterial relative of the eukaryotic mitochondria. *E. coli*, however, appears to be a reasonable model, albeit with several differences from the endosymbiont that became the mitochondria. We support the hypothesis that eukaryotic mitochondria were derived from an  $\alpha$ -proteobacterial endosymbiont within an Asgard Archaea [49]. Eukaryotes, however, arose as a complex set of genetic fusions of multiple Archaea and multiple Bacteria. For the purposes of this paper, we trace tRNA U34, A34 $\rightarrow$ I34 and G34 $\rightarrow$ Q34 modifications through evolution. We discuss maintenance of the Ile-Met sector. Maintenance of 1-codon sectors (i.e., for Met and Trp) in evolution was difficult and was abandoned during evolution of mitochondria [5]. We consider modifications of anticodon position 37 [18,50]. We combine these data with evolution of aaRS enzymes and analyses of tRNAomes. To our knowledge, these issues have largely not been raised or have not been integrated in this manner in published papers. We consider our presentation to be highly informative to describe the major advances in evolution of the genetic code through the natural biological history of Earth.

### 4. Ancient Archaea

In this paper, we present or approximate the genetic codes of several reference organisms including some related data. Figure 2 shows an approximation of the *Pyrococcus furiosis* genetic code. Because of missing tRNA modification data, some information has been taken from or inferred from other Archaea. At the time of writing, significant tRNA modification data is available for *Pyrococcus furiosis*, *Methanocaldococcus jannachii*, *Methanococcus maripaludis*, *Sulfolobus acidocaldarius* and *Haloferax volcanii* [3,4,11]. The genetic code is presented as a 64-assignment code. Codon sequence surrounds the table. Anticodon data is enriched with tRNA modification data mostly for the wobble base (tRNA-34). The amino acid and structural class (class I or II; structural subclasses A-E) of the aminoacyl tRNA synthetase (aaRS) enzymes was included. Anticodons that are not utilized in an organism or domain may be shown in red with strikethrough. To follow the narrative of this paper, all of these data are necessary to consider in order to compare genetic codes relevant to the generation of Eukarya and mitochondria.

First of all, A34, in which A is unmodified, is rarely or never allowed in Archaea [14]. Rather, in Archaea, G34 appears to always be utilized. As a wobble base, G34 has the advantage of pairing with codon wobble U, as a G~U wobble pair, or else with codon wobble C, as a Watson-Crick G=C pair. At the base of code evolution, U34 appears to seldom or never be unmodified, specifically by a methylation-based modification at the 5-carbon of U34 (cm<sup>5</sup>U34-based modifications). For the precise chemistry of tRNA modifications, please refer to the Modomics Database [25,31-34]. We propose that cm<sup>5</sup>U-based modifications (i.e., cnm<sup>5</sup>U in *P. furiosis*) suppress superwobbling that is observed for 4-codon sectors in mitochondrial tRNAs [1,2,5]. A cnm<sup>5</sup>U34 tRNA, therefore, is likely confined to read codon wobble A and G. Superwobbling, by contrast, would allow unmodified U34 to read A, G, C and U, which would prevent evolution of 2-codon sectors. To evolve 2-codon genetic code sectors (i.e., for columns 1, 3 and 4), therefore, required cm<sup>5</sup>U-based modifications.

		U	C	A	G	2nd				
1A1	U	Phe-IIC	GAA/ <b>AAA</b>	Ser-IIA	GGA/ <b>AGA</b>	Tyr-IC	GUA/ <b>AUA</b>	Cys-IB	GCA/ <b>ACA</b>	U
1A2	U	Phe-IIC	GAA	Ser-IIA	GGA	Tyr-IC	GUA	Cys-IB	GCA	C
1B1	U	Leu-IA	<b>cnm<sup>5</sup>UAA</b>	Ser-IIA	<b>cnm<sup>5</sup>UGA</b>	<b>STOP</b>	<b>UUA</b>	<b>STOP</b>	<b>UCA</b>	A
1B2	U	Leu-IA	CAA	Ser-IIA	CGA	<b>STOP</b>	<b>CUA</b>	Trp-IC	CCA	G
2A1	C	Leu-IA	GAG/ <b>AAG</b>	Pro-IIA	GGG/ <b>AGG</b>	His-IIA	GUG/ <b>AUG</b>	Arg-ID	GCG/ <b>ACG</b>	U
2A2	C	Leu-IA	GAG	Pro-IIA	GGG	His-IIA	GUG	Arg-ID	GCG	C
2B1	C	Leu-IA	<b>cnm<sup>5</sup>UAG</b>	Pro-IIA	<b>cnm<sup>5</sup>UGG</b>	Gln	<b>cnm<sup>5</sup>UUG</b>	Arg-ID	<b>cnm<sup>5</sup>UCG</b>	A
2B2	C	Leu-IA	CAG	Pro-IIA	CGG	Gln	<b>ac<sup>4</sup>CUG</b>	Arg-ID	CCG	G
3A1	A	Ile-IA	GAU/ <b>AUA</b>	Thr-IIA	GGU/ <b>AGU</b>	Asn-IIA	GUU/ <b>AUU</b>	Ser-IIA	GCU/ <b>ACU</b>	U
3A2	A	Ile-IA	GAU	Thr-IIA	GGU	Asn-IIA	GUU	Ser-IIA	GCU	C
3B1	A	Ile-IA	<b>agm<sup>2</sup>CAU/<b>UAU</b></b>	Thr-IIA	<b>cnm<sup>5</sup>UGU</b>	Lys-IE	<b>cnm<sup>5</sup>UUU</b>	Arg-ID	<b>cnm<sup>5</sup>UCU</b>	A
3B2	A	Met-IA	<b>CmAU/CAU</b>	Thr-IIA	CGU	Lys-IE	CUU	Arg-ID	CCU	G
4A1	G	Val-IA	GAC/ <b>AAC</b>	Ala-IIID	GGC/ <b>AGG</b>	Asp-IIID	GUC/ <b>AUC</b>	Gly-IIA	GCC/ <b>ACG</b>	U
4A2	G	Val-IA	GAC	Ala-IIID	GGC	Asp-IIID	GUC	Gly-IIA	GCC	C
4B1	G	Val-IA	<b>cnm<sup>5</sup>UAC</b>	Ala-IIID	<b>cnm<sup>5</sup>UGC</b>	Glu-IB	<b>cnm<sup>5</sup>UUC</b>	Gly-IIA	<b>cnm<sup>5</sup>UCC</b>	A
4B2	G	Val-IA	CAC	Ala-IIID	CGC	Glu-IB	<b>ac<sup>4</sup>CUC</b>	Gly-IIA	CCC	G
row	1st	aa-aaRS	Ac	aa-aaRS	Ac	aa-aaRS	Ac	aa-aaRS	Ac	3rd
		1		2		3		4		

Figure 2. The genetic code in Archaea (i.e., *Pyrococcus furiosus*). Genetic code columns (tRNA-35) are labelled 1-4. The leftmost table column gives row designations. Row 1-4 numbers indicate the tRNA-36 base. Codon bases (1<sup>st</sup>, 2<sup>nd</sup>, 3<sup>rd</sup>) are shaded pale yellow. tRNA-34 bases are indicated with modifications in bold type. Amino acids and aaRS structural classes and subclasses are shown (i.e., Phe-IIC indicates tRNA<sup>Phe</sup> is charged by PheRS-IIC) (aa-aaRS). GAA/**AAA** indicates anticodon (Ac) data. Anticodon GAA reads codons UUU and UUC, and anticodon AAA is not utilized. Color highlighting is meant to emphasize particular table features and evolution of aaRS enzymes through Earth's history in Figures 2-6. Data were modeled on *P. furiosus* but tRNA modification data are not complete so some data were inferred or utilized from other Archaea. Color shading is meant to be largely consistent in Figures 2-6.

Furthermore, 1-codon sectors were difficult to evolve and maintain. Consider the Ile/Met 4-codon sector, in which Met occupies a 1-codon (AUG) sector. We posit that the 4-codon Ile/Met sector was originally a 4-codon Ile sector that Met invaded, eliminating the Ile UAU anticodon [35-37]. In Archaea and Bacteria, Ile utilizes a CAU anticodon. In some Archaea, C34 is modified to 2-agmatidine (agm<sup>2</sup>C) to read codon AUA (Ile) but not codon AUG (Met) [4,51-53]. Note that a **cnm<sup>5</sup>UAU** anticodon would read both AUA (Ile) and AUG (Met), causing miscoding. Met utilizes two tRNAs, tRNA<sup>Met</sup> (i.e., CmAU) for elongation and tRNA<sup>iMet</sup> (i.e., unmodified CAU) for initiation. A very similar strategy is utilized to maintain the 1-codon Met box in most or all prokaryotes [25,51,54-57]. The Trp 1-codon sector (UGG) is read by the Trp anticodon CCA that is specific for codon UGG. The UCA anticodon is not utilized, because Trp shares a 2-codon box with a stop codon (UGA) that is recognized by a protein release factor that binds to the mRNA UGA stop codon to terminate translation on the ribosome [58]. Anticodon **cnm<sup>5</sup>UCA** would read codons UGA and UGG, causing miscoding and suppressing translation stops. This explains why Trp utilizes anticodon CCA, rather than **cnm<sup>5</sup>UCA**, to read codon UGG.

GlnRS-IB was a eukaryotic innovation that was transferred from Eukarya to Archaea and Bacteria by horizontal gene transfer [49,59]. Some archaeal and bacterial species, therefore, lack GlnRS-IB and instead use GluRS-IB to convert tRNA<sup>Gln</sup> to Glu-tRNA<sup>Gln</sup>. In these organisms, an amidotransferase converts Glu-tRNA<sup>Gln</sup> to Gln-tRNA<sup>Gln</sup> for translation [60,61]. So, GlnRS-IB in Archaea and Bacteria was a later acquisition in evolution (i.e., perhaps ~1.5 to 2.5 billion years ago). In Archaea, GluRS-IB, LysRS-IE and GlnRS-IB (from Eukarya) are closely related aaRS enzymes [35-37]. In some cases, the historic structural sub classifications for aaRS are deceptive. LysRS-IE is more closely related to GluRS-IB and GlnRS-IB than any of these three aaRS enzymes are to CysRS-IB. Similarly, AspRS-IIIB,

AsnRS-IIB and HisRS-IIA are reasonably closely related aaRS enzymes. We posit that a pre-LUCA AspRS-IIA evolved to AspRS-IIB to suppress tRNA charging errors, before evolution of AsnRS-IIB from AspRS-IIB. These homologies create a striped pattern of aaRS relatedness in column 3, indicative of the mode by which column 3 sectored [35-37]. The striped pattern in Archaea is somewhat disrupted by evolution of LysRS-IIB in Bacteria to replace archaeal LysRS-IE.

## 5. Ancient Bacteria

As a model organism for an ancient Bacterium, we selected *Thermus thermophilus* (Figure 3). Unfortunately, to date, there is too much missing tRNA modification data for *T. thermophilus*, so, perhaps, the analysis we present can be refined in the future. Although data are currently missing, we posit a 5-carbon  $\text{cm}^5\text{U34}$ -based modification to suppress superwobbling and to support the existence of 2-codon genetic code sectors. In column 4, the Arg 4-codon sector may be an intermediate in evolution of the  $\text{A34} \rightarrow \text{I34}$  modification. *T. thermophilus* tRNA<sup>Arg</sup> encodes anticodon ACG and lacks a tRNA with a GCG anticodon. *T. thermophilus*, however, appears to lack the enzyme expected to convert  $\text{A34} \rightarrow \text{I34}$  (tRNA adenosine deaminase). Currently, we do not know whether an unknown modification of A34 is present in *T. thermophilus*. Unmodified Arg (UCG) would read the entire 4-codon box. Modified anticodon  $\text{cm}^5\text{UCG}$  would be expected to read CGA and CGG Arg codons. Anticodon CCG reads the CGG Arg codon. Precisely how *T. thermophilus* reads the Arg 4-codon box, therefore, does not appear to be currently reported. It is possible that *T. thermophilus* represents an intermediate stage in evolution of the Arg (ACG  $\rightarrow$  ICG) anticodon present in most Bacteria [14].

	U	C	A	G	2nd					
1A1	U	Phe-IIC	GAA/ <b>AAA</b>	Ser-IIA	GGA/ <b>AGA</b>	Tyr-IC	<b>QGUA/AUA</b>	Cys-IB	GCA/ <b>ACA</b>	U
1A2	U	Phe-IIC	GAA	Ser-IIA	GGA	Tyr-IC	<b>QGUA</b>	Cys-IB	GCA	C
1B1	U	Leu-IA	<b>cm<sup>5</sup>UAA</b>	Ser-IIA	<b>cm<sup>5</sup>UGA</b>	<b>STOP</b>	<b>UUA</b>	<b>STOP</b>	<b>UCA</b>	A
1B2	U	Leu-IA	CAA	Ser-IIA	CGA	<b>STOP</b>	<b>CUA</b>	Trp-IC	CCA	G
2A1	C	Leu-IA	<b>GAG/AAG</b>	Pro-IIA	<b>GGG/AGG</b>	His-IIA	<b>QGUG/AUG</b>	Arg-ID	<b>ACG/GCG</b>	U
2A2	C	Leu-IA	GAG	Pro-IIA	GGG	His-IIA	<b>QGUG</b>	Arg-ID	ACG	C
2B1	C	Leu-IA	<b>cm<sup>5</sup>UAG</b>	Pro-IIA	<b>cm<sup>5</sup>UGG</b>	Gln	<b>cm<sup>5</sup>UUG</b>	Arg-ID	UCG	A
2B2	C	Leu-IA	CAG	Pro-IIA	CGG	Gln	CUG	Arg-ID	CCG	G
3A1	A	Ile-IA	<b>GAU/AAU</b>	Thr-IIA	<b>GGU/AGU</b>	Asn-IIB	<b>QGUU/AUU</b>	Ser-IIA	<b>GCU/ACU</b>	U
3A2	A	Ile-IA	GAU	Thr-IIA	GGU	Asn-IIB	<b>QGUU</b>	Ser-IIA	GCU	C
3B1	A	Ile-IA	<b>k<sup>2</sup>CAU/UAU</b>	Thr-IIA	<b>cm<sup>5</sup>UGU</b>	Lys-IIIB	<b>cm<sup>5</sup>UUU</b>	Arg-ID	<b>cm<sup>5</sup>UCU</b>	A
3B2	A	Met-IA	<b>m<sup>5</sup>CAU/CAU</b>	Thr-IIA	CGU	Lys-IIIB	CUU	Arg-ID	CCU	G
4A1	G	Val-IA	<b>GAC/AAC</b>	Ala-IID	<b>GGC/AGC</b>	Asp-IIB	<b>QGUC/AUC</b>	Gly-IIA	<b>GCC/ACC</b>	U
4A2	G	Val-IA	GAC	Ala-IID	GGC	Asp-IIB	<b>QGUC</b>	Gly-IIA	GCC	C
4B1	G	Val-IA	<b>cm<sup>5</sup>UAC</b>	Ala-IID	<b>cm<sup>5</sup>UGC</b>	Glu-IB	<b>cm<sup>5</sup>UUC</b>	Gly-IIA	<b>cm<sup>5</sup>UCC</b>	A
4B2	G	Val-IA	CAC	Ala-IID	CGC	Glu-IB	CUC	Gly-IIA	CCC	G
row	1st	aa-aaRS	Ac	aa-aaRS	Ac	aa-aaRS	Ac	aa-aaRS	Ac	3rd
		1		2		3		4		

Figure 3. The genetic code in ancient Bacteria (i.e., *Thermus thermophilus*). GAA/**AAA** indicates anticodon GAA is utilized and AAA is not, to encode Phe. **QGUA/AUA** indicates the  $\text{G34} \rightarrow \text{Q34}$  modification and AUA is not utilized. LysRS-IIB is a bacterial innovation. **cm<sup>5</sup>UAA** for Leu indicates that the precise 5-carbon U modification to suppress superwobbling is not currently reported for *T. thermophilus*. Some tRNA modification data were inferred by identifying enzymes in *T. thermophilus*. It is not clear to us at the time of writing how the Arg 4-codon box is read.

In column 1, the Ile/Met sector is maintained in much the same manner as in Archaea, although, using a slightly different modification. In *T. thermophilus*, tRNA lysidine (34) synthetase (TilS) is present, so it appears *T. thermophilus* utilizes the 2-lysidine Ile ( $\text{k}^2\text{CAU}$ ) modification [25,51,54-56]. The 2-lysidine modification is chemically similar to the 2-

agmatidine modification in Archaea. 2-lysidine is utilized to read Ile codon (AUA) but not Met codon (AUG). The UAU anticodon is not utilized, because cm<sup>2</sup>U34 would read both codons AUA (Ile) and AUG (Met). The elongator tRNA<sup>Met</sup> (CAU) has a lightly modified C34 (i.e., CmAU). As in Archaea, the initiator tRNA<sup>iMet</sup> (CAU) is unmodified.

In column 3, *T. thermophilus* utilizes a type II tRNA<sup>Tyr</sup>, with a longer V-loop (14-nt; the primordial length of the type II V-loop) [62]. *T. thermophilus* TyrRS-IC interacts with the V-loop tip as a determinant in Tyr placement to form Tyr-tRNA<sup>Tyr</sup>. Although the corresponding tRNAs have not been analyzed for modifications, *T. thermophilus* encodes enzymes for queuosine modification of column 3 tRNAs. Bacterial LysRS-IIB replaces archaeal LysRS-IE. LysRS-IIB is derived in evolution from AspRS-IIB, probably by duplication and repurposing of the gene copy [35]. So, even when an aaRS enzyme is replaced by a very different aaRS in evolution (i.e., LysRS-IE (Archaea) → LysRS-IIB (Bacteria)), evolution of the replacement aaRS may arise within the same column (column 3). Replacement of archaeal LysRS-IE with bacterial LysRS-IIB breaks the striped pattern observed for the simpler archaeal genetic code (compare Figures 2 and 3, column 3). We posit that Archaea, which have a simpler genetic code, are older organisms than Bacteria (compare Figures 2 and 3) [39,63]. *T. thermophilus* has a GlyRS-IIA and a ProRS-IIA that lacks an editing active site, similar to GlyRS-IIA and ProRS-IIA in Archaea. Later in bacterial evolution, GlyRS-IID and ProRS-IIA (i.e., with an added editing active site) evolved. More derived Bacteria utilize CmoA and CmoB enzymes to generate the cmo<sup>5</sup>U modification found in 4-codon sectors in columns 1 and 2 of the *E. coli* genetic code (i.e., Val, Ser, Pro, Thr and Ala) (Figure 4). *T. thermophilus* lacks a detectable CmoA or CmoB homolog. Some *Rickettsiales* utilize CmoA and CmoB, but many do not. In mitochondria, unmodified U34 (superwobbling) is utilized to read 4-codon sectors. Also, CmoA and CmoB were probably missing in the bacterial endosymbiont that became the mitochondria.

## 6. Derived Bacteria

Because of available tRNA modification data, our model organism for a more derived Bacterium is generally *Escherichia coli* (Figure 4) [30]. In this regard, we would prefer to also show full information for the nearest relative of the α-proteobacterial species (i.e., *Rickettsiales*) that became the mitochondria, but we cannot identify these data, and, because of horizontal gene transfers, a modern *Rickettsiales* might not be an apt comparison to the mitochondria. We posit that the 5-carbon of U34 is often modified in Bacteria to suppress superwobbling and to maintain 2-codon sectors. tRNA-34 modification data tend to evolve in columns, as might be expected for enzymes that bind the tRNA anticodon to add a modification. Columns represent the central position tRNA-35 of the anticodon.

Interestingly, in columns 1 and 2, the cmo<sup>5</sup>U34 modification is found in tRNAs encoding Val, Ser, Pro, Thr and Ala [25,64,65]. The cmo<sup>5</sup>U34 modification, therefore, is found in 4-codon sectors and was expected to read codons ending in wobble A, G and U but not C. For tRNA<sup>Pro</sup> (cmo<sup>5</sup>UGG), however, this single tRNA<sup>Pro</sup> (cmo<sup>5</sup>UGG) supports viability of *Salmonella*, indicating that cmo<sup>5</sup>U34 anticodons can potentially read the entire Pro 4-codon box. In *Bacillus subtilis*, tRNA<sup>Leu</sup> (UAG), in which U34 appears to be unmodified, may utilize superwobbling [30].

In column 4, tRNA<sup>Arg</sup> (ACG → ICG), encoded A34 is modified to inosine (I34) by deamination [14-16]. Interestingly, tRNA<sup>Arg</sup> (GCG), which is favoured in Archaea, is not utilized. When A34 is converted to I34, the corresponding G34 anticodon is not utilized. Anticodon I34 reads codon wobble bases U, C and A but not G. To read the 4-codon Arg box, tRNA<sup>Arg</sup> (ICG), (mnm<sup>5</sup>UCG) and (CCG) are utilized. tRNA<sup>Arg</sup> (mnm<sup>5</sup>UCG) probably reads codons CGA and CGG. Also, in column 4, GlyRS-IIA may be replaced with GlyRS-IID in some derived Bacteria (i.e., *E. coli*). In α-Proteobacteria, GlyRS-IIA is utilized, as in *T. thermophilus* and Archaea. Not surprisingly, GlyRS-IID is utilized in plant chloroplasts (i.e., from *Cyanobacteria*), although GlyRS-IIA, not GlyRS-IID, is utilized in the plant mitochondria [49].

		U	C	A	G	2nd				
1A1	U	Phe-IIIC	GAA/AAA	Ser-IIA	GGA/AGA	Tyr-IC	QGUU/AUA	Cys-IB	GCA/ACA	U
1A2	U	Phe-IIIC	GAA	Ser-IIA	GGA	Tyr-IC	QGUU	Cys-IB	GCA	C
1B1	U	Leu-IA	cmnm <sup>5</sup> UmAA	Ser-IIA	cmo <sup>5</sup> UGA	STOP	UUA	STOP	UCA	A
1B2	U	Leu-IA	CmAA	Ser-IIA	CGA	STOP	CUA	Trp-IC	CCA	G
2A1	C	Leu-IA	GAG/AAG	Pro-IIA	GGG/AGG	His-IIA	QGUG/AUG	Arg-ID	IACG/GCG	U
2A2	C	Leu-IA	GAG	Pro-IIA	GGG	His-IIA	QGUG	Arg-ID	IACG	C
2B1	C	Leu-IA	UAG	Pro-IIA	cmo <sup>5</sup> UGG	Gln	mnm <sup>5</sup> s <sup>2</sup> UUG	Arg-ID	mnm <sup>5</sup> UCG/IACG	A
2B2	C	Leu-IA	CAG	Pro-IIA	CGG	Gln	CUG	Arg-ID	CCG	G
3A1	A	Ile-IA	GAU/AAU	Thr-IIA	GGU/AGU	Asn-IIB	QGUU/AUU	Ser-IIA	GCU/ACU	U
3A2	A	Ile-IA	GAU	Thr-IIA	GGU	Asn-IIB	QGUU	Ser-IIA	GCU	C
3B1	A	Ile-IA	k <sup>2</sup> CAU/UAU	Thr-IIA	cmo <sup>5</sup> UGU	Lys-IIB	mnm <sup>5</sup> s <sup>2</sup> UUU	Arg-ID	mnm <sup>5</sup> UCU	A
3B2	A	Met-IA	m <sup>5</sup> CAU/CAU	Thr-IIA	CGU	Lys-IIB	CUU	Arg-ID	CCU	G
4A1	G	Val-IA	GAC/AAC	Ala-IID	GGC/AGC	Asp-IIB	gluQGUC/AUC	Gly-IID	GCC/AGC	U
4A2	G	Val-IA	GAC	Ala-IID	GGC	Asp-IIB	gluQGUC	Gly-IID	GCC	C
4B1	G	Val-IA	cmo <sup>5</sup> UAC	Ala-IID	cmo <sup>5</sup> UGC	Glu-IB	mnm <sup>5</sup> s <sup>2</sup> UUC	Gly-IID	mnm <sup>5</sup> UCC	A
4B2	G	Val-IA	CAC	Ala-IID	CGC	Glu-IB	CUC	Gly-IID	CCC	G
		1st aa-aaRS	Ac	aa-aaRS	Ac	aa-aaRS	Ac	aa-aaRS	Ac	3rd
		1		2		3		4		

Figure 4. The genetic code in derived Bacteria (i.e., *Escherichia coli*). Innovations include: 1) ProRS-IIA takes on additional bacterial features; 2) Arg ACG→ICG/GCG is utilized (*T. thermophilus* appears to lack tRNA adenosine deaminase); and 3) GlyRS-IIA can be replaced in some Bacteria by GlyRS-IID. As in *T. thermophilus*, LysRS-IIB and type II tRNA<sup>Tyr</sup> are utilized. This table is based on incomplete tRNA modification data. *E. coli* appears not to utilize Lys anticodon CUU.

In column 1, the Ile/Met 4-codon sector is essentially as described for Archaea and ancient Bacteria. Ile anticodon GAU reads codons AUU and AUC. Ile anticodon k<sup>2</sup>CAU (k<sup>2</sup>C for 2-lysidine modification of C) reads codon AUA (Ile) but not AUG (Met) [25,51,54]. Anticodon UAU is not utilized because even a cm<sup>5</sup>UAU would read both AUA (Ile) and AUG (Met) causing miscoding. Met utilizes tRNA<sup>Met</sup> (m<sup>5</sup>CAU) (elongator Met) and tRNA<sup>iMet</sup> (unmodified CAU) (initiator Met). Maintaining 1-codon sectors presents problems. For instance, in mitochondria, Ile and Met occupy 2-codon sectors to minimize the size of the tRNAome and its supporting proteome [5].

In column 3, queuosine modification for G34 (G34→Q34) is utilized [23-25]. Interestingly, the G34→Q34 column 3 modification is passed forward into the eukaryotic cytosol and also into mitochondria. All G34 anticodons in column 3 are modified G34→Q34. There can be further modification of queuosine to glutamyl-queuosine (tRNA<sup>Asp</sup> (gluQGUC)). As in *T. thermophilus*, tRNA<sup>Tyr</sup> is a type II tRNA with a longer V-loop. As expected, this feature of tRNA<sup>Tyr</sup> goes forward to the mitochondria but not the eukaryotic cytosol. LysRS-IIB is utilized in most Bacteria in place of archaeal LysRS-IE. *E. coli* appears to lack tRNA<sup>Lys</sup> (CUU). Apparently, tRNA<sup>Lys</sup> (mnm<sup>5</sup>s<sup>2</sup>UUU) reads both Lys codons AAA and AAG, as expected.

## 7. Mitochondria

Mitochondria were evolved from an  $\alpha$ -proteobacterial endosymbiont, perhaps, a *Rickettsiales*. The genetic code for human mitochondria is shown in Figure 5 [5]. Because of human health issues, better tRNA modification data are available for human mitochondrial tRNAs than for most Eukarya. Also, human mitochondria utilize only 22-tRNAs, so humans, vertebrates and animals have a significantly reduced mitochondrial tRNAome. We believe the data shown in Figure 5 are essentially complete and accurate.

		U	C	A	G	2nd
1A1	U	Phe-IIC GAA/AAA	Ser-IIA UGA	Tyr-IC QGUA/AUA	Cys-IB GCA/ACA	U
1A2	U	Phe-IIC GAA	Ser-IIA UGA	Tyr-IC QGUA	Cys-IB GCA	C
1B1	U	Leu-IA $\tau m^5$ UAA	Ser-IIA UGA	STOP UUA	Trp-IC $\tau m^5$ UCA	A
1B2	U	Leu-IA $\tau m^5$ UAA/CAA	Ser-IIA UGA	STOP CUA	Trp-IC $\tau m^5$ UCA	G
2A1	C	Leu-IA UAG/AAG	Pro-IIA UGG	His-IIA QGUG/AUG	Arg-ID UCG	U
2A2	C	Leu-IA UAG/GAG	Pro-IIA UGG	His-IIA QGUG	Arg-ID UCG	C
2B1	C	Leu-IA UAG	Pro-IIA UGG	Gln $\tau m^5$ s <sup>2</sup> UUG	Arg-ID UCG	A
2B2	C	Leu-IA UAG/CAG	Pro-IIA UGG	Gln $\tau m^5$ s <sup>2</sup> UUG	Arg-ID UCG	G
3A1	A	Ile-IA GAU/AAU	Thr-IIA UGU	Asn-IIB QGUU/AUU	Ser-IIA Ser-IIA	GCU/ACU
3A2	A	Ile-IA GAU	Thr-IIA UGU	Asn-IIB QGUU	Ser-IIA	GCU
3B1	A	Met-IA $f^5$ CAU/UAU	Thr-IIA UGU	Lys-IIB $\tau m^5$ s <sup>2</sup> UUU	STOP UCA	A
3B2	A	Met-IA $f^5$ CAU	Thr-IIA UGU	Lys-IIB $\tau m^5$ s <sup>2</sup> UUU	STOP UCA	G
4A1	G	Val-IA UAC/AAC	Ala-IID UGC	Asp-IIB QGUC/AUC	Gly-IIA UCC	U
4A2	G	Val-IA UAC/GAC	Ala-IID UGC	Asp-IIB QGUC	Gly-IIA UCC	C
4B1	G	Val-IA UAC	Ala-IID UGC	Glu-IB $\tau m^5$ s <sup>2</sup> UUC	Gly-IIA UCC	A
4B2	G	Val-IA UAC/CAC	Ala-IID UGC	Glu-IB $\tau m^5$ s <sup>2</sup> UUC	Gly-IIA UCC	G
row	1st	aa-aaRS Ac	aa-aaRS Ac	aa-aaRS Ac	aa-aaRS Ac	3rd
		1	2	3	4	

Figure 5. The genetic code in human mitochondria. A major strategy to shrink the mitochondrial tRNAome was superwobbling (beige shading). In mitochondria, Met, Ile and Trp utilize 2-codon sectors. The distribution of stop codons has changed. GlnRS-IB is not imported into human mitochondria. G34→Q34 modifications are utilized in column 3.  $\tau$  indicates taurine modifications. Many unused anticodons have not been struck out in this figure (except in column 1). It appears that the human mitochondrial code may be completely and accurately reported [5].

The main strategy for shrinking the mitochondrial tRNAome is “superwobbling” or 4-way wobbling, in which a single unmodified U34 tRNA reads an entire 4-codon box [1,2,5]. This strategy is used for all 4-codon boxes, including 4-codon boxes encoding Leu, Val, Ser, Pro, Thr, Ala, Arg and Gly (beige shading in Figure 4). In column 3, G34→Q34 modifications are utilized (light green shading in Figure 5). 2-codon boxes with U34 utilize a modified U34, as expected, to restrict superwobbling, which would cause miscoding. Evolution of specific modifications generally aligns in columns, as expected. Human mitochondria include no 1-codon sectors (i.e., to encode Met and Trp) [5]. Instead, atypically, 2-codon sectors are utilized for Ile, Met and Trp. Because a stop codon (UGA) was lost in forming a Trp 2-codon sector, the loss was compensated by converting AGG and AGA, which in Bacteria are Arg codons, into mitochondrial stop codons. Human mitochondria do not import GlnRS-IB. Instead, GluRS-IB is utilized to synthesize Glu-tRNA<sup>Gln</sup>, which is converted to Gln-tRNA<sup>Gln</sup> by an amidotransferase. The bacterial mitochondrial ancestor did not encode GlnRS-IB, which was a eukaryotic innovation transferred to Archaea and Bacteria by horizontal gene transfers [49]. Archaeal *P. furiosis* also lacks GlnRS-IB and uses a similar tRNA<sup>Gln</sup> charging strategy. Mitochondria utilize LysRS-IIB, which was derived initially from a bacterial source. Not all mitochondrial and chloroplast tRNAomes, tRNA modifications and collections of aaRS enzymes are the same, so human mitochondria are an example without complete generality.

## 8. The eukaryotic cytosol

In the eukaryotic cytosol, the genetic code reflects the fusion of an Asgard Archaea and the  $\alpha$ -proteobacterial endosymbiont that became the mitochondria [8-10,66] (Figure 6). A major feature in evolution of the eukaryotic cytosol is the expansion of the A34→I34 strategy (beige shading in Figure 6). All 4-codon sectors except that encoding glycine utilize the A34→I34 modification and, also, suppression of the corresponding G34 anticodon

[14]. We suspect that the Gly 4-codon sector did not adopt the A34→I34 modification strategy because of evolutionary pressures to adjust rates of translation. It appears that the Gly GCC anticodon may have been better balanced with the mn<sup>5</sup>UCC and CCC anticodons. Although *E. coli* does not do this, some Bacteria encode A34 in 4-codon sectors other than Arg (ACG→ICG), but, generally, in these cases, A34 does not appear to be converted to inosine [14,16]. To prevent miscoding, the A34→I34 modification strategy can only occur in 3- (Ile) or 4-codon sectors, because I34 recognizes codon wobble bases U, C and A.

		U	C	A	G	2nd
1A1	U	Phe-IIC	GmAA/AAA	Ser-IIA	IAGA/GGA	Tyr-IC
1A2	U	Phe-IIC	GmAA	Ser-IIA	IAGA	Tyr-IC
1B1	U	Leu-IA	ncm <sup>5</sup> UmAA	Ser-IIA	ncm <sup>5</sup> UGA/IAGA	STOP
1B2	U	Leu-IA	m <sup>5</sup> CAA	Ser-IIA	CGA	STOP
2A1	C	Leu-IA	IAAG/GAG	Pro-IIA	IAGG/GGG	His-IIA
2A2	C	Leu-IA	IAAG	Pro-IIA	IAGG	His-IIA
2B1	C	Leu-IA	UAG/IAAG	Pro-IIA	ncm <sup>5</sup> UGG/IAGC	Gln-IB
2B2	C	Leu-IA	CAG	Pro-IIA	CGG	Gln-IB
3A1	A	Ile-IA	IAAU/GAU	Thr-IIA	IAGU/GGU	Asn-IIB
3A2	A	Ile-IA	IAAU	Thr-IIA	IAGU	Asn-IIB
3B1	A	Ile-IA	ΨUAΨ/IAAU	Thr-IIA	ncm <sup>5</sup> UGU/IAGC	Lys-IIB
3B2	A	Met-IA	CmAU/CAU	Thr-IIA	CGU	Lys-IIB
4A1	G	Val-IA	IAAC/GAC	Ala-IID	IAGC/GGC	Asp-IIB
4A2	G	Val-IA	IAAC	Ala-IID	IAGC	Asp-IIB
4B1	G	Val-IA	ncm <sup>5</sup> UAC/IAAC	Ala-IID	ncm <sup>5</sup> UGC/IAGC	Glu-IB
4B2	G	Val-IA	CAC	Ala-IID	CGC	Glu-IB
row	1st	aa-aaRS	Ac	aa-aaRS	Ac	aa-aaRS
		1	2	3	4	3rd

Figure 6. The genetic code in the eukaryotic cytosol (i.e., human). Shading and symbols are as in Figures 2-5. ΨUAΨ indicates ΨAΨ (Ψ for pseudouridine).

In column 1, the Ile/Met 4-codon sector underwent some eukaryotic cytosol-specific changes. The Ile anticodon AAU→IAU modification is utilized, allowing the reading of Ile codons AUU, AUC and AUA. Also, in Eukaryotes, anticodon UAU→ΨAΨ (Ψ for pseudouridine) can be used to read codon AUA (Ile) but not AUG (Met) [30]. In Prokaryotes, generally, UAU is not utilized even with modification (Figures 2-4). In column 3, G34 is modified to Q34 or a modified Q34 (i.e., galactosyl- or mannosyl-queuosine) [23,24]. Because queuosine in column 3 is a bacterial innovation, the eukaryotic cytosol takes on significant bacterial characteristics in the genetic fusion(s) that resulted in eukaryogenesis. LysRS-IIB is another bacterial innovation that is utilized in the eukaryotic cytosol. Apparently, LysRS-IE, derived from an Asgard archaeal partner in the fusion, was rejected. GlyRS-IIA could be derived from an Asgard Archaea, an  $\alpha$ -Proteobacteria or by horizontal gene transfer from another archaeal or bacterial source.

The eukaryotic cytosol does not utilize the cmo<sup>5</sup>U34 modification found in some Bacteria but not others (columns 1 and 2; compare Figures 4 and 6). Probably, the cmo<sup>5</sup>U34 modification was absent in the bacterial endosymbiont that became the mitochondria. We posit that optimal balanced reading of 4-codon boxes may be tuned by coevolution of tRNA sequences and anticodon modifications. Therefore, the cmo<sup>5</sup>U34 modification may be more compatible paired with synonymous G34 anticodons, as observed in *E. coli* for Val, Ser, Pro, Thr and Ala (Figure 4). By contrast, in Eukarya, the ncm<sup>5</sup>U34 modification may be more compatible paired with synonymous I34 anticodons (Figure 6). This could help explain why Gly utilizes anticodons GCC (rather than ICC, which does not appear to be utilized), ncm<sup>5</sup>UmCC and CCC anticodons in Eukarya (Figure 6). The ncm<sup>5</sup>UmCC Gly anticodon probably is restricted to read Gly codons GGG and GGA.

## 9. Sources of Eukaryotic and mitochondrial aaRS enzymes

Table I reflects work in progress toward understanding how human cytoplasmic and mitochondrial aaRS enzymes may have evolved through the complex genetic fusion(s) that generated Eukarya [49]. The story is tangled because of: 1) (sometimes multiple) horizontal gene transfers; 2) multiple archaeal and bacterial contributions to the eukaryotic genetic make-up; 3) eukaryotic genetic innovations; and 4) coevolution of cytosolic and mitochondrial tRNAs and aaRS enzymes. Generally, cytosolic tRNAs are thought to have archaeal origins and mitochondrial tRNAs probably have an  $\alpha$ -proteobacterial origin. Interestingly, tracing mitochondrial aaRS to  $\alpha$ -proteobacterial origins has been challenging, indicating many diverse bacterial contributions to Eukarya evolution [59,67,68]. In plants, several aaRS enzymes are co-targeted to the mitochondria and the chloroplasts, and chloroplast aaRS, in some cases, appear to have been derived from a cyanobacterial source [68]. Also, there are apparent discrepancies relating to the proteobacterial sourcing of mitochondrial aaRS [59,67,68]. A full and reliable accounting of the sourcing of aaRS enzymes in the eukaryotic cytosols (i.e., animals and plants) and in mitochondria and chloroplast organelles does not appear to yet be available. Also, nearest apparent bacterial relatives of most mitochondrial and chloroplast aaRS have not been unambiguously reported [49].

**Table I.** Human aaRS enzymes (and genes) in the cytosol and mitochondria. PMW indicates *Parvarchaeota*, *Micrarchaeota*, and *Woesearchaeota* [49]. The mitochondria utilizes GluRS-IB to generate Glu-tRNA<sup>Gln</sup> and a transamidase to generate Gln-tRNA<sup>Gln</sup> for translation.

aaRS	Cyto	Cyto/Mito	Mito
AlaRS-IID	AARS (Bacteria)		AARS2 (Bacteria)
ArgRS-ID	RARS (Bacteria)		RARS2 (Bacteria)
AsnRS-IIB	NARS (Archaea)		NARS2 (Bacteria)
AspRS-IIB	DARS (Deinococcus-Thermus; Bacteria)		DARS2 (Bacteria)
CysRS-IB	CARS (Archaea)		CARS2 (Archaea)
GlnRS-IB	QARS (Eukarya)		Transamidation
GluRS-IB	EPRS (PMW; Archaea)		EARS2 (Bacteria)
GlyRS-IIA		GARS (Euryarchaeota; Archaea)	
HisRS-IIA	HARS (Archaea)		HARS2 (Bacteria)
IleRS-IA	IARS (Lentisphaera; Bacteria)		IARS2 (Bacteria)
LeuRS-IA	LARS (PMW; Archaea)		LARS2 (Bacteria)
LysRS-IIB		KARS (Bacteria)	
MetRS-IA	MARS (Archaea)		MARS2 (Bacteria)
PheRS-IIC	FARSA + FARSB (Euryarchaeota?; Archaea)		FARS2 (Bacteria)
ProRS-IIA	EPRS (Archaea)		PARS2 (Bacteria)
SerRS-IIA	SARS (TACK; Archaea)		SARS2 (Bacteria)
ThrRS-IIA	TARS (i.e., Gemmatimonadetes?; Bacteria)		TARS2 (Bacteria)
TrpRS-IC	WARS (PMW; Archaea)		WARS2 (Bacteria)
TyrRS-IC	YARS (Archaea)		YARS2 (Bacteria)
ValRS-IA	VARS (Deltaproteobacteria?; Bacteria)		VARS2 (Bacteria)

Mitochondrial aaRS enzymes are encoded within the eukaryotic cell nucleus. For two aaRS, the gene encoding the cytoplasmic aaRS and the mitochondrial aaRS is the same (GlyRS-IIA (GARS) and LysRS-IIB (KARS)). In most cases, separate genes encoding the cytoplasmic and mitochondrial aaRS are utilized (Table I). Mitochondrial aaRS enzymes are expected to include a mitochondrial targeting sequence. We conclude the following. Many cytosolic eukaryotic aaRS enzymes appear to be bacterial in origin (i.e., 7 cytosolic aaRS enzymes: AlaRS-IID (AARS), ArgRS-ID (RARS), AspRS-IIB (DARS), IleRS-IIA (IARS), LysRS-IIB (KARS), ThrRS-IIA (TARS) and ValRS-IIA (VARS)). In the cases in which there are separate cytoplasmic and mitochondrial aaRS genes, the cytoplasmic aaRS gene is likely to have an archaeal origin and the mitochondrial gene invariably appears to have a bacterial origin (i.e., AsnRS-IIB (NARS and NARS2), GluRS-IB (EPRS and EARS2), HisRS-IIA (HARS and HARS2), LeuRS-IIA (LARS and LARS2), MetRS-IIA (MARS and MARS2); PheRS-IIC $\alpha$  and PheRS-IIC $\beta$  (FARSA, FARSB and FARS2), ProRS-IIA (EPRS and PARS2), SerRS-IIA (SARS and SARS2), TrpRS-IC (WARS and WARS2) and TyrRS-IC (YARS and YARS2)). In human cells, EPRS is a hybrid gene encoding both GluRS-IB and ProRS-IIA. Twelve cytosolic aaRS enzymes appear to have an archaeal origin (i.e., 12 cytosolic aaRS enzymes: AsnRS-IIB (NARS), CysRS-IB (CARS), GluRS-IB (EPRS), GlyRS-IIA (GARS), HisRS-IIA (HARS), LeuRS-IIA (LARS), MetRS-IIA (MARS), PheRS-IIC $\alpha$ / $\beta$  (FARSA and FARSB), ProRS-IIA (EPRS), SerRS-IIA (SARS), TrpRS-IC (WARS) and TyrRS-IC (YARS)). The CARS gene appears to have split into cytosolic CARS and mitochondrial CARS2 by gene duplication and divergence. As noted above, GlnRS-IB is not imported into human mitochondria. In the eukaryotic cytosol, GlnRS-IB appears to be a eukaryotic innovation that was transferred to Bacteria and Archaea by multiple horizontal gene transfers [49,59]. Some cytosolic aaRS genes appear to have undergone multiple horizontal gene transfers. Examples include AlaRS-IID (AARS), AsnRS-IIB (NARS), ArgRS-ID (RARS), CysRS-IB (CARS), HisRS-IIA (HARS), MetRS-IIA (MARS), ProRS-IIA (EPRS) and TyrRS-IC (YARS). Because of complex genetics, horizontal gene transfers and divergent evolution, there may be significant differences comparing eukaryotic cytosols, mitochondria and chloroplasts from very different species. It appears that for the first eukaryotes to have survived may have required multiple and complex horizontal gene transfers and/or multiple endosymbioses.

## 10. tRNA modifications are as old as LUCA

We consider *Pyrococcus furiosis* to be a reasonable reference organism for LUCA. *P. furiosis* includes an *Elp3* homolog that may encode tRNA-U34 cm<sup>5</sup>U methylase that initiates the cnm<sup>5</sup>U34 modification (Figure 2). The *Elp3* enzyme class is as ancient as LUCA. These enzymes utilize S-adenosylmethionine, an iron-sulphur complex, acetyl coenzyme A and radical intermediates to methylate the 5-carbon of U34 [69-71]. The cm<sup>5</sup>U34 reaction appears to include multiple steps and cooperation of the S-adenosylmethionine and the lysine acetyltransferase homology (coenzyme A-binding) active sites. S-adenosylmethionine is converted to a 5' deoxyadenosine radical. Acetyl-CoA is bound in the lysine acetyltransferase homology domain. An acetyl radical may then be formed and attached at the C5 position of U34. In Figure 7, the related *E. coli* enzyme RlmN methylase is shown that modifies the 2-carbon of tRNA-A37 [72,73]. The RlmN images were selected because they better emphasize some properties of these ancient enzymes. The image in Figure 7B is a detail and different orientation than that shown in Figure 7A. The (β-α)<sub>6</sub> partial barrel that binds S-adenosylmethionine was derived from a (β-α)<sub>8</sub> TIM barrel (TIM for triose phosphate isomerase). The partial barrel domain is identified by 6-parallel β-sheets with intervening α-helices in an open barrel shape. These ancient enzymes include a linked lysine acetyltransferase active site. The coenzyme A-binding region of the lysine acetyltransferase homology domain is identified in the image by antiparallel β-sheets (Figure 7A). Because *Elp3* homologs are older than LUCA, TIM barrels, S-adenosylmethionine, Fe<sub>4</sub>-S<sub>4</sub> cages, lysine acetyltransferases, coenzyme A and cm<sup>5</sup>U34-based modifications must be older than LUCA [74,75]. We posit that cm<sup>5</sup>U34-based tRNA modifications, which were

required to form 2-codon genetic code sectors, were required to evolve the genetic code, which must also be older than LUCA. Because modifications of the tRNA-37 position were important or essential to read the tRNA-36 position, we posit that (probably) t<sup>6</sup>A37 and m<sup>1</sup>G37 modifications are likely older than LUCA (see below).

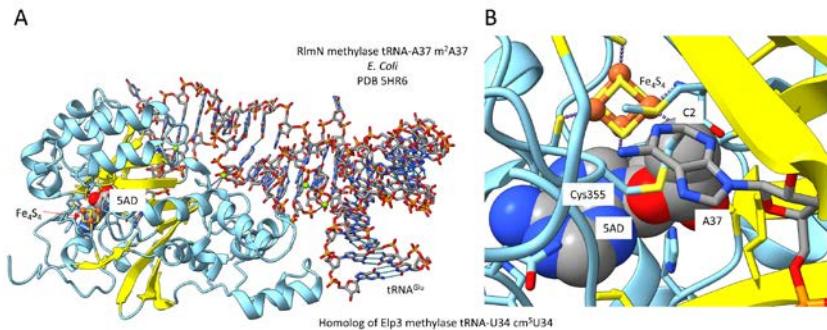


Figure 7. Elp3 (tRNA-cm<sup>5</sup>U34) is an ancient enzyme. The Elp3 homolog RlmN (tRNA-m<sup>2</sup>A37) methylase is shown. A) A view of the RlmN structure. B) A detail and rotated view.  $\beta$ -sheets are yellow. The Fe<sub>4</sub>S<sub>4</sub> cage is indicated. A 5'-deoxyadenosine (5AD) radical is formed from S-adenosylmethionine (space-filling representation). The radical reaction mechanism of RlmN methylase involves a covalent intermediate linking Cys355 and m<sup>2</sup>A37. In Archaea, Elp3 may function somewhat differently. Enzymes of this class include an S-adenosylmethionine methylase domain and a lysine acetyl transferase homology domain that binds acetyl coenzyme A.

To potentially gain insights into tRNA-A37 and -G37 modifications, we visualized the genetic code for Archaea along with reported tRNA-37 modifications (Figure 8). We strongly support the idea that Archaea are the most ancient organisms on Earth and the most similar to LUCA [39,48,63]. Because of missing data, we combined results for tRNA-37 modifications from a number of archaeal species. We conclude the following. At the base of genetic code evolution, the major determinant of tRNA-37 modifications was the identity of the tRNA-36 base. As a result, similar or identical tRNA-37 modifications tend to cluster in genetic code rows (rows 1-4). This result makes sense because tRNA-36 and tRNA-37 are adjacent bases. The most-bulky ancient tRNA-37 modifications (i.e., t<sup>6</sup>A37 and hn<sup>6</sup>A37) are associated with tRNA-U36 (row 3) indicating that U36 may have required stabilization during early code evolution. TRNA-m<sup>1</sup>G37 modifications appear important or essential for reading tRNA-A36 (row 1). Of course, in principle, the identity of tRNA-37 could relate to the reading of the first codon position in mRNA instead of the tRNA-36 position, but we do not favor this idea. It appears to us that mRNA evolution generally chased tRNA evolution and that the genetic code evolved around the tRNA anticodon and the anticodon delimiting base tRNA-37. Also, tRNA-37 modifying enzymes can read the tRNA-36 base directly but not the complementary codon base. All through row 3 (tRNA-U36), tRNA-37 t<sup>6</sup>A, hn<sup>6</sup>A and ms<sup>2</sup>hn<sup>6</sup>A are found. One exception is tRNA<sup>iMet</sup>, for which the anticodon loop is unmodified. From this comparison, it appears to us that tRNA-37 modifications may be most important to support translation elongation rather than to support initiation. Further discrimination of tRNA<sup>Ile</sup> (CAU), tRNA<sup>Met</sup> (CAU) and tRNA<sup>iMet</sup> (CAU) is evident in the acceptor stems of the tRNAs [35].

		U	C	A	G	
		Ser-IIA	GGA(m <sup>1</sup> G/mimG)/AGA	Tyr-IC	GU(A(imG-14)/AUU	Cys-IB
1A1	U	Phe-IIC	GAA(m <sup>1</sup> G/mimG)/AAA	Ser-IIA	GGA(m <sup>1</sup> G/mimG)/AGA	Cys-IB
1A2	U	Phe-IIC	GAA(m <sup>1</sup> G/mimG)	Ser-IIA	GGA(m <sup>1</sup> G/mimG)	Cys-IB
1B1	U	Leu-IA	cnm <sup>5</sup> UAA(m <sup>1</sup> G/imG)	Ser-IIA	cnm <sup>5</sup> UGA(m <sup>1</sup> G)	STOP
1B2	U	Leu-IA	CAA	Ser-IIA	CGA	STOP
2A1	C	Leu-IA	GAG/AAG	Pro-IIA	GGG(m <sup>1</sup> G)/AGG	Arg-ID
2A2	C	Leu-IA	GAG	Pro-IIA	GGG(m <sup>1</sup> G)	Arg-ID
2B1	C	Leu-IA	cnm <sup>5</sup> UAG(A/m <sup>1</sup> G)	Pro-IIA	cnm <sup>5</sup> UGG(m <sup>1</sup> G)	Arg-ID
2B2	C	Leu-IA	CAG	Pro-IIA	CGG	Arg-ID
3A1	A	Ile-IA	GAU(t <sup>6</sup> A/hn <sup>6</sup> A)/AAU	Thr-IIA	GGU(t <sup>6</sup> A/hn <sup>6</sup> A)/AGU	Ser-IIA
3A2	A	Ile-IA	GAU(t <sup>6</sup> A/hn <sup>6</sup> A)	Thr-IIA	GGU(t <sup>6</sup> A/hn <sup>6</sup> A)	Ser-IIA
3B1	A	Ile-IA	agm <sup>2</sup> CAU(t <sup>6</sup> A/hn <sup>6</sup> A)/UAU	Thr-IIA	cnm <sup>5</sup> UGU(hn <sup>6</sup> A)	Arg-ID
3B2	A	Met-IA	CmAU(t <sup>6</sup> A/hn <sup>6</sup> A/ms <sup>2</sup> hn <sup>6</sup> A)/CAUA	Thr-IIA	CGU(t <sup>6</sup> A)	Arg-ID
4A1	G	Val-IA	GACA/AAC	Ala-IID	GGC(A/U)/AGC	Gly-IIA
4A2	G	Val-IA	GACA	Ala-IID	GGC(A/U)	GCC
4B1	G	Val-IA	cnm <sup>5</sup> UACA	Ala-IID	cnm <sup>5</sup> UGCA	Gly-IIA
4B2	G	Val-IA	CACA	Ala-IID	CGCA	Gly-IIA
row 1st	aa-aaRS		Ac	aa-aaRS	Ac	aa-aaRS
			1	2	3	4
						3rd

Figure 8. tRNA-37 modifications in Archaea. The tRNA-34 and tRNA-37 modifications are indicated in bold type. tRNA-37 modifications track the tRNA-36 position (rows 1-4). Row 1 (light blue) and 3 (light green) numbers are shaded for emphasis.

According to tRNA anticodon preference rules, the genetic code evolved around the tRNA anticodon. At the wobble position tRNA-34, G was favored over C/U. At anticodon positions 2 and 3, the preference rules are C>G>U>>A, and preferences are much stronger for the 3<sup>rd</sup> anticodon position (tRNA-36), which, early in code evolution, was a wobble position [35-37]. In keeping with these rules, unmodified tRNA-A37 appears favorable for row 4 (tRNA-C36), and C is the most favored tRNA-36 base (Figure 8). Although data are missing, it appears that tRNA-37 modifications can also be absent for row 2 (tRNA-G36). By contrast, in Archaea, row 3 (tRNA-U36) appears to be the most heavily modified for tRNA-37. We posit that tRNA-t<sup>6</sup>A37 may be among the most ancient row 3 modifications. Notably, t<sup>6</sup>A37 and hn<sup>6</sup>A37 are large N-6 modifications of A37 that may be important for stabilization of tRNA-U36 during translation elongation. Row 1, tRNA-A36, was the last row to fill during evolution of the genetic code. Row 1 is modified for tRNA-37. We posit that tRNA-m<sup>1</sup>G37 may be the most ancient row 1 modification. Because m<sup>1</sup>G37 (row 1) appears to be a smaller modification than t<sup>6</sup>A or hn<sup>6</sup>A37 (row 3), we posit that tRNA-A36 may have been easier to stabilize than tRNA-U36 after suppression of tRNA-36 wobbling (i.e., by EF-Tu, 30S ribosomal closing and tRNA-37 modifications). Also, there is the difference in the identity of the t<sup>6</sup>A37 and m<sup>1</sup>G37 bases. Removing the tRNA-m<sup>1</sup>G37 modification increases the frameshifting of a near-cognate tRNA in the ribosome P-site [28].

Preference rules for the tRNA anticodon may also partially explain why the glycine 4-codon sector did not evolve the A34→I34 modification in Eukaryotes. According to anticodon preference rules, Gly (GCC) is the most favored anticodon in the genetic code [35-37]. This may partly explain why the unmodified GCC anticodon was favored over a modified ICC anticodon for the glycine 4-codon sector in Eukarya. Consideration of anticodon preference rules appears to reinforce our model for evolution of the genetic code, our interpretations of tRNA anticodon loop modifications and our hypothesis that the genetic code evolved around the reading of the tRNA anticodon on the primitive pre-LUCA ribosome.

## 11. Partial redundancy and overlap in translation functions

Because of their ancient evolution and central importance to life, very early, translation systems evolved multiply overlapping, partly redundant and mutually-reinforcing systems. Such redundancy and overlap are observed in: 1) translational fidelity and frame maintenance; 2) tRNA sequence and modification; and 3) aaRS enzyme selectivity in tRNA charging. Because translation systems were central to life and evolution of the genetic code, functional redundancy and, also, backed-up, resilient functions were necessary to evolve stable systems. On the ribosome, translational accuracy and maintenance of the translation frame appear to be partially reinforcing systems. Specifically, translational

accuracy and frame maintenance involve: 1) EF-Tu GTPase; 2) the 16S rRNA “latch” (30S ribosomal subunit closing mechanism); 3) a mRNA bend between the P-site and A-site codons; and 4) modifications of the tRNA-37 base [28,45,46]. EF-Tu is the most important factor in translational accuracy. EF-Tu binds the aminoacylated tRNA (aa-tRNA) and docks it on the ribosome. If the tRNA anticodon-mRNA codon interaction is cognate, EF-Tu hydrolyzes GTP to close the conformation of the ribosome 30S subunit (also referred to as closing the 16S rRNA latch). Once the latch is closed, EF-Tu releases the cognate A-site aa-tRNA to accommodate into the peptidyl transferase center for peptide bond transfer. Accommodation requires a surprisingly large motion of the 3'-end of the aa-tRNA. Figure 9 shows a detail of a catalytic ribosome structure (PDB 5IBB) with the P-site (peptidyl-site) and A-site (aminoacyl-site) tRNAs. To avoid confusion, only the decoding center is shown in the image, not the peptidyl transferase center, and only the anticodon loops of the P-site and A-site tRNAs are shown. The 16S rRNA latch (G530~A1492 and A1493; *Thermus thermophilus* numbering) is shown in its closed conformation. The mRNA bends between the P-site and A-site codons. The bend (or “kink”) orients the 3'-ends of the tRNAs in the peptidyl transferase center, but the bend also separates the P-site and A-site tRNA anticodons in the decoding center [76-78]. Separation of the P-site and A-site anticodons in the decoding center has multiple effects. First, the bend in the mRNA prevents collision of the two anticodon loops. Notably, without the bend, A-site tRNA-37 might collide with the P-site tRNA. Second, separation of the P-site and A-site tRNAs helps the tRNAs to maintain the translation frame by acting as ratchet pawls. Closing the latch maintains the accuracy of translation by confirming the codon-anticodon interaction but also helps to maintain the frame. Modifications at the tRNA-37 position help to delineate the A-site anticodon and to maintain the translation reading frame. Notably, mutations that disable tRNA-37 modifications can cause slippage of the translation frame [28]. Bulky 37 modifications are associated most strongly with U36 (row 3) and A36 (row 1) anticodons, indicating that, among other features, tRNA-37 modifications help to read otherwise less stable codon-anticodon interactions (Figure 8).

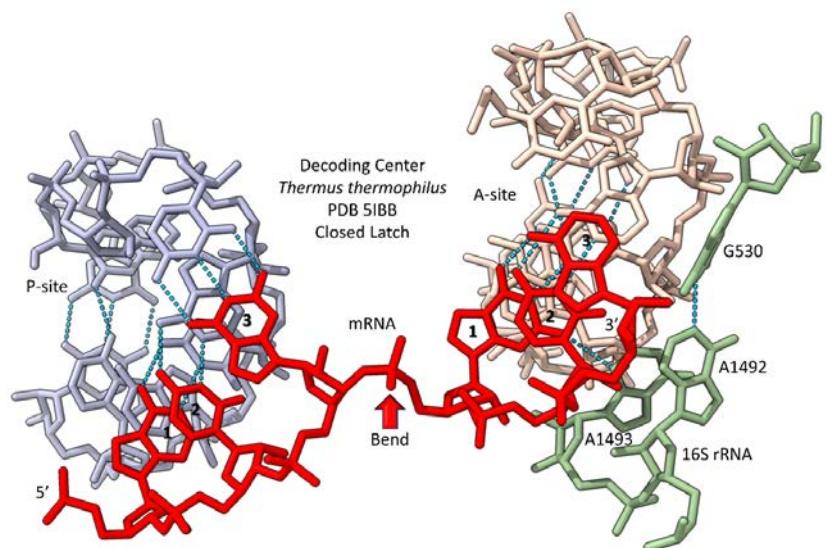


Figure 9. The decoding center of the *T. thermophilus* ribosome during peptide bond synthesis. Colors: grey) P-site tRNA anticodon loop; beige) A-site tRNA anticodon loop; sea green) the “latch”; and red) mRNA. A bend in the mRNA that separates the P-site and A-site codons and anticodons is indicated (red arrow).

The tRNA anticodon loop has a highly specialized sequence with modifications that affect anticodon readout and loop dynamics (Figure 1). Also, the anticodon loop is a target for multiple interactions with modifying enzymes and the cognate aaRS. Thus, any particular sequence or modification can have multiple purposes and interactions. Mutations,

therefore, can have complex and unanticipated effects. The anticodon immediately follows a U-turn following a U, in the 7-nt anticodon loop. The primordial tRNA anticodon loop sequence was close to 32-CU/BNNA-38 (/ indicates a U-turn; B indicates G, C or U (not A); N indicates any base) [35-37,79]. Modifications are common at positions 32, 34, 37 and 38 [18,25]. A weak interaction (i.e., a C~A reverse Hoogsteen pair) is often observed between positions 32 and 38. The C32~A38 interaction may help to preserve the U-turn loop conformation that is important to maintain the codon-anticodon interaction. So, tRNA anticodon loop modifications, sequences and dynamics all are evolved features that affect translational accuracy and output. We consider anticodon loop features to be complex, with overlapping inputs and outputs (i.e., sequences and modifications) that are evolved for different species and for individual tRNAs.

Matching a cognate tRNA to its cognate aaRS is also a problem with multiple inputs [29]. Notably, aaRS enzymes may read: 1) the discriminator base (XCCA-3'; X is the discriminator); 2) the acceptor stem; 3) the anticodon loop; 4) the tRNA elbow (where the D loop and the T loop interact); 5) expanded V-loops in type II tRNAs; and 6) tRNA modifications. We posit that aaRS recognition of their cognate tRNA, therefore, is a product of multiple partially overlapping determinants and anti-determinants. Table I indicates how cognate tRNAs and aaRS enzymes may have been sorted after genetic fusion of multiple Archaea and multiple Bacteria to form Eukarya.

## 12. Conclusions

We strongly support the model that the genetic code evolved around the reading of the tRNA anticodon on the primitive pre-LUCA ribosome [35-37]. Analyses of modifications at the tRNA-34 and -37 anticodon loop positions support this concept. Suppression of wobbling at the tRNA-36 position was essential to evolve the code.

Some of the conclusions of this paper are shown schematically in Figure 10. The presentation in this paper was partly organized around work of others [18,25]. We wished to expand the previous presentations to make it easier for non-experts in tRNA modification and anticodon readout to shape a detailed understanding. We also wanted to emphasize the problem of code evolution and devolution in mitochondria as one that helps explain ancient pre-LUCA evolution and also eukaryogenesis [5]. Figure 10 indicates that, in outline, evolution of life on Earth was simple with a small number of main branches. We advocate for the model that LUCA evolved first to Archaea. Archaea gave rise to Bacteria [39,48,63]. Fusion of an Asgard Archaea and an  $\circ$  -Proteobacteria (i.e., *Rickettsiales*) gave rise to Eukarya, with division and establishment of separate and partly overlapping translation systems for the eukaryotic cytosol and the mitochondria [6,10,43]. Many other archaeal and bacterial genetic inputs were likely during eukaryogenesis, but, at the time of writing, these other gene transfers are somewhat less completely understood (Table I) [49].

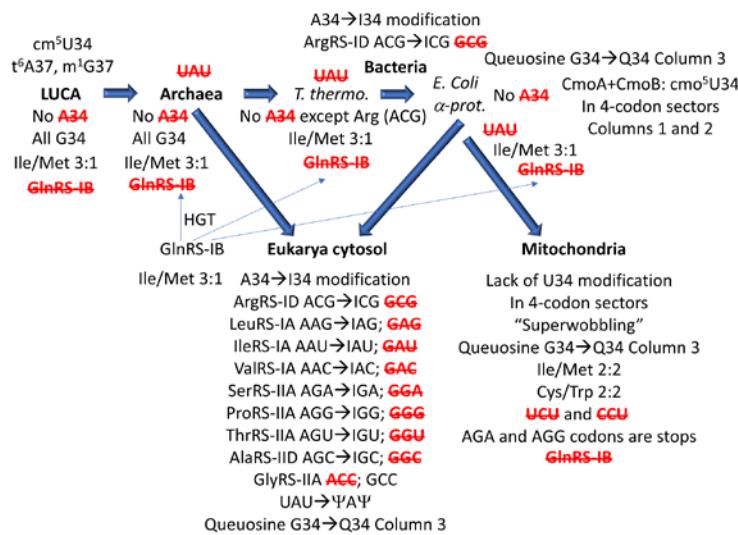


Figure 10. Evolution of tRNA-34 wobble modifications. Superwobbling in mitochondria indicates that cm<sup>5</sup>U34-based modifications were necessary to generate 2-codon sectors to evolve the LUCA code. Red strikethrough indicates that an anticodon is not utilized. Ψ indicates pseudouridine. In mitochondria, 2-codon sectors are utilized to encode Ile, Met and Trp. HGT indicates horizontal gene transfer.

We consider analysis of the evolution of genetic codes and tRNA-34 modifications through Earth's history to support our narrative (Figures 2-6). The simplest genetic code is that of Archaea (Figure 2), indicating that Archaea is closest to LUCA [39,48,63]. Generally, unmodified A34 is not allowed in Archaea, and only G34 is utilized. This fact alone indicates how genetic code degeneracy evolved. Degeneracy evolved through natural processes of the evolution of the reading of the tRNA anticodon on the primitive ribosome. To evolve the genetic code, universal or near universal cm<sup>5</sup>U34-based modifications were necessary to suppress superwobbling (4-way wobbling) and to, thus, support evolution of 2-codon genetic code sectors. Lacking 2-codon sectors, the genetic code would have been limited to a maximum of 16-amino acids.

Translation systems evolved through ancient Bacteria to more derived Bacteria. To date, too much tRNA modification data remains unreported for *Thermus thermophilus*. The missing *T. thermophilus* data will enhance this discussion. More derived Bacteria are genetically diverse with many innovations. In some derived Bacteria, G34 anticodons in 4-codon boxes appear to pair with the cm<sup>5</sup>U34 modification (Val, Ser, Pro, Thr and Ala), unmodified UAG (Leu) and mnm<sup>5</sup>UCC (Gly) (Figure 4). The emergence of the A34 → I34 modification is relevant. The A34 → I34 innovation is associated with suppression of the otherwise preferred G34 anticodon (Figure 6). The A34 → I34 modification expanded in Eukarya. In 3- and 4-codon boxes, I34 anticodons may partner with particular U34 modifications (i.e., ncm<sup>5</sup>U34 and mcm<sup>5</sup>U34, in Eukarya). The G34 → Q34 (Q for queuosine) modification in genetic code column 3 arose in Bacteria and was transmitted to the eukaryotic cytosol and to mitochondria.

Tracing the evolution of the Ile/Met 4-codon sector through evolution is instructive. Maintaining 1-codon sectors for Met and Trp in the genetic code required proteome support. Probably, for this reason, mitochondria abandoned 1-codon sectors (Figure 5) to simplify the tRNAome and its supporting proteome [5]. In prokaryotes, we posit that Met invaded a 4-codon Ile sector during genetic code evolution, suppressing use of the UAU anticodon and resulting in C34 modifications to read Ile (i.e., CAU → agm<sup>2</sup>CAU and k<sup>2</sup>CAU). The 2-agmatidine modification of C34 found in Archaea and the related 2-lysidine modification in Bacteria read codon AUA (Ile) but not codon AUG (Met). In Eukarya, the Ile anticodon modification (UAU → ΨAU) arose, rescuing Ile anticodon UAU.

We posit that 4-codon sectors of the genetic code were balanced using different evolved strategies in different organisms to utilize, generally, 3-tRNAs to read 4-codons. This balance was mostly achieved by adjusting use of G34 or A34-derived and U34 anticodons. In Archaea, G34 and  $\text{cm}^5\text{U34}$ -based anticodons (i.e.,  $\text{cnm}^5\text{U34}$ ) were utilized (Figure 2). In some derived Bacteria, G34 and  $\text{cm}^5\text{U34}$  anticodons were partnered for columns 1 and 2 of the code (4-codon sectors). In column 4, anticodon ICG partners with  $\text{mn}^5\text{UCG}$  to encode Arg, and GCC partners with  $\text{mn}^5\text{UCC}$  to encode Gly (Figure 4). According to anticodon preference rules, Gly (GCC) is expected to be the most favoured anticodon in the genetic code. Gly (GCC) is associated with unmodified tRNA-A37 in Archaea (Figure 8), possibly reflecting the preferred anticodon GCC status. In Eukarya, diverse strategies were evolved for balancing 3- and 4-codon sectors (Figure 6). Very clearly, anticodons that are not utilized in organisms are very important for maintaining balanced reading of tRNAs (Figures 2-6). In mitochondria, 4-codon sectors utilize unmodified U34 to read an entire 4-codon box, indicating that small mitochondrial genome size was more important than optimization of balancing multiple tRNAs for the most rapid and efficient reading of the 4-codon sectors (Figure 5).

We posit that the genetic code evolved around the reading of the tRNA anticodon on the primitive pre-LUCA ribosome. Analysis of tRNA wobble modifications strongly supports the idea that the genetic code evolved around the reading of the anticodon wobble position. Code degeneracy arose from wobbling at the 34 and 36 positions, as previously described [35-37]. Wobbling limits coding to pyrimidine-purine discrimination, so, only 2-assignments were possible at a tRNA wobble position, and to evolve 1-codon sectors posed difficulties with miscoding and anticodon ambiguity. tRNA-37 modifications evolved to help lock down the anticodon 36 position, in part, to suppress wobbling at position 36. Also, wobbling at tRNA-36 was suppressed by evolution of EF-Tu and the 16S rRNA latch (Figures 8 and 9). Analysis of how the genetic code devolved in evolution of the mitochondria strongly supports these views. We do not find the concept of late wobbling evolution to be credible [13,80]. We posit that the genetic code evolved and sectorized largely around the reading of tRNA wobble positions.

Column 3 of the genetic code is split entirely into 2-codon sectors. We have posited that initially column 3 was divided into alternating 2-codon Asp and Glu sectors [35-37]. Our model explains the striped pattern of related aaRS enzymes in Archaea column 3 (Figure 2). According to our model for code evolution, tRNA-U34 modification (i.e.,  $\text{cm}^5\text{U34}$ ) may have been necessary to suppress superwobbling at tRNA-U34 and to achieve the 8-amino acid fractionation of the code. According to our model, therefore,  $\text{cm}^5\text{U34}$ -based modifications may have been necessary to achieve a genetic code including 8-amino acids. Alternatively, only tRNAs with 34-GU-35 (Asp) and 34-CU-35 (Glu) may have initially been utilized. In this case, C34 may have required modification to read mRNA wobble 3A. We conclude that tRNA wobble modifications appear to have been necessary as early as at the 8-amino acid stage of genetic code evolution.

The model we support for evolution of life on Earth is a fairly well-accepted model (Figure 10). The analysis we present, therefore, appears to be straightforward and reasonable. Our work with the initial evolution of the genetic code is also very consistent with our current analysis [35-37]. As noted, the analyses that we present will be enhanced by the acquisition of additional tRNA modification data.

We imagine eukaryogenesis proceeding through a tense evolutionary bottleneck from FECA to LECA (first to last eukaryotic common ancestors). It appears to us that eukaryogenesis was tortured, involving many endosymbiotic and other large horizontal gene transfer events, only some of which resulted in identified eukaryotic organelles. Apparently, contributions were made to the process by many archaeal and many bacterial genes and, also, the genetic fusions were balanced by many compensating eukaryotic innovations. The FECA to LECA bottleneck is reflected in the evolution of aaRS enzymes through eukaryogenesis (Table I) [49]. Clearly, genes were transferred between many

different organisms, including the horizontal transfer of the gene encoding GlnRS-IB from Eukarya to Archaea and Bacteria.

### 13. Future work

Specific goals for future work include: 1) obtain additional tRNA modification data (i.e., for *Pyrococcus furiosus* and *Thermus thermophilus*); 2) Improve the data underlying Table I (obtain optimal aaRS enzyme evolutionary sourcing for: 1) animals; 2) plants; 3) mitochondria; and 4) chloroplasts); 3) improve the description of evolution of tRNA-34 modifications and modification enzymes; and 4) improve the description of evolution of tRNA-37 modifications and modification enzymes. These additional data would enhance the narrative presented here.

**Funding:** This research received no external funding.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Alkatib, S.; Scharff, L.B.; Rogalski, M.; Fleischmann, T.T.; Matthes, A.; Seeger, S.; Schottler, M.A.; Ruf, S.; Bock, R. The contributions of wobbling and superwobbling to the reading of the genetic code. *PLoS Genet* **2012**, *8*, e1003076, doi:10.1371/journal.pgen.1003076.
2. Rogalski, M.; Karcher, D.; Bock, R. Superwobbling facilitates translation with reduced tRNA sets. *Nat Struct Mol Biol* **2008**, *15*, 192-198, doi:10.1038/nsmb.1370.
3. Wolff, P.; Villette, C.; Zumsteg, J.; Heintz, D.; Antoine, L.; Chane-Woon-Ming, B.; Droogmans, L.; Grosjean, H.; Westhof, E. Comparative patterns of modified nucleotides in individual tRNA species from a mesophilic and two thermophilic archaea. *RNA* **2020**, *26*, 1957-1975, doi:10.1261/rna.077537.120.
4. Phillips, G.; de Crecy-Lagard, V. Biosynthesis and function of tRNA modifications in Archaea. *Curr Opin Microbiol* **2011**, *14*, 335-341, doi:10.1016/j.mib.2011.03.001.
5. Suzuki, T.; Yashiro, Y.; Kikuchi, I.; Ishigami, Y.; Saito, H.; Matsuzawa, I.; Okada, S.; Mito, M.; Iwasaki, S.; Ma, D., et al. Complete chemical structures of human mitochondrial tRNAs. *Nat Commun* **2020**, *11*, 4269, doi:10.1038/s41467-020-18068-6.
6. Eme, L.; Spang, A.; Lombard, J.; Stairs, C.W.; Ettema, T.J.G. Archaea and the origin of eukaryotes. *Nat Rev Microbiol* **2017**, *15*, 711-723, doi:10.1038/nrmicro.2017.133.
7. Roger, A.J.; Munoz-Gomez, S.A.; Kamikawa, R. The Origin and Diversification of Mitochondria. *Curr Biol* **2017**, *27*, R1177-R1192, doi:10.1016/j.cub.2017.09.015.
8. Eme, L.; Ettema, T.J.G. The eukaryotic ancestor shapes up. *Nature* **2018**, *562*, 352-353, doi:10.1038/d41586-018-06868-2.
9. Spang, A.; Eme, L.; Saw, J.H.; Caceres, E.F.; Zaremba-Niedzwiedzka, K.; Lombard, J.; Guy, L.; Ettema, T.J.G. Asgard archaea are the closest prokaryotic relatives of eukaryotes. *PLoS Genet* **2018**, *14*, e1007080, doi:10.1371/journal.pgen.1007080.
10. Martijn, J.; Vosseberg, J.; Guy, L.; Offre, P.; Ettema, T.J.G. Deep mitochondrial origin outside the sampled alphaproteobacteria. *Nature* **2018**, *557*, 101-105, doi:10.1038/s41586-018-0059-5.
11. Yu, N.; Jora, M.; Solivio, B.; Thakur, P.; Acevedo-Rocha, C.G.; Randau, L.; de Crecy-Lagard, V.; Addepalli, B.; Limbach, P.A. tRNA Modification Profiles and Codon-Decoding Strategies in *Methanocaldococcus jannaschii*. *J Bacteriol* **2019**, *201*, doi:10.1128/JB.00690-18.

12. Grosjean, H.; Gaspin, C.; Marck, C.; Decatur, W.A.; de Crecy-Lagard, V. RNomics and Modomics in the halophilic archaea *Haloferax volcanii*: identification of RNA modification genes. *BMC Genomics* **2008**, *9*, 470, doi:10.1186/1471-2164-9-470.

13. Yarus, M. Crick Wobble and Superwobble in Standard Genetic Code Evolution. *J Mol Evol* **2021**, *89*, 50-61, doi:10.1007/s00239-020-09985-7.

14. Srinivasan, S.; Torres, A.G.; Ribas de Pouplana, L. Inosine in Biology and Disease. *Genes (Basel)* **2021**, *12*, doi:10.3390/genes12040600.

15. Rafels-Ybern, A.; Torres, A.G.; Camacho, N.; Herencia-Ropero, A.; Roura Frigole, H.; Wulff, T.F.; Raboteg, M.; Bordons, A.; Grau-Bove, X.; Ruiz-Trillo, I., et al. The Expansion of Inosine at the Wobble Position of tRNAs, and Its Role in the Evolution of Proteomes. *Mol Biol Evol* **2019**, *36*, 650-662, doi:10.1093/molbev/msy245.

16. Rafels-Ybern, A.; Torres, A.G.; Grau-Bove, X.; Ruiz-Trillo, I.; Ribas de Pouplana, L. Codon adaptation to tRNAs with Inosine modification at position 34 is widespread among Eukaryotes and present in two Bacterial phyla. *RNA Biol* **2018**, *15*, 500-507, doi:10.1080/15476286.2017.1358348.

17. Torres, A.G.; Pineyro, D.; Filonava, L.; Stracker, T.H.; Batlle, E.; Ribas de Pouplana, L. A-to-I editing on tRNAs: biochemical, biological and evolutionary implications. *FEBS Lett* **2014**, *588*, 4279-4286, doi:10.1016/j.febslet.2014.09.025.

18. El Yacoubi, B.; Bailly, M.; de Crecy-Lagard, V. Biosynthesis and function of posttranscriptional modifications of transfer RNAs. *Annu Rev Genet* **2012**, *46*, 69-95, doi:10.1146/annurev-genet-110711-155641.

19. Pernod, K.; Schaeffer, L.; Chicher, J.; Hok, E.; Rick, C.; Geslain, R.; Eriani, G.; Westhof, E.; Ryckelynck, M.; Martin, F. The nature of the purine at position 34 in tRNAs of 4-codon boxes is correlated with nucleotides at positions 32 and 38 to maintain decoding fidelity. *Nucleic Acids Res* **2020**, *48*, 6170-6183, doi:10.1093/nar/gkaa221.

20. Vik, E.S.; Nawaz, M.S.; Strom Andersen, P.; Fladeby, C.; Bjoras, M.; Dalhus, B.; Alseth, I. Endonuclease V cleaves at inosines in RNA. *Nat Commun* **2013**, *4*, 2271, doi:10.1038/ncomms3271.

21. Muller, M.; Legrand, C.; Tuorto, F.; Kelly, V.P.; Atlasi, Y.; Lyko, F.; Ehrenhofer-Murray, A.E. Queuine links translational control in eukaryotes to a micronutrient from bacteria. *Nucleic Acids Res* **2019**, *47*, 3711-3727, doi:10.1093/nar/gkz063.

22. Tuorto, F.; Legrand, C.; Cirzi, C.; Federico, G.; Liebers, R.; Muller, M.; Ehrenhofer-Murray, A.E.; Dittmar, G.; Grone, H.J.; Lyko, F. Queuosine-modified tRNAs confer nutritional control of protein translation. *EMBO J* **2018**, *37*, doi:10.15252/embj.201899777.

23. Tuorto, F.; Lyko, F. Genome recoding by tRNA modifications. *Open Biol* **2016**, *6*, doi:10.1098/rsob.160287.

24. Vinayak, M.; Pathak, C. Queuosine modification of tRNA: its divergent role in cellular machinery. *Biosci Rep* **2009**, *30*, 135-148, doi:10.1042/BSR20090057.

25. Agris, P.F.; Eruysal, E.R.; Narendran, A.; Vare, V.Y.P.; Vangaveti, S.; Ranganathan, S.V. Celebrating wobble decoding: Half a century and still much is new. *RNA Biol* **2018**, *15*, 537-553, doi:10.1080/15476286.2017.1356562.

26. Bjork, G.R.; Durand, J.M.; Hagervall, T.G.; Leipuviene, R.; Lundgren, H.K.; Nilsson, K.; Chen, P.; Qian, Q.; Urbonavicius, J. Transfer RNA modification: influence on translational frameshifting and metabolism. *FEBS Lett* **1999**, *452*, 47-51, doi:10.1016/s0014-5793(99)00528-1.

27. Bjork, G.R.; Wikstrom, P.M.; Bystrom, A.S. Prevention of translational frameshifting by the modified nucleoside 1-methylguanosine. *Science* **1989**, *244*, 986-989, doi:10.1126/science.2471265.

28. Hoffer, E.D.; Hong, S.; Sunita, S.; Maehigashi, T.; Gonzalez, R.L.J.; Whitford, P.C.; Dunham, C.M. Structural insights into mRNA reading frame regulation by tRNA modification and slippery codon-anticodon pairing. *Elife* **2020**, *9*, doi:10.7554/eLife.51898.

29. Perona, J.J.; Gruic-Sovulj, I. Synthetic and editing mechanisms of aminoacyl-tRNA synthetases. *Top Curr Chem* **2014**, *344*, 1-41, doi:10.1007/128\_2013\_456.

30. Sajek, M.P.; Wozniak, T.; Sprinzl, M.; Jaruzelska, J.; Barciszewski, J. T-psi-C: user friendly database of tRNA sequences and structures. *Nucleic Acids Res* **2020**, *48*, D256-D260, doi:10.1093/nar/gkz922.

31. Boccaletto, P.; Machnicka, M.A.; Purta, E.; Piatkowski, P.; Baginski, B.; Wirecki, T.K.; de Crecy-Lagard, V.; Ross, R.; Limbach, P.A.; Kotter, A., et al. MODOMICS: a database of RNA modification pathways. 2017 update. *Nucleic Acids Res* **2018**, *46*, D303-D307, doi:10.1093/nar/gkx1030.

32. Machnicka, M.A.; Milanowska, K.; Osman Oglou, O.; Purta, E.; Kurkowska, M.; Olchowik, A.; Januszewski, W.; Kalinowski, S.; Dunin-Horkawicz, S.; Rother, K.M., et al. MODOMICS: a database of RNA modification pathways--2013 update. *Nucleic Acids Res* **2013**, *41*, D262-267, doi:10.1093/nar/gks1007.

33. Czerwoniec, A.; Dunin-Horkawicz, S.; Purta, E.; Kaminska, K.H.; Kasprzak, J.M.; Bujnicki, J.M.; Grosjean, H.; Rother, K. MODOMICS: a database of RNA modification pathways. 2008 update. *Nucleic Acids Res* **2009**, *37*, D118-121, doi:10.1093/nar/gkn710.

34. Dunin-Horkawicz, S.; Czerwoniec, A.; Gajda, M.J.; Feder, M.; Grosjean, H.; Bujnicki, J.M. MODOMICS: a database of RNA modification pathways. *Nucleic Acids Res* **2006**, *34*, D145-149, doi:10.1093/nar/gkj084.

35. Lei, L.; Burton, Z.F. Evolution of the genetic code. *Transcription* **2021**, *12*, 28-53, doi:10.1080/21541264.2021.1927652.

36. Lei, L.; Burton, Z.F. Evolution of Life on Earth: tRNA, Aminoacyl-tRNA Synthetases and the Genetic Code. *Life (Basel)* **2020**, *10*, doi:10.3390/life10030021.

37. Kim, Y.; Opron, K.; Burton, Z.F. A tRNA- and Anticodon-Centric View of the Evolution of Aminoacyl-tRNA Synthetases, tRNAomes, and the Genetic Code. *Life (Basel)* **2019**, *9*, doi:10.3390/life9020037.

38. Pak, D.; Du, N.; Kim, Y.; Sun, Y.; Burton, Z.F. Rooted tRNAomes and evolution of the genetic code. *Transcription* **2018**, *9*, 137-151, doi:10.1080/21541264.2018.1429837.

39. Long, X.; Xue, H.; Wong, J.T. Descent of Bacteria and Eukarya From an Archaeal Root of Life. *Evol Bioinform Online* **2020**, *16*, 1176934320908267, doi:10.1177/1176934320908267.

40. Youle, R.J. Mitochondria-Striking a balance between host and endosymbiont. *Science* **2019**, *365*, doi:10.1126/science.aaw9855.

41. Lopez-Garcia, P.; Eme, L.; Moreira, D. Symbiosis in eukaryotic evolution. *J Theor Biol* **2017**, *434*, 20-33, doi:10.1016/j.jtbi.2017.02.031.

42. Kaiser, F.; Krautwurst, S.; Salentin, S.; Haupt, V.J.; Leberecht, C.; Bittrich, S.; Labudde, D.; Schroeder, M. The structural basis of the genetic code: amino acid recognition by aminoacyl-tRNA synthetases. *Sci Rep* **2020**, *10*, 12647, doi:10.1038/s41598-020-69100-0.

43. Zaremba-Niedzwiedzka, K.; Caceres, E.F.; Saw, J.H.; Backstrom, D.; Juzokaite, L.; Vancaester, E.; Seitz, K.W.; Anantharaman, K.; Starnawski, P.; Kjeldsen, K.U., et al. Asgard archaea illuminate the origin of eukaryotic cellular complexity. *Nature* **2017**, *541*, 353-358, doi:10.1038/nature21031.

44. Shi, H.; Moore, P.B. The crystal structure of yeast phenylalanine tRNA at 1.93 Å resolution: a classic structure revisited. *RNA* **2000**, *6*, 1091-1105, doi:10.1017/s1355838200000364.

45. Loveland, A.B.; Demo, G.; Grigorieff, N.; Korostelev, A.A. Ensemble cryo-EM elucidates the mechanism of translation fidelity. *Nature* **2017**, *546*, 113-117, doi:10.1038/nature22397.

46. Loveland, A.B.; Demo, G.; Korostelev, A.A. Cryo-EM of elongating ribosome with EF-Tu\*GTP elucidates tRNA proofreading. *Nature* **2020**, *584*, 640-645, doi:10.1038/s41586-020-2447-x.

47. Quigley, G.J.; Suddath, F.L.; McPherson, A.; Kim, J.J.; Sneden, D.; Rich, A. The molecular structure of yeast phenylalanine transfer RNA in monoclinic crystals. *Proc Natl Acad Sci U S A* **1974**, *71*, 2146-2150, doi:10.1073/pnas.71.5.2146.

48. Lei, L.; Burton, Z.F. Early Evolution of Transcription Systems and Divergence of Archaea and Bacteria *Frontiers in Molecular Biosciences* **2021**, *8*, doi:10.3389/fmolb.2021.651134.

49. Furukawa, R.; Nakagawa, M.; Kuroyanagi, T.; Yokobori, S.I.; Yamagishi, A. Quest for Ancestors of Eukaryal Cells Based on Phylogenetic Analyses of Aminoacyl-tRNA Synthetases. *J Mol Evol* **2017**, *84*, 51-66, doi:10.1007/s00239-016-9768-2.

50. Agris, P.F.; Narendran, A.; Sarachan, K.; Vare, V.Y.P.; Eruysal, E. The Importance of Being Modified: The Role of RNA Modifications in Translational Fidelity. *Enzymes* **2017**, *41*, 1-50, doi:10.1016/bs.enz.2017.03.005.

51. Suzuki, T.; Numata, T. Convergent evolution of AUA decoding in bacteria and archaea. *RNA Biol* **2014**, *11*, 1586-1596, doi:10.4161/15476286.2014.992281.

52. Satpati, P.; Bauer, P.; Aqvist, J. Energetic tuning by tRNA modifications ensures correct decoding of isoleucine and methionine on the ribosome. *Chemistry* **2014**, *20*, 10271-10275, doi:10.1002/chem.201404016.

53. Mandal, D.; Kohrer, C.; Su, D.; Russell, S.P.; Krivos, K.; Castleberry, C.M.; Blum, P.; Limbach, P.A.; Soll, D.; RajBhandary, U.L. Agmatidine, a modified cytidine in the anticodon of archaeal tRNA(Ile), base pairs with adenosine but not with guanosine. *Proc Natl Acad Sci U S A* **2010**, *107*, 2872-2877, doi:10.1073/pnas.0914869107.

54. Sonawane, K.D.; Sambhare, S.B. The influence of hypermodified nucleosides lysidine and t(6)A to recognize the AUA codon instead of AUG: a molecular dynamics simulation study. *Integr Biol (Camb)* **2015**, *7*, 1387-1395, doi:10.1039/c5ib00058k.

55. Nakanishi, K.; Bonnefond, L.; Kimura, S.; Suzuki, T.; Ishitani, R.; Nureki, O. Structural basis for translational fidelity ensured by transfer RNA lysidine synthetase. *Nature* **2009**, *461*, 1144-1148, doi:10.1038/nature08474.

56. Grosjean, H.; Bjork, G.R. Enzymatic conversion of cytidine to lysidine in anticodon of bacterial isoleucyl-tRNA--an alternative way of RNA editing. *Trends Biochem Sci* **2004**, *29*, 165-168, doi:10.1016/j.tibs.2004.02.009.

57. Soma, A.; Ikeuchi, Y.; Kanemasa, S.; Kobayashi, K.; Ogasawara, N.; Ote, T.; Kato, J.; Watanabe, K.; Sekine, Y.; Suzuki, T. An RNA-modifying enzyme that governs both the codon and amino acid specificities of isoleucine tRNA. *Mol Cell* **2003**, *12*, 689-698, doi:10.1016/s1097-2765(03)00346-0.

58. Burroughs, A.M.; Aravind, L. The Origin and Evolution of Release Factors: Implications for Translation Termination, Ribosome Rescue, and Quality Control Pathways. *Int J Mol Sci* **2019**, *20*, doi:10.3390/ijms20081981.

59. Brindefalk, B.; Viklund, J.; Larsson, D.; Thollesson, M.; Andersson, S.G. Origin and evolution of the mitochondrial aminoacyl-tRNA synthetases. *Mol Biol Evol* **2007**, *24*, 743-756, doi:10.1093/molbev/msl202.

60. Bhaskaran, H.; Perona, J.J. Two-step aminoacylation of tRNA without channeling in Archaea. *J Mol Biol* **2011**, *411*, 854-869, doi:10.1016/j.jmb.2011.06.039.

61. Perona, J.J. Two-step pathway to aminoacylated tRNA. *Structure* **2005**, *13*, 1397-1398, doi:10.1016/j.str.2005.09.003.

62. Yaremcuk, A.; Krikliwyi, I.; Tukalo, M.; Cusack, S. Class I tyrosyl-tRNA synthetase has a class II mode of cognate tRNA recognition. *EMBO J* **2002**, *21*, 3829-3840, doi:10.1093/emboj/cdf373.

63. Kim, K.M.; Caetano-Anolles, G. The evolutionary history of protein fold families and proteomes confirms that the archaeal ancestor is more ancient than the ancestors of other superkingdoms. *BMC Evol Biol* **2012**, *12*, 13, doi:10.1186/1471-2148-12-13.

64. Nasvall, S.J.; Chen, P.; Bjork, G.R. The wobble hypothesis revisited: uridine-5-oxyacetic acid is critical for reading of G-ending codons. *RNA* **2007**, *13*, 2151-2164, doi:10.1261/rna.731007.

65. Nasvall, S.J.; Chen, P.; Bjork, G.R. The modified wobble nucleoside uridine-5-oxyacetic acid in tRNAPro(cmo5UGG) promotes reading of all four proline codons in vivo. *RNA* **2004**, *10*, 1662-1673, doi:10.1261/rna.7106404.

66. Eme, L.; Spang, A.; Lombard, J.; Stairs, C.W.; Ettema, T.J.G. Archaea and the origin of eukaryotes. *Nat Rev Microbiol* **2018**, *16*, 120, doi:10.1038/nrmicro.2017.154.

67. Brindefalk, B.; Ettema, T.J.; Viklund, J.; Thollesson, M.; Andersson, S.G. A phylometagenomic exploration of oceanic alphaproteobacteria reveals mitochondrial relatives unrelated to the SAR11 clade. *PLoS One* **2011**, *6*, e24457, doi:10.1371/journal.pone.0024457.

68. Brandao, M.M.; Silva-Filho, M.C. Evolutionary history of *Arabidopsis thaliana* aminoacyl-tRNA synthetase dual-targeted proteins. *Mol Biol Evol* **2011**, *28*, 79-85, doi:10.1093/molbev/msq176.

69. Abbassi, N.E.; Biela, A.; Glatt, S.; Lin, T.Y. How Elongator Acetylates tRNA Bases. *Int J Mol Sci* **2020**, *21*, doi:10.3390/ijms21218209.

70. Lin, T.Y.; Abbassi, N.E.H.; Zakrzewski, K.; Chramiec-Glabik, A.; Jemiola-Rzeminska, M.; Rozycki, J.; Glatt, S. The Elongator subunit Elp3 is a non-canonical tRNA acetyltransferase. *Nat Commun* **2019**, *10*, 625, doi:10.1038/s41467-019-08579-2.

71. Glatt, S.; Zabel, R.; Kolaj-Robin, O.; Onuma, O.F.; Baudin, F.; Graziadei, A.; Taverniti, V.; Lin, T.Y.; Baymann, F.; Seraphin, B., et al. Structural basis for tRNA modification by Elp3 from *Dehalococcoides mccartyi*. *Nat Struct Mol Biol* **2016**, *23*, 794-802, doi:10.1038/nsmb.3265.

72. Schwalm, E.L.; Grove, T.L.; Booker, S.J.; Boal, A.K. Crystallographic capture of a radical S-adenosylmethionine enzyme in the act of modifying tRNA. *Science* **2016**, *352*, 309-312, doi:10.1126/science.aad5367.

73. Blue, T.C.; Davis, K.M. Computational Approaches: An Underutilized Tool in the Quest to Elucidate Radical SAM Dynamics. *Molecules* **2021**, *26*, doi:10.3390/molecules26092590.

74. Martin, W.F.; Weiss, M.C.; Neukirchen, S.; Nelson-Sathi, S.; Sousa, F.L. Physiology, phylogeny, and LUCA. *Microb Cell* **2016**, *3*, 582-587, doi:10.15698/mic2016.12.545.

75. Weiss, M.C.; Sousa, F.L.; Mrrijvac, N.; Neukirchen, S.; Roettger, M.; Nelson-Sathi, S.; Martin, W.F. The physiology and habitat of the last universal common ancestor. *Nat Microbiol* **2016**, *1*, 16116, doi:10.1038/nmicrobiol.2016.116.

76. Keedy, H.E.; Thomas, E.N.; Zaher, H.S. Decoding on the ribosome depends on the structure of the mRNA phosphodiester backbone. *Proc Natl Acad Sci U S A* **2018**, *115*, E6731-E6740, doi:10.1073/pnas.1721431115.

77. Demeshkina, N.; Jenner, L.; Westhof, E.; Yusupov, M.; Yusupova, G. A new understanding of the decoding principle on the ribosome. *Nature* **2012**, *484*, 256-259, doi:10.1038/nature10913.

78. Selmer, M.; Dunham, C.M.; Murphy, F.V.t.; Weixlbaumer, A.; Petry, S.; Kelley, A.C.; Weir, J.R.; Ramakrishnan, V. Structure of the 70S ribosome complexed with mRNA and tRNA. *Science* **2006**, *313*, 1935-1942, doi:10.1126/science.1131127.

79. Burton, Z.F. The 3-Minihelix tRNA Evolution Theorem. *J Mol Evol* **2020**, *88*, 234-242, doi:10.1007/s00239-020-09928-2.

---

80. Yarus, M. Evolution of the Standard Genetic Code. *J Mol Evol* **2021**, *89*, 19–44, doi:10.1007/s00239-020-09983-9.