How the Early Genetic Code Was Established?: Inference from the Analysis of Extant Animal Mitochondrial Decoding Systems

Kimitsuna Watanabe and Shin-ichi Yokobori

Contents

Abstract Mitochondria are intracellular organelles in eukaryotic cells that have their own genome and translational apparatus. The vertebrate mitochondrial decoding system is thought to be the simplest among all extant living systems and to have originated by retrogression from the universal decoding system, induced mainly by genome economisation and directional mutation pressure during mitochondrial evolution. Thus, it is reasonable to speculate that the vertebrate genetic code table is a typical model for the early genetic code table.

In some metazoan mitochondrial decoding systems, it was found that unmodified anticodons of tRNA have the potential to base-pair with cognate codons

K. Watanabe (✉)
Department of Applied Life Sciences, School of Life Sciences, Tokyo University of Pharmacy and Life Sciences, 1432-1 Horinouchi, Hachioji, Tokyo 192-0392, Japan

Department of Biotechnology, Graduate School of Agricultural and Life Sciences, The University of Tokyo, 1-1-1 Yayoi, Bunkyo-ku, Tokyo 113-8657, Japan
e-mail: kim-wata@toyaku.ac.jp; kwatanab@ft.catv.ne.jp

S. Yokobori
Department of Applied Life Sciences, School of Life Sciences, Tokyo University of Pharmacy and Life Sciences, 1432-1 Horinouchi, Hachioji, Tokyo 192-0392, Japan

V.A. Erdmann et al. (eds.), *Chemical Biology of Nucleic Acids*, RNA Technologies,
DOI 10.1007/978-3-642-54452-1_2,

and retain flexibility at the wobble pair, suggesting that the early decoding system consisted of unmodified RNA prior to the emergence of RNA-modifying systems. Competition would likely have occurred between G_{34}-tRNA having GNN anticodons and U_{34}-tRNA having UNN anticodons (or C_{34}-tRNA having CNN anticodons for AUR codons) in their binding to the ribosomal A site, which would have resulted in the discrimination between NNY codons and NNR codons in the two-codon sets. Thus, the early genetic code table would have been established in such a way that eight family box codons were deciphered by U_{34}-tRNA, and eight NNY and six NNR codons in the two-codon sets were deciphered by G_{34}-tRNA and U_{34}-tRNA (or C_{34}-tRNA), respectively.

This review describes the characteristics of an early decoding system inferred from the genetic code of present vertebrate mitochondria, and how the present universal decoding system may have originated from the early decoding system during evolutionary history.

Keywords Genetic code • Mitochondria • tRNA anticodon • Family box • 2-Codon set

1 Introduction

The genetic code is the critical principle underlying the cellular interpretation of genetic information and is considered common to all living systems (the universal genetic code; Osawa 1995). However, some non-universal codons are utilised by animal mitochondria (Barrell et al. 1979; Anderson et al. 1981) and free-living organisms such as *Mycoplasma* spp. (Yamao et al. 1985). It is thus hypothesised that the genetic code is changeable and species dependent (Osawa et al. 1992). Here, the early genetic code preceding the universal genetic code is deduced through inspection of the decoding systems of numerous metazoan mitochondria (Watanabe 2010).

Mitochondria are proposed to have originated from aerobic bacteria engulfed by ancestral host bacteria (Margulis 1970). Molecular phylogenetic analysis of ribosomal RNA sequences and genes housed in the mitochondrial genome indicates that the mitochondrial ancestor is α-proteobacterial (Andersson et al. 1998). The ancestral host cell for mitochondria is thought to be a cell wall-less archaeon such as thermoplasmas (Margulis 1993). The mitochondrial genome houses very many fewer genes than bacterial genomes; for example, there are 37 human mitochondrial genes, while most bacteria have more than 1,000 genes. Correspondingly, most of the proteins that function within the mitochondria are encoded by the nuclear genome and are synthesised in the cytoplasm, indicating large-scale gene transfer from the mitochondrial to the nuclear genome (Lang et al. 1977). The mitochondrial genome size is thus diminished, varies depending on animal phylum,

and is only 20,000 bases in metazoan mitochondria (NCBI DB for Organelle genome). This diminution is called "genome economisation". Genome economisation is often accompanied by AT pressure, in which the base composition of DNA is biased towards AT richness by evolutionary mutation pressure. AT pressure would have been in effect during the evolutionary transition from α-proteobacteria, via proto-mitochondria, to the present metazoan mitochondrial state.

On the basis of the simplest anticodon composition of vertebrate mitochondria among various extant decoding systems, Jukes (1983) proposed that the mitochondrial genetic code was brought about by retrogression from the universal genetic code to the early genetic code through genomic economisation and accompanying AT pressure. Osawa (1995) proposed a simple early genetic code and its possible development into the present universal code; the simple code consists of 23 tRNA species for 61 sense codons and is based on the vertebrate mitochondrial genetic code.

Here, the evolution of mitochondrial genetic codes is re-examined using the most recent experimental results on mitochondrial tRNAs from various animal phyla. An early genetic code table and concomitant decoding systems are inferred, assuming retrogression from the present systems possessing the universal genetic code.

2 Comparison of Genetic Apparatus in Mitochondrial and Non-mitochondrial Decoding Systems

Table 1 shows the comparison of translation apparatus of vertebrate mitochondria as a representative of mitochondria, of *Mycoplasma capricolum* as a representative of free-living organisms possessing the simplest decoding system, and of *Escherichia coli* as a representative of organisms possessing the universal genetic code. Although the ribosomal components of *M. capricolum* and *E. coli* are almost the same, those of vertebrate mitochondria differ, having rRNA about half the size of *E. coli* rRNA and approximately 50 % more ribosomal proteins than *E. coli*. The functional domains composing the peptidyltransferase centre in the large ribosomal subunit (Suzuki et al. 2001a) and the decoding centre in the small ribosomal subunit (Suzuki et al. 2001b) are well conserved between vertebrate mitochondria and *E. coli*.

The number of tRNA species varies, with 22 in vertebrate mitochondria, 29 in *M. capricolum* and 48 in *E. coli*. One tRNA species exists per family box and two-codon set in vertebrate mitochondria, but multiple tRNA species exist in *M. capricolum* and *E. coli*. *M. capricolum* and *E. coli* possess two different species of $tRNA^{Met}$, one for initiation and one for elongation; however, vertebrate mitochondria possess a single species of $tRNA^{Met}$ that performs both functions (Takemoto et al. 2009). *E. coli* uses the universal genetic code, but

Table 1 Comparison of translation apparatus from vertebrate mitochondria and the free-living organisms *Mycoplasma capricolum* and *Escherichia coli*

	Mitochondria	Free-living organisms	
	Vertebrate mitochondria	*Mycoplasma capricolum*	*Escherichia coli*
	Non-universal		Universal
Genetic code	UGA = Trp, AUA = Met, AGA/G = Stop	UGA = Trp	
	60 sense codons	62 sense codons	61 sense codons
	4 stop codons	2 stop codons	3 stop codons
tRNA	22 species	29 species	42–48 species[a]
$tRNA^{Met}$	1 species	2 species	2 species
Ribosome	55S (28S + 39S)	70S (30S + 50S)	70S (30S + 50S)
rRNA	2,513 nt.	4,527 nt.	4,566 nt.
	12S + 16S	5S + 16S + 23S	5S + 16S + 23S
rProtein	78 species	51 species	54 species

[a]There are 42 different anticodons and a total of 48 tRNAs with different body sequences in *E. coli* (Näsvall et al. 2007)

M. capricolum and vertebrate mitochondria use non-universal genetic codes. In *M. capricolum*, the universal UGA stop codon becomes a Trp codon (Yamao et al. 1985), and in vertebrate mitochondria the AGA/AGG Arg codons become stop codons and the AUA Ile codon changes to a Met codon (Barrell et al. 1979; Anderson et al. 1981). The vertebrate mitochondrial system is the simplest genetic code system among all extant organisms so far examined (Suzuki et al. 2011a).

3 Characteristics of Mitochondrial Genetic Code Systems

The most prominent characteristic of animal mitochondrial genetic code systems is that they use non-universal genetic codes: (1) UGA, Stop becomes Trp in all animal mitochondria (Barrell et al. 1979); (2) AUA, Ile becomes Met in most metazoan mitochondria, excepting Echinodermata, Hemichordata, Platyhelminthes, Cnidaria, Ctenophora, Placozoa, and Porifera (Watanabe 2010); (3) AGR, Arg becomes Ser in most invertebrate mitochondria (Himeno et al. 1987), Arg becomes Gly in ascidian mitochondria (Yokobori et al. 1993), and Arg becomes Stop in vertebrate mitochondria (Anderson et al. 1981); (4) AAA, Lys becomes Asn in Echinodermata (Himeno et al. 1987) and Platyhelminth mitochondria (Telford et al. 2000); and (5) UAA, Stop becomes Tyr in a nematode, *Radopholus similis* (Jacob et al. 2009) and a planarian (Bessho et al. 1992) mitochondria. In the last case, no structural study on the corresponding tRNA was performed, and no release factor (RF) relevant to the UAA codon was identified, so discussion on the codon–anticodon relationship was not possible in this instance.

Modified nucleosides in the tRNA anticodons are associated with mitochondrial codon changes in all cases (Watanabe and Yokobori 2011): (1) The anticodon wobble position (the 34th) of $tRNA^{Trp}$ changes to 5-carboxymethyluridine ($cmnm^5U$) in nematode mitochondria (Sakurai et al. 2005) and to 5-taurinomethyluridine (τm^5U) in ascidian (Suzuki et al. 2011b), molluscan (Ohira et al. 2013), and vertebrate (Suzuki et al. 2002, 2011a) mitochondria; (2) The anticodon wobble position of $tRNA^{Met}$ changes to 5-formylcytidine (f^5C) in vertebrate (Moriya et al. 1994), arthropod (Tomita et al. 1999a), and nematode mitochondria (Watanabe et al. 1994) and to τm^5U in ascidian mitochondria (Suzuki et al. 2011b); (3) The anticodon wobble nucleoside of $tRNA^{Gly}$ for AGR codons becomes τm^5U in ascidian mitochondria (Suzuki et al. 2011b). The anticodon wobble nucleoside of $tRNA^{Ser}$ becomes m^7G in most invertebrate mitochondria (Matsuyama et al. 1998; Tomita et al. 1998) and unmodified U in nematode mitochondria (Watanabe et al. 1994); (4) The starfish mitochondrial $tRNA^{Asn}$ that deciphers AAA as Asn possesses the anticodon GΨU (Ψ, pseudouridine) (Tomita et al. 1999b); this is unusual in that it is a change at the second anticodon position. In vitro *E. coli* translation experiments indicate that $tRNA^{Asn}_{G\Psi U}$ has an approximately twofold higher translational efficiency than $tRNA^{Asn}_{GUU}$ (Tomita et al. 1999b). In other invertebrate mitochondria such as *Drosophila melanogaster*, $tRNA^{Lys}$ interprets AAR as Lys and has the anticodon CUU (Tomita et al. 1999a). Therefore, $tRNA^{Lys}_{CUU}$ should read AAR codons in a manner similar to the case of squid mitochondrial $tRNA^{Met}_{CAU}$ described subsequently.

The previously mentioned five nucleoside species ($cmnm^5U$, τm^5U, m^7G, Q, and f^5C) or seven species including the 2-thiolated forms of $cmnm^5U$ ($cmnm^5s^2U$) and τm^5U (τm^5s^2U) are all the modified nucleosides found in the anticodon wobble position of metazoan mitochondrial tRNAs. As shown in Table 2, $cmnm^5(s^2)U$, $\tau m^5(s^2)U$ and f^5C permit base pairing only with purine nucleosides (R_3) excluding base pairing with pyrimidine nucleosides (Y_3) of the cognate codons, probably through fixing their conformations in the C3′-endo form (Yokoyama and Nishimura 1995; Takemoto et al. 2009). This was proved experimentally using *E. coli* and bovine mitochondrial in vitro translation systems (Takai et al. 1994; Kurata et al. 2008; Takemoto et al. 2009). Queosine (Q, a G derivative), which forms base pairing only with pyrimidines, is sometimes found in the anticodon wobble position of *D. melanogaster* mitochondrial $tRNA^{Asn}_{GUU}$, but not in the anticodon wobble position of starfish mitochondrial $tRNA^{Asn}_{G\Psi U}$ (Tomita et al. 1999a, b). This indicates that *D. melanogaster* mitochondrial $tRNA^{Asn}_{G/QUU}$ restricts codon recognition to AAY codons, but starfish mitochondrial $tRNA^{Asn}_{G\Psi U}$ can decode not only AAY codons but also the AAA codon. It is likely that the m^7G located at the wobble position of $tRNA^{Ser}$ of most invertebrate mitochondria allows decoding of all four family box codons (Matsuyama et al. 1998; Tomita et al. 1998), although this speculation has not been verified experimentally (Watanabe and Yokobori 2011).

Nucleoside modifications often provide mechanisms for variable codon recognition; however, unmodified anticodon nucleosides have also been shown to decode cognate codons in some metazoan mitochondria (Table 2). For example, U_{34} was

Table 2 Mitochondrial pairing possibilities between the anticodon first position (N_{34}, wobble nucleotide) of tRNA and the codon third position (N_3)

Pairing rule	N_{34} of tRNA		N_3 of codon	Existence
Unmodified	U	⟷	U, C, A, G	All tRNAs belonging to family box of almost all animals
	A	⟷	U, C, A, G	Nematode $tRNA^{Arg}_{ACG}$
	G	⟷	U, C, A	Insect $tRNA^{Ser}_{GCU}$ for AGY/A codons
	C	⟷	A, G	Squid $tRNA^{Met}_{CAU}$, Insect $tRNA^{Lys}_{CUU}$
Modified	$cmnm^5(s^2)U$	⟷	A, G	Nematode $tRNA^{Trp}$
	$\tau m^5(s^2)U$	⟷	A, G	Ascidian $tRNA^{Trp}$, vertebrate $tRNA^{Trp}$, squid $tRNA^{Trp}$
				Ascidian $tRNA^{Met}$, ascidian $tRNA^{Gly}_{U*CU}$, vertebrate $tRNA^{Leu}_{U*AA}$
				Vertebrate tRNAs for NNR codons in the third column of codon table
	m^7G	⟷	U, C, A, G	Most invertebrate $tRNA^{Ser}_{GCU}$ for AGN codons
	Q	⟷	U, C	Vertebrate tRNAs for NNY codons in the third column of codon table
	f^5C	⟷	A, G	Vertebrate $tRNA^{Met}$, arthropod $tRNA^{Met}$, nematode $tRNA^{Met}$

shown experimentally to base-pair with all four nucleosides at the third codon position (Suzuki et al. 2011a). Unmodified U_{34} can adopt flexible conformations of either the C3′-endo or C2′-endo form, and this allows base pairing with pyrimidines in addition to purines. Unmodified A_{34} has a similar role in nematode mitochondrial $tRNA^{Arg}_{ACG}$ (Watanabe et al. 1997). Unmodified G is thought to be able to base-pair with A as well as C and U, because the AGG codon is absent in some metazoan mitochondria, such as those from *D. melanogaster*, and the anticodon wobble position of the relevant $tRNA^{Ser}_{GCU}$ is unmodified G (Tomita et al. 1999a). Therefore, the unmodified GCU anticodon of $tRNA^{Ser}_{GCU}$ should read the three codons AGU, AGC, and AGA as Ser. Recently, it was found that squid (*Loligo bleekeri*) mitochondrial $tRNA^{Met}$ possesses unmodified C at the anticodon wobble position (Ohira et al. 2013), suggesting that, in this case, the AUA codon is deciphered by the CAU anticodon via noncanonical A_3–C_{34} pairing. The possibility of such base-pair formation had previously been reported in the case of *D. melanogaster* $tRNA^{Lys}_{CUU}$ deciphering AAR codons (Tomita et al. 1999a).

Two mitochondrial RFs, mtRF1a and mtRF1, have been identified to date; mtRF1a recognises UAG and UAA codons (Soleimanpour-Lichaei et al. 2007), and mtRF1, which arose from a duplication of mtRF1a, probably binds the

ribosome when the A site is devoid of mRNA (Huynen et al. 2012). An RF that recognised the UGA codon would have been lost in the transition process from pre-mitochondrial ancestral bacteria (α-proteobacteria) to the present mitochondria (Andersson et al. 1998). There has been no report identifying any RF able to recognise AGA/AGG codons directly. Instead, mtRF1a recognising UAR codons in vertebrate mitochondria (Soleimanpour-Lichaei et al. 2007) are thought to be used for translation termination of the AGR codons, as these would be changed to the UAG codon by -1 frame-shifting (Temperley et al. 2010). However, this hypothesis may not be generally applicable as the necessary U may not always be located prior to the AGR codons in the genome.

4 The Codon–Anticodon Relationship in Animal Mitochondria

As shown in previous sections, six non-universal codons (the UAA case is not discussed here) have been identified in animal mitochondria, and these vary over the course of animal evolution. As already mentioned, we have elucidated that the anticodons of tRNAs deciphering these non-universal codons ($tRNA^{Trp}$ for UGA, $tRNA^{Met}$ for AUA, $tRNA^{Asn}$ for AAA, and $tRNA^{Ser}$ and $tRNA^{Gly}$ for AGR) are all modified. However, besides these modified nucleosides involved in decoding non-universal codons, it has been found that unmodified nucleosides also exist at the wobble position of tRNA in the codon–anticodon interaction in some of the mitochondrial translation systems; unmodified U_{34} exists in $tRNA^{Ser}$ deciphering AGR codons in nematode (*Ascaris suum*) mitochondria (Watanabe et al. 1994). Apart from non-universal codon decoding, all tRNAs for the family box codons possess unmodified U_{34}. Unmodified U_{34} is also known to be present in the non-mitochondrial free-living organism systems such as *M. capricolum* (Inagaki et al. 1995) and decodes all four codons in the family boxes. Using an in vitro translation system of *M. capricolum*, it was experimentally verified that threonyl-$tRNA^{Thr}_{UGU}$ and alanyl-$tRNA^{Ala}_{UGC}$, both possessing unmodified U in the anticodon wobble position, could translate all four ACN and GCN codons, respectively (Andachi et al. 1989; Osawa 1995; Inagaki et al. 1995). An in vivo experiment in a plastid using knockout mutants for the $tRNA^{Gly}$ with G_{34}-containing anticodon in the GGN family box showed that the remaining $tRNA^{Gly}$ with U_{34}-containing anticodon was sufficient for survival of the plastid (Rogaiski et al. 2008), which also indicates that U_{34}-$tRNA^{Gly}$ can decipher all the GGN codons. In addition, unmodified A_{34} is present in the nematode (*Ascaris suum*) mitochondrial $tRNA^{Arg}$ deciphering GCN codons (Watanabe et al. 1997), in budding yeast (*Saccharomyces cerevisiae*) mitochondrial $tRNA^{Arg}_{ACU}$ deciphering CGN codons (Sibler et al. 1986), and in *Mycoplasma* spp. $tRNA^{Thr}_{AGU}$ deciphering ACN codons (Andachi et al. 1987). The translation activity of threonyl-$tRNA^{Thr}_{AGU}$ of *M. capricolum* toward all four ACN codons was experimentally verified (Inagaki

et al. 1995). G_{34}-tRNA^{Ser} (G_{34} unmodified) deciphers three codons (AGU/AGC/AGA) in *D. melanogaster* mitochondria (Tomita et al. 1999a), C_{34}-tRNA^{Met} (C_{34} unmodified) deciphers AUR codons in squid (*Loligo bleekeri*) (Ohira et al. 2013), and C_{34}-tRNA^{Lys} (C_{34} unmodified) deciphers AAR codons in *D. melanogaster* (Tomita et al. 1999a). These data indicate that unmodified U_{34} (or, rarely, A_{34}) at the anticodon wobble position (N_{34}) can potentially pair with all four nucleosides at the codon third position (N_3), unmodified G_{34} can potentially pair with U_3, C_3, and A_3, and unmodified C_{34} can potentially pair with A_3 and G_3. Taken together, this strongly suggests that all codons were deciphered by unmodified tRNA in the early genetic code system; this would be prior to the emergence of the universal genetic code and RNA-modifying systems, the existence of which is difficult to imagine in the beginning of the genetic code systems.

Table 2 summarises the pairing patterns between anticodon first nucleoside (N_{34}) and codon third nucleoside (N_3) so far elucidated in animal mitochondrial translation systems.

5 How Was the Mitochondrial Decoding System Generated?

The mitochondrial decoding system is putatively derived from retrogression of the universal decoding system. This retrogression was mainly induced by genome economisation and directional mutation pressure, accompanying various events such as genetic code changes, reduction of tRNA species, and loss of modification enzymes.

Two main hypotheses have been proposed to explain genetic code changes: the "codon capture" hypothesis assuming an "unassigned codon" (Osawa and Jukes 1989) and the "ambiguous intermediate" hypothesis, which assumes two different tRNAs recognising a single codon (Schultz and Yarus 1994). The codon capture hypothesis successfully explains a codon change of the UGA Trp codon to a stop codon in *Mollicutes* (Osawa et al. 1992), but there has been little experimental evidence in support of the ambiguous intermediate hypothesis. Simulation studies on codon reassignment (Sengupta et al. 2007) suggest that the pathway of codon reassignment in favour of either "codon capture" or "ambiguous codon" depends on initial conditions such as genome size and number of codons involved. For example, as mitochondria possess a small genome size and a limited number of involved codons, the "codon capture" hypothesis can account for all the codon changes concerning the non-universal codons found in animal mitochondria (Watanabe and Yokobori 2011).

If codon change and the concomitant reduction of the number of tRNA anticodon species are regarded as the first event in the retrogression process of the mitochondrial decoding system, the second event is the removal of modification enzymes. The following events are proposed during the process of altering the

universal genetic code to the vertebrate mitochondrial code. After removal of U_{34}-modification enzymes for U*NN anticodons of tRNA in family boxes during evolution, the resultant UNN anticodons (U unmodified) would become translatable for all four codons; this is observed in the present genetic codes for vertebrate mitochondria and in *Mycoplasma* spp. The GNN anticodons and CNN anticodons that had translated NNY and NNG codons would thus become unnecessary and be eliminated. In two-codon sets, U*NN anticodons for NNY codons and GNN or QNN anticodons for NNR codons would be retained because NNY and NNR codons code for different amino acids. In addition, $tRNA^{Arg}$ for AGR codons would be eliminated, probably by genome economisation. Once mtRF1a (if it recognised UAG codon generated by -1 frame-shifting of the AGR codons) or a new RF (recognising AGR codons directly) emerged, it would result in the change of AGR to a stop codon. This combination of events would finally result in 22 anticodon species and retention of modified nucleosides, as observed at present in vertebrate mitochondria.

6 The Early Genetic Code System

Vertebrate mitochondria use 60 sense codons and four stop codons and have a total of 22 tRNA species, including a single tRNA for eight family boxes and 14 - two-codon sets. One single $tRNA^{Met}$ functions as both initiator and elongator. The vertebrate mitochondrial genetic code system is the simplest genetic code system observed to date among extant organisms and can be used as a model of the early genetic code (Table 3). As noted previously, RNAs engaged in the early genetic code system would likely have been unmodified. All the family box codons are deciphered by U_{34}-tRNA having UNN anticodons. The two-codon sets are divided into two categories; the upper half (NNY codons) are read by GNN anticodons and the lower half (NNR codons) are read by UNN anticodons or CNN anticodons for AUR. How then would NNY codons have been distinguished from NNR codons in the early translation system? As illustrated in Fig. 1, U_{34}-tRNA has the potential to decipher any of the four possible codons, G_{34}-tRNA has a potential to decipher NNY and NNA codons, and C_{34}-tRNA has a potential to decipher NNY codons. Competition would likely have occurred between G_{34}-tRNA and U_{34}-tRNA (or C_{34}-tRNA for AUR codons) in their binding to the ribosomal A site. To separate the upper half and lower half codons in the two-codon sets, the binding affinity of G_{34}-tRNA with NNY codons at the ribosomal A site must be stronger than that of U_{34}-tRNA with NNY codons in the upper half, whereas the binding affinity of U_{34}-tRNA (or C_{34}-tRNA) with NNR codons must be stronger than that of G_{34}-tRNA with NNA codons in the lower half. Specifically, the base pairing between G_{34} and U_3 and between G_{34} and C_3 must be stronger than that between U_{34} and U_3 and between U_{34} and C_3. Likewise, the base pairing between U_{34} (or C_{34}) and A_3 must be stronger than that between G_{34} and A_3. These speculations are reasonable as R–Y pairing should be stronger than Y–Y or R–R pairing, and this is experimentally

Table 3 An early genetic code table proposed in this study

Codon	Anticodon	Amino acid	Codon	Anticodon	Amino acid	Codon	Anticodon	Amino acid	Codon	Anticodon	Amino acid
UUU	GAA	Phe	**UCU**	**UGA**	**Ser**	UAU	GUA*3	Tyr	UGU	GCA	Cys
UUC			**UCC**			UAC			UGC		
UUA	UAA*1	Leu	**UCA**			UAA		Stop	*UGA*	*UCA*1*	*Trp*
UUG			**UCG**			UAG			*UGG*		
CUU	**UAG**	**Leu**	**CCU**	**UGG**	**Pro**	CAU	GUG*3	His	**CGU**	**UCG**	**Arg**
CUC			**CCC**			CAC			**CGC**		
CUA			**CCA**			CAA	UUG*4	Gln	**CGA**		
CUG			**CCG**			CAG			**CGG**		
AUU	GAU	Ile	**ACU**	**UGU**	**Thr**	AAU	GUU*3	Asn	AGU	GCU	Ser
AUC			**ACC**			AAC			AGC		
AUA	*CAU*2*	*Met*	**ACA**			AAA	UUU*4	Lys	*AGA*		*Stop*
AUG			**ACG**			AAG			*AGG*		
GUU	**UAC**	**Val**	**GCU**	**UGC**	**Ala**	GAU	GUC*3	Asp	**GGU**	**UCC**	**Gly**
GUC			**GCC**			GAC			**GGC**		
GUA			**GCA**			GAA	UUC*4	Glu	**GGA**		
GUG			**GCG**			GAG			**GGG**		

Family box codons and specific codons related to codon change are shown by bold and italic emphases, respectively. This code is same as that in vertebrate mitochondria, with the exception that the following nucleosides are modified in vertebrate mitochondria: *1, 5-taurinomethyluridine (τm^5U); *2, 5-formylcytidine (f^5C); *3, queosine; *4, 5-taurinomethyl-2-thiouridine (τm^5s^2U)

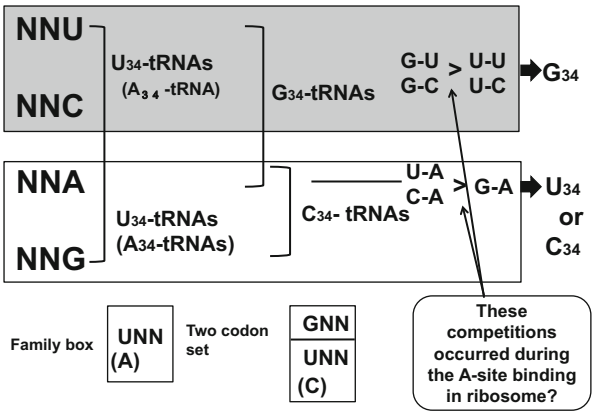

Fig. 1 Decoding scheme of two-codon set codons by competition between U_{34}-tRNA and G_{34}-tRNA (or C_{34}-tRNA for AUR codons) in the early genetic code system

verifiable. In summary, it is proposed that the early genetic code table would have consisted of eight family boxes, 14 two-codon sets, and four stop codons and would have been translated by 22 tRNA species. It remains to be experimentally determined whether all 64 codons were used in the early genetic code. UUR/CUN codons for Leu, CGN/AGR codons for Arg, and UCN/AGY codons for Ser are redundant, so the removal of one or several redundant codons would not diminish the number of different amino acids (20 species) available for protein synthesis. One of these redundancies is exploited in vertebrate mitochondria, where AGR codons become stop codons.

7 From the Early Decoding System to the Universal Decoding System

It is conceivable that the early decoding system could have evolved to the present universal decoding system by almost the reverse process of the evolution of the mitochondrial decoding system (Osawa 1995; Osawa et al. 1992). The process would, therefore, involve genome enlargement due to gene duplication, concomitant GC pressure, emergence of new tRNA genes accompanied by nucleoside modification, and the resulting reassignment of three kinds of codon: UGA from Trp to stop codon, AUA from Met to Ile, and AGR from stop codon to Arg. The progression of these reassignment processes is proposed as follows. For the UGA codon, the anticodon wobble position of $tRNA^{Trp}$ may have changed from U to C by GC pressure, so that the resultant CCA anticodon became readable only by the UGG codon, and the UGA codon became unassigned. An RF recognising the UGA codon (e.g. RF2 in *E. coli*) could then emerge and capture the UGA codon, transforming it into a stop codon. Acetylation enzyme modification of C_{34} of $tRNA^{Met}_{CAU}$ to ac^4C_{34} would lead the resulting $tRNA^{Met}_{ac^4CAU}$ to become readable only for the AUG codon, and the AUA codon would become unassigned. In fact,

E. coli $tRNA^{Met}_{m}$ (elongator) possesses such an ac^4CAU anticodon, which reads the AUG codon as Met, but does not read the AUA codon (Stern and Schulman 1978). Subsequently, the AUA codon could be captured by $tRNA^{Ile}_{LAU}$ (L (or k^2C), lysidine, 2-lysylcytidine present in *E. coli* and *Mycoplasma* $tRNA^{Ile}$) or $tRNA^{Ile}_{agm^2CAU}$ (agm^2C, 2-agmatinylcytidine present in Archaea), which originated from $tRNA^{Met}_{CAU}$ via the L-forming enzyme TilS (Soma et al. 2003) or the agm^2C-forming enzyme, TiaS (Ikeuchi et al. 2010). Thus, the AUA codon would be deciphered as Ile (Muramatsu et al. 1988; Ikeuchi et al. 2010). Although TilS and TiaS belong to quite different enzyme families, the chemical structures of L and agm^2C are quite similar, so that $tRNA^{Ile}_{LAU}$ and $tRNA^{Ile}_{agm^2CAU}$ could have acquired decoding ability for the AUA codon by convergent evolution. For AGR codons, the RF recognising the codons could be eliminated from the genome, then $tRNA^{Ser}_{GCU}$ could became readable for the AGA codon in addition to AGY codons, and the AGG codon would become unassigned. Genome enlargement would allow emergence of $tRNA^{Arg}_{UCU}$, which could capture the AGA codon as well as the AGG codon as Arg, because $tRNA^{Arg}_{UCU}$ would prevail over $tRNA^{Ser}_{GCU}$ in the binding of AGA codon, as discussed previously.

During this process, the number of tRNAs involved in the translation system would increase, probably initially to the number used in the *Mycoplasma* system (28 species), and finally to the number in the *E. coli* system (42–48 species). This would occur in parallel with nucleoside modifications, particularly G_{34}-tRNA modification to Q_{34}-tRNA corresponding to NNY codons in the upper half of two-codon sets, and modification of the wobble U of U_{34}-tRNA to various U derivatives, such as mnm^5U, mnm^5s^2U, $cmnm^5U$, $cmnm^5U_m$, or $cmnm^5s^2U$ in the lower half of two-codon sets, as in *Mycoplasma* and *E. coli* systems. Modification of G_{34} and U_{34} would be necessary for strict discrimination of UUY and UUR codon boxes by G_{34}-tRNA and U_{34}-tRNA.

To summarise, Fig. 2 shows a phylogenetic tree of three domains of life, mitochondrial origination from α-proteobacteria, and the emergence of universal and non-universal genetic codes. It should be emphasised that early decoding was performed by unmodified RNA, and the universal decoding system emerged after the development of RNA modification.

8 Concluding Remarks and Perspective

In this review, the following concepts were newly deduced from the study of metazoan mitochondrial decoding systems.

1. The change of AGR codons during the retrogression process was found to follow a route from Arg (universal code) to stop codons (vertebrate mitochondria) via Ser codons (most invertebrate mitochondria) (Gly codons in *Cnidarian* mitochondria seem to be a sub-route). This concept is useful for considering the

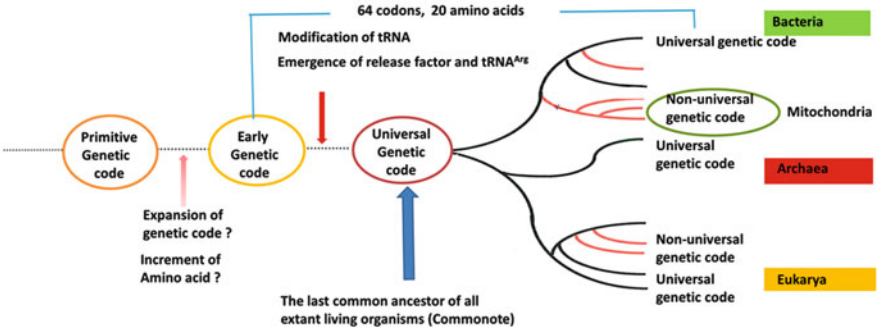

Fig. 2 Three domains of life indicating the position of mitochondria in the evolutionary process. The assignment of 20 amino acids to 64 codons would have been conserved from the early genetic code system. The early genetic system utilised unmodified RNA and developed into the universal code system after the emergence of RNA modification

evolutional processes by which AGR codon-use changed from the early code to the present universal code.

2. Examples from mitochondria indicate that any of the unmodified nucleosides at the anticodon wobble position of tRNA are available for the decoding of the corresponding codon at the ribosomal A site. It is thus possible to infer that unmodified tRNA had deciphered all codons in the early genetic code system, prior to the emergence of the universal genetic code and RNA-modifying systems.
3. The most plausible explanation for the discrimination of the NNY codon from the NNR codon frame in the two-codon sets is that the discrimination is fulfilled by competition between G_{34}-tRNA and U_{34}-tRNA (or C_{34}-tRNA for AUR codons) in their binding to the ribosomal A site. Thus, the early genetic code table would have been established in such a way that eight family box codons were deciphered by U_{34}-tRNA, and eight NNY and six NNR codons in the two-codon sets were deciphered by G_{34}-tRNA and U_{34}-tRNA (or C_{34}-tRNA in part), respectively. This genetic code table is the simplest among the extant living organisms, consisting of eight family boxes and 14 two-codon sets, decoded by 22 species of tRNA.

Studies of the mitochondrial genetic code and decoding apparatus from different animal phyla, alongside data from organisms using the universal code, have produced new insights into the evolution of non-universal genetic systems. Examination of such evolutionary traces from the mitochondrial genetic code systems of additional animal phyla would be useful to allow further elucidation of the early decoding system and the origins of the genetic code in the RNA world.

Acknowledgements We thank Dr. Tsutomu Suzuki of the University of Tokyo for supplying us with analytical data of mitochondrial tRNAs and his wholehearted support of this study. We also thank Dr. Akira Muto, Hirosaki University, for *Mycoplasma* ribosome data, and Dr. Takeo Suzuki, University of Tokyo, for *E. coli* tRNA data.

References

Andachi Y, Yamao F, Iwami M et al (1987) Occurrence of unmodified adenine and uracil at the first position of anticodon in threonine tRNAs in *Mycoplasma capricolum*. Proc Natl Acad Sci U S A 84:7398–7402

Andachi Y, Yamao F, Muto A et al (1989) Codon recognition patterns as deduced from sequences of the complete set of transfer RNA species in *Mycoplasma capricolum*: resemblance to mitochondria. J Mol Biol 209:37–54

Anderson S, Bankier AT, Barerll BG et al (1981) Sequence and organization of the human mitochondrial genome. Nature 290:457–465

Andersson SGE, Zomorodipour A, Andersson JO et al (1998) The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria. Nature 396:133–143

Barrell BG, Bankier AT, Drouin J (1979) A different genetic code in human mitochondria. Nature 282:189–194

Bessho Y, Ohama T, Osawa S (1992) Planarian mitochondria II. The unique genetic code as deduced from cytochrome c oxydase subunit I gene sequence. J Mol Evol 34:331–335

Himeno H, Masaki H, Kawai T et al (1987) Unusual genetic codes and a novel gene structure for $tRNA^{Ser}_{AGY}$ in starfish mitochondria. Gene 56:219–230

Huynen MA, Duarte I, Chrzanowska-Lightowlers ZMA et al (2012) Structure based hypothesis of a mitochondrial ribosome rescue mechanism. Biol Direct 7:14

Ikeuchi Y, Kimura S, Numata T et al (2010) Agmatine-containing cytidine in tRNA anticodon essential for AUA decoding in Archaea. Nat Chem Biol 6:277–282

Inagaki Y, Kojima A, Bessho Y et al (1995) Translation of synonymous codons in family boxes by *Mycoplasma capricolum* tRNAs with unmodified uridine or adenosine at the first anticodon position. J Mol Biol 251:486–492

Jacob JEM, Vanholme B, Van Leeuwen T et al (2009) A unique genetic code change in the mitochondrial genome of the parasitic nematode *Radopholus similis*. BMC Res Notes 2:192

Jukes TH (1983) Evolution of the amino acid code: inferences from mitochondrial codes. J Mol Evol 19:219–225

Kurata S, Weixlbaumer A, Ohtsuki T et al (2008) Modified uridines with C5-methylene substituents at the first position of the tRNA anticodon stabilize U-G wobble pairing during decoding. J Biol Chem 283:18801–18811

Lang BF, Burger G, O'Kelly CJ et al (1977) An ancestral mitochondrial DNA resembling a eubacterial genome in miniature. Nature 387:493–497

Margulis L (1970) Origin of eukaryotic cells. Yale University Press, New Haven, CT

Margulis L (1993) Symbiosis in cell evolution: microbial communities in the Archean and Proterozoic eons, 2nd edn. W. H. Freeman, New York

Matsuyama S, Ueda T, Crain PF et al (1998) A novel wobble rule found in starfish mitochondria. Presence of 7-methylguanosine at the anticodon wobble position expands decoding capability of tRNA. J Biol Chem 273:3363–3368

Moriya J, Yokogawa T, Wakita K et al (1994) A novel modified nucleoside found at the first position of the anticodon of methionine tRNA from bovine liver mitochondria. Biochemistry 33:2234–2239

Muramatsu T, Nishikawa K, Nemoto F et al (1988) Codon and amino acid specificities of a transfer RNA are both converted by a single post-transcriptional modification. Nature 294:187–188

Näsvall SJ, Chen P, Björk GR (2007) The wobble hypothesis revisited: uridine-5-oxyacetic acid is critical for reading of G-ending codons. RNA 13:1–14

NCBI DB for Organelle genome. http://www.ncbi.nlm.gov/genomes/OrganelleResource,cgi?taxid=33208

Ohira T, Suzuki T, Miyauchi K et al (2013) Decoding mechanism of non-universal genetic codes in *Loligo bleekeri* mitochondria. J Biol Chem 288:7645–7652

Osawa S (1995) Evolution of the genetic code. Oxford University Press, Tokyo

Osawa S, Jukes TH (1989) Codon reassignment (codon capture) in evolution. J Mol Evol 28:271–278

Osawa S, Jukes TH, Watanabe K et al (1992) Recent evidence for evolution of the genetic code. Microbiol Rev 56:229–264

Rogaiski M, Karcher D, Bock R (2008) Superwobbling facilitates translation with reduced tRNA sets. Nat Struct Mol Biol 15:192–198

Sakurai M, Ohtsuki T, Suzuki T et al (2005) Unusual usage of wobble modifications in mitochondrial tRNAs of the nematode *Ascaris suum*. FEBS Lett 579:2767–2772

Schultz DW, Yarus M (1994) Transfer RNA mutation and the malleability of the genetic code. J Mol Biol 235:1377–1380

Sengupta S, Yang X, Higgs PG (2007) The mechanisms of codon reassignments in mitochondrial genetic codes. J Mol Evol 64:662–688

Sibler AP, Dirheimer G, Martin RP (1986) Codon reading patterns in *Saccharomyces cerevisiae* mitochondria based on sequences of mitochondrial tRNAs. FEBS Lett 194:131–138

Soleimanpour-Lichaei HR, Kühl I, Gaisne M et al (2007) mtRF1a is a human mitochondrial translation release factor decoding the major termination codons UAA and UAG. Mol Cell 27:745–757

Soma A, Ikeuchi Y, Kanemasa S et al (2003) An RNA-modifying enzyme that governs both the codon and amino acid specificities of isoleucine tRNA. Mol Cell 12:689–698

Stern L, Schulman LH (1978) The role of the minor base N^4-acetylcytidine in the function of the *Escherichia coli* non-initiator methionine transfer RNA. J Biol Chem 253:6132–6139

Suzuki T, Terasaki M, Takemoto-Hori C et al (2001a) Structural compensation for the deficit of rRNA with proteins in the mammalian mitochondrial ribosome. J Biol Chem 276:21724–21736

Suzuki T, Terasaki M, Takemoto-Hori C et al (2001b) Proteomic analysis of the mammalian mitochondrial ribosome. J Biol Chem 276:33181–33195

Suzuki T, Suzuki T, Wada T et al (2002) Taurine as a constituent of mitochondrial tRNAs: new insights into the function of taurine and human mitochondrial diseases. EMBO J 21:6581–6589

Suzuki T, Nagao A, Suzuki T (2011a) Human mitochondrial tRNAs: biogenesis, function, structural aspects, and diseases. Annu Rev Genet 45:299–329

Suzuki T, Miyauchi K, Suzuki T et al (2011b) Taurine-containing uridine modifications in tRNA anticodons are required to decipher non-universal genetic codes in ascidian mitochondria. J Biol Chem 286:35494–35498

Takai K, Horie N, Yamaizumi Z et al (1994) Recognition of UUN codons by two leucine tRNA species from *Escherichia coli*. FEBS Lett 344:31–34

Takemoto C, Spremulli LL, Benkowski LA et al (2009) Unconventional decoding of the AUA codon as methionine by mitochondrial $tRNA^{Met}$ with the anticodon f^5CAU as revealed with a mitochondrial *in vitro* translation system. Nucleic Acids Res 37:1616–1627

Telford MJ, Herniou EA, Russel RB et al (2000) Changes in mitochondrial genetic codes as phylogenetic characters: two examples from the flatworms. Proc Natl Acad Sci U S A 97:11359–11364

Temperley R, Richter R, Dennerlein S et al (2010) Hungry codons promote frameshifting in human mitochondrial ribosomes. Science 327:301

Tomita K, Ueda T, Watanabe K (1998) 7-methylguanosine at the anticodon wobble position of squid mitochondrial $tRNA^{Ser}GCU$: molecular basis for assignment of AGA/AGG codons as serine in invertebrate mitochondria. Biochim Biophys Acta 1399:78–82

Tomita K, Ueda T, Ishiwa S, Crain PF, McCloskey JA, Watanabe K (1999a) Codon reading patterns in *Drosophila melanogaster* mitochondria based on their tRNA sequences: a unique wobble rule in animal mitochondria. Nucleic Acids Res 27:4291–4297

Tomita K, Ueda T, Watanabe K (1999b) The presence of pseudouridine in the anticodon alters the genetic code: a possible mechanism for assignment of the AAA lysine codon as asparagine in echinoderm mitochondria. Nucleic Acids Res 27:1683–1689

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

Watanabe K, Yokobori S (2011) tRNA modification and genetic code variations in animal mitochondria. J Nucleic Acids 2011:Article ID 623095

Watanabe Y, Tsurui H, Ueda T et al (1994) Primary and higher order structures of nematode (*Ascaris suum*) mitochondrial tRNAs lacking either the T or D stem. J Biol Chem 269:22902–22906

Watanabe Y, Tsurui H, Ueda T et al (1997) Primary sequence of mitochondrial $tRNA^{Arg}$ of a nematode *Ascaris suum*: occurrence of unmodified adenosine at the first position of the anticodon. Biochim Biophys Acta 1350:119–122

Yamao F, Muto A, Kawauchi Y et al (1985) UGA is read as tryptophan in *Mycoplasma capricolum*. Proc Natl Acad Sci U S A 82:2306–2309

Yokobori S, Ueda T, Watanabe K (1993) Codons AGA and AGG are read as glycine in ascidian mitochondria. J Mol Evol 36:1–8

Yokoyama S, Nishimura S (1995) Modified nucleosides and codon recognition. In: Söll D, RajBhandary UL (eds) tRNA: structure, biosynthesis and function. ASM, Washington, DC, pp 207–223

http://www.springer.com/978-3-642-54451-4

Chemical Biology of Nucleic Acids
Fundamentals and Clinical Applications
Erdmann (Deceased), V.A.; Markiewicz, W.T.;
Barciszewski, J. (Eds.)
2014, XI, 599 p. 231 illus., 84 illus. in color., Hardcover
ISBN: 978-3-642-54451-4