

Review

Evolving genetic code

By Takeshi OHAMA,^{*1} Yuji INAGAKI,^{*2} Yoshitaka BESSHO^{*3} and Syozo OSAWA^{*4,*5,†}

(Communicated by Takao SEKIYA, M.J.A.)

Abstract: In 1985, we reported that a bacterium, *Mycoplasma capricolum*, used a deviant genetic code, namely UGA, a “universal” stop codon, was read as tryptophan. This finding, together with the deviant nuclear genetic codes in not a few organisms and a number of mitochondria, shows that the genetic code is not universal, and is in a state of evolution. To account for the changes in codon meanings, we proposed the codon capture theory stating that all the code changes are non-disruptive without accompanied changes of amino acid sequences of proteins. Supporting evidence for the theory is presented in this review. A possible evolutionary process from the ancient to the present-day genetic code is also discussed.

Keywords: genetic code, frozen accident theory, unassigned or nonsense codon, codon capture, variability of the genetic code, evolution of the genetic code

Introduction

The genetic code is essential to all forms of life and is of fundamental importance to the whole of biology. Until relatively recently, the code was thought to be invariable, frozen, in all organisms, because of the way in which any change would produce widespread alteration in the amino acid sequences of proteins. The universality of the genetic code was first challenged in 1979, when mammalian mitochondria were found to use a code that deviated somewhat from the “universal”.¹⁾ It was thought that the change in the code

happened to be tolerable in mitochondria, because of their small genome (see below).

In 1985, our research group in Nagoya University, Japan, found that a bacterium, *Mycoplasma capricolum*, used a deviant genetic code, namely that UGA, a universal stop codon, was read as Trp.²⁾ At about the same time, several workers announced that some ciliated protozoans used UAR (R = A or G) as Gln codons. At present, there are known considerable numbers of departures from the “universal” code in the nuclear as well as the mitochondrial codes (for refs., see Osawa *et al.*³⁾, Osawa⁴⁾) It is therefore misleading to think that “the genetic code is strikingly (or nearly) universal, but there exist some exceptions”. Such a description may be found in many text books. In reality, the genetic code is obviously not universal, and the deviant codes should not be treated as mere exceptions. Then the codon capture theory was proposed to account for the changes in codon meanings.^{5)–7)} The theory was based on experimental and theoretical studies conducted by us, in addition to the available data at that time. In short, the variations result from reassignment of codons, which takes place by disappearance of codon (unassigned codon) from coding sequences, followed by its reappearance in a new role. In other words, unassigned codons have the potential for reassignment. Simultaneously, an altered tRNA anticodon

^{*1} Kochi University of Technology, Department of Environmental System Engineering, 185 Miyanokuchi, Tosayamada-cho, Kaimi-Shi, Kochi 782-8502, Japan.

^{*2} University of Tsukuba, Center for Computational Sciences, Institute of Biological Sciences, Tsukuba, Ibaraki 305-8577, Japan.

^{*3} Genomic Sciences Center, Yokohama Institute, RIKEN, 1-7-22, Suehiro-cho, Tsurumi, Yokohama 230-0045, Japan.

^{*4} 1003, 2-4-7, Ushita-Asahi, Higashi-ku, Hiroshima 732-0067, Japan.

^{*5} Recipients of the Japan Academy Prize in 1992.

† Correspondence should be addressed: S. Osawa, 1003, 2-4-7, Ushita-Asahi, Higashi-ku, Hiroshima 732-0067, Japan (e-mail: osawasyozo@nifty.com).

Abbreviations: Phe: phenylalanine; Leu: leucine; Ile: isoleucine; Met: methionine; Val: valine; Ser: serine; Pro: proline; Thr: threonine; Ala: alanine; Tyr: tyrosine; His: histidine; Gln: glutamine; Asn: Asparagine; Lys: lysine; Asp: aspartic acid; Glu: glutamic acid; Cys: cysteine; Trp: tryptophan; Arg: arginine; Gly: glycine.

must also appear. The general model for code changes is that a codon disappears from coding sequences, typically as a result of directional (GC/AT-biased) mutation pressure. The codon reappeared, in many cases as a result of a change of directional mutation pressure, and acquired a new meaning. This can result from a change in an anticodon, or from a change in aminoacylation of a tRNA, or from a change in codon-anticodon pairing. All these changes are non-disruptive, because there is no change in the amino acid sequences of the proteins.

As mentioned above, the theory suggests that the present-day genetic codes, that are more or less specific to various organismic kingdoms, are derived with a series of non-disruptive changes from the code that was reached just before diversification of the common progenote of the present-day organisms.

In this article, we review the recent evidence for evolution of the genetic code mainly based on our own studies.

Frozen-Accident theory

Crick⁸⁾ proposed the frozen-accident theory as follows:

This theory states that the code is universal because at the present time any change would be lethal, or at least very strongly selected against. This is because in all organisms (with the possible exception of certain viruses) the code determines (by reading the mRNA) the amino acid sequences of so many highly evolved protein molecules that any change to these would be highly disadvantageous unless to correct the “mistakes” produced by altering the code. This accounts for the fact that the code does not change. To account for it being the same in all organisms, one must assume that all life evolved from a single organism (more strictly, from a single closely interbreeding population). In its extreme form, the theory implies that the allocation of codons to amino acids at this point was entirely a matter of “chance”.

This theory is based on the assumption that the primary nucleotide sequences (codon) are kept unchanged upon a change of the codon meaning, for example, changing codon AAA from Lys to Asn. This certainly results in widespread disruption of proteins by making unacceptable amino acid substitutions (Lys to Asn in this case) in all the

proteins. In other words, upon the code change, nucleotide sequences in the genes are kept unchanged, while amino acid sequences are subjected to intolerable alterations.

There is some evidence for freezing of the code in one respect: the same amino acids are in all codes, including mitochondrial and chloroplast codes. However, the known changes in the meaning of codons are difficult to reconcile with the frozen-accident theory as discussed below.

Distribution of the code changes

In 1979, it was reported that vertebrate mitochondria use AUA for Met and UGA for Trp, instead of Ile and stop, respectively, in the “universal code”.¹⁾ Indeed, as noted in Introduction, there are known considerable numbers of deviant mitochondrial codes in multicellular animals (see Yokobori *et al.*⁹⁾ for a review) as well as in unicellular eukaryotes.¹⁰⁾ To account for these changes, it was proposed that the mitochondrial genomes are much smaller (10 or so genes) than the nuclear genomes, and mitochondria can probably tolerate changes in the code that would be unacceptable to a larger and more complex system. However, the tolerance explanation for the mitochondrial code changes is no longer tenable, because, as mentioned above, it was discovered that the nuclear genome of *Mycoplasma capricolum* uses UGA as a Trp codon and in certain ciliated protozoans, UAA and UAG (= UAR) code for Gln. It is now known that 8 species of Mollicutes (eubacteria), including 7 species of *Mycoplasma* and one species *Spiroplasma*, use UGA as a Trp codon. CUG (Leu) is read as Ser in six species of *Candida* (yeasts), UAR (stop) is used for the codons of Gln in several species of ciliated protozoans (*Tetrahymena*, *Paramecium*, *Stylonicia*, *Oxytricus*) and UGA (stop) for Cys in a ciliate *Euplotes*. Two species of unicellular green alga (*Acetabularia*) also use UAR as Gln codons. For the references for the nuclear code changes until 1995, see Osawa.⁴⁾ More code changes have since been reported as follows: UGA as Trp in ciliates, *Colpoda inflata* and *Blepharisma americanum*¹¹⁾; UAA as Glu in three peritrich species, *Vorticella microstoma*, *Optisthonica henneguyi* and *O. matiensis*¹²⁾; UAR as Gln in a subgroup of diplomonads (e.g., *Hexamita inflata*),¹³⁾ and in the oxymonad *Streblomatrix*.¹⁴⁾ Since ciliates, the ulvophycean green algae,

diplomonads, and oxymonads are distantly related to each other, the use of UAR as Gln codons occurred independently in different phylogenetic lineages. Such a widespread occurrence of the deviant codes in nuclear as well as mitochondrial genomes clearly indicates that the genetic code is neither universal nor frozen. It is most likely that the deviant codes (including those in mitochondria) originated from what was used in a single progenote population for the present-day organisms.

Unassigned or nonsense codons

An unassigned or nonsense codon may be defined as a codon which does not exist in the genome of a certain organism, because of a lack of the corresponding tRNA or release factor. The unassigned codon may be produced typically by the extreme directional mutation pressure, which has been exerted on the entire genomic DNA toward A/T predominating over G/C (AT-pressure) or toward G/C over A/T (GC-pressure),¹⁵⁾ so that the genomic G+C content is more or less, and sometimes extremely, uneven among various organisms (e.g., see Muto and Osawa).¹⁶⁾

The genomic A+T content of *Mycoplasma capricolum* is a little more than 75%, the highest of all organisms that have been examined. Here, codons NNU and NNA (N = U, C, A, or G) greatly predominate over codons NNC and NNG. As illustrated in Fig. 1, there is a strong correlation between the codon usage and the relative amount of isoacceptor tRNAs.¹⁷⁾ As tRNAs having anticodon UNN (and Thr ACU)¹⁸⁾ in family boxes can translate most of the four codons by four-way wobbling as in mammalian mitochondria,¹⁹⁾ these are not included in the figure. Codon CGG, an Arg codon does not exist among more than 5,800 codons in *Mycoplasma capricolum*. Also undetected is the tRNA with anticodon CCG for codon CGG. It is probable that, under extremely strong AT-pressure, codon CGG was totally converted to its synonymous codons CGU, CGA or AGR, which are read by anticodons ICG (I = inosine; A on DNA) or *UCU [in this case, *U = 5-carboxymethylamino-methyl U (cmnm⁵U); see below]. As a result, tRNA Arg CCG has been removed from the genome along with disappearance of codon CGG. Thus, CGG has become an unassigned codon, which therefore does not exist in the genome of *Mycoplasma capricolum*. The idea of CGG being an unassigned codon is

strongly supported in the following experiment.²⁰⁾ In a cell-free system prepared from *Mycoplasma capricolum*, translation of synthetic mRNA containing in-frame CGG codons (see top of Fig. 2), does not result in 'read-through' to codons (Tyr) beyond CGG, i.e., translation ceases just before CGG. The synthesized peptide is tightly attached to 70S ribosomes and is released upon further incubation with puromycin. Thus the peptide is in the P-site of the ribosome in the form of peptidyl-tRNA, leaving the A-site empty (Fig. 2b). When UAA stop codons are used instead of CGG, no read-through occurs beyond UAA, but the synthesized peptide is released from 70S ribosomes (Fig. 2c and d), presumably by release factor 1 (RF-1).

Contrary to *Mycoplasma capricolum*, the genomic GC content of *Micrococcus luteus* is 75%, the highest of all the organisms known to date. The codons ending with G or C comprise 95 to 100% of all codons.²¹⁾ Fig. 3 shows a strong correlation between synonymous codon usage and the amount of the corresponding tRNAs (anticodon) in *Micrococcus luteus*.²²⁾ Generally speaking, a large amount of tRNAs with anticodons GNN and CNN translate the abundantly existing NNC and NNG codons, where as the amount of anticodon *UNN[*U = 5-methyl-2-thiouridine (xm⁵s²U; X = -CH₃, -CH₃COO⁻, etc) or an equivalent modified U^{22a)} is very small or not detectable in accordance with disuse of NNA codons. A small amount of NNU codons is utilized, because anticodon GNN pairs strongly with NNC and only weakly with NNU by wobble. It is then highly probable that some of the NNA are unassigned codons. In fact, *in vitro* experiments, similar to those of *Mycoplasma capricolum*, show that in-frame AGA and AUA codons do not result in 'read-through' to codon AGA or AUA, and that the synthesized peptides are attached to 70S ribosomes. The in-frame stop codon UGA causes the release of peptides from the ribosomes.²³⁾ Thus, at least AGA and AUA, and perhaps some other NNA codons (not tested in the *in vitro* system) would be unassigned codons.

The mitochondrial genome of the yeast *Torulopsis glabrata* is AT-rich and neither CGN nor the corresponding tRNA has been found.²⁴⁾ Also in mitochondrial genome of the chlorophyte alga *Prototheca wickerhamii* is AT-rich and codon UAG, UGA and CGG are not used at all, with a lack of tRNA for Arg CGG codon, indicating that CGG

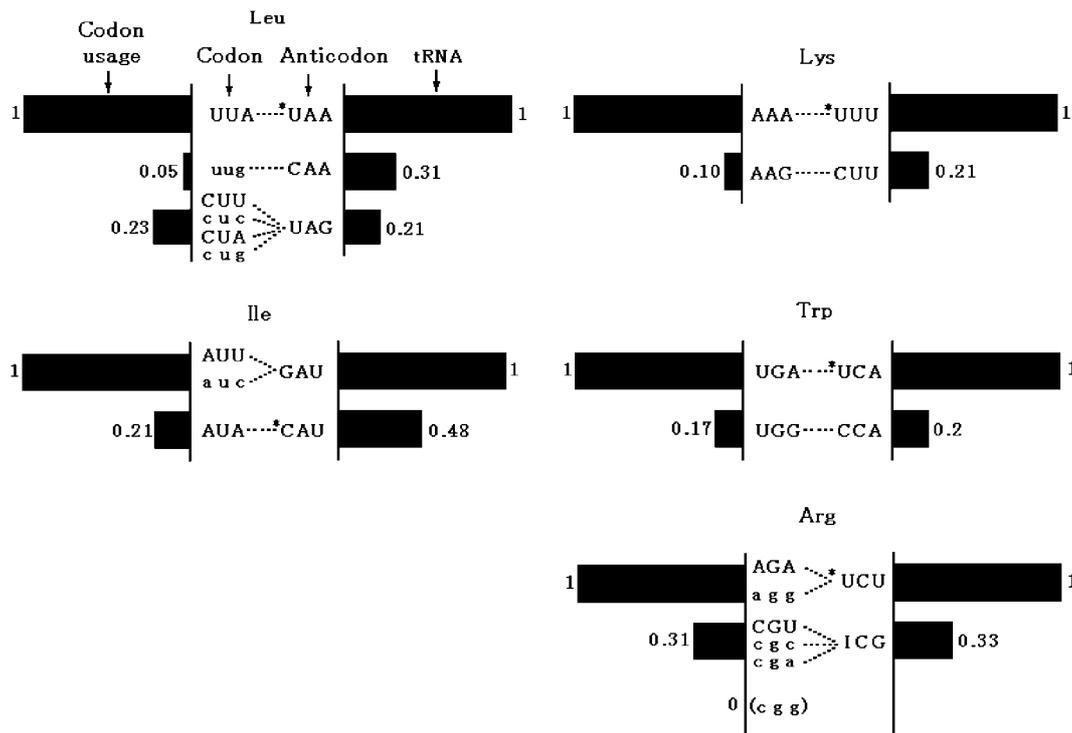


Fig. 1. Correlation between codon usage and the relative amount of isoacceptor tRNAs in *Mycoplasma capricolum*.

The most abundant codons and anticodons were expressed as 1. Codons in capital letters and lower letters show more than and less than 25% usage in the synonymous codons, respectively. Codon-anticodon pairing is shown by dotted lines. The total 5,814 codons were examined. In this figure, however, 4,452 codons were used, because the rest 1,362 codons are in family boxes (see the text) and Met AUG.

is an unassigned codon. There are more examples of unassigned codons in mitochondrial genomes, such as AAA in a species of hemichordate mitochondria,²⁵⁾ etc. (see also Yokobori *et al.*).⁹⁾

Although the codon usage and the amount of the responsible tRNA run in parallel, this does not mean that the directional mutation pressure is the sole factor to govern this phenomenon. At least in eubacteria, the codon usage is indeed influenced strongly by the directional mutation pressure, but is also biased to a considerable extent by tRNA constraint, because the codon usage does not go exactly in parallel with the G+C content of the spacer regions that would most probably be free from the tRNA-constraint.⁵⁾ It is then reasonable to assume that in *Mycoplasma capricolum* and *Micrococcus luteus*, unassigned codons would have been produced by the exertion of extremely high AT- or GC-pressure primarily on codons and are biased by tRNA to a considerable extent. However, it is unlikely that when a certain codon becomes unas-

signed, the corresponding tRNA disappears first, followed by the removal of the codon, because, without the tRNA, the codon would be untranslatable and therefore unable to remain in the genes.

Contrary to the generally accepted view, in that all organisms including mitochondria and chloroplasts use the familiar genetic code table consisting of 64 codons including three stop codons, the occurrence of unassigned codons implies that some life forms use fewer than 64 codons. In addition, it is worthwhile to note that stop codons are often referred to as nonsense codons, but the stop codons are not nonsensical; they function to release the peptides from ribosomes by interacting with release factors (RFs), whereas the nonsense or unassigned codon does not appear in the genome. Even if the nonsense codon appears by chance, it is immediately selected against, because it has neither function as a stop codon because of inability to interact with RF, nor is it translated to any amino acid due to the lack of the corresponding tRNA.

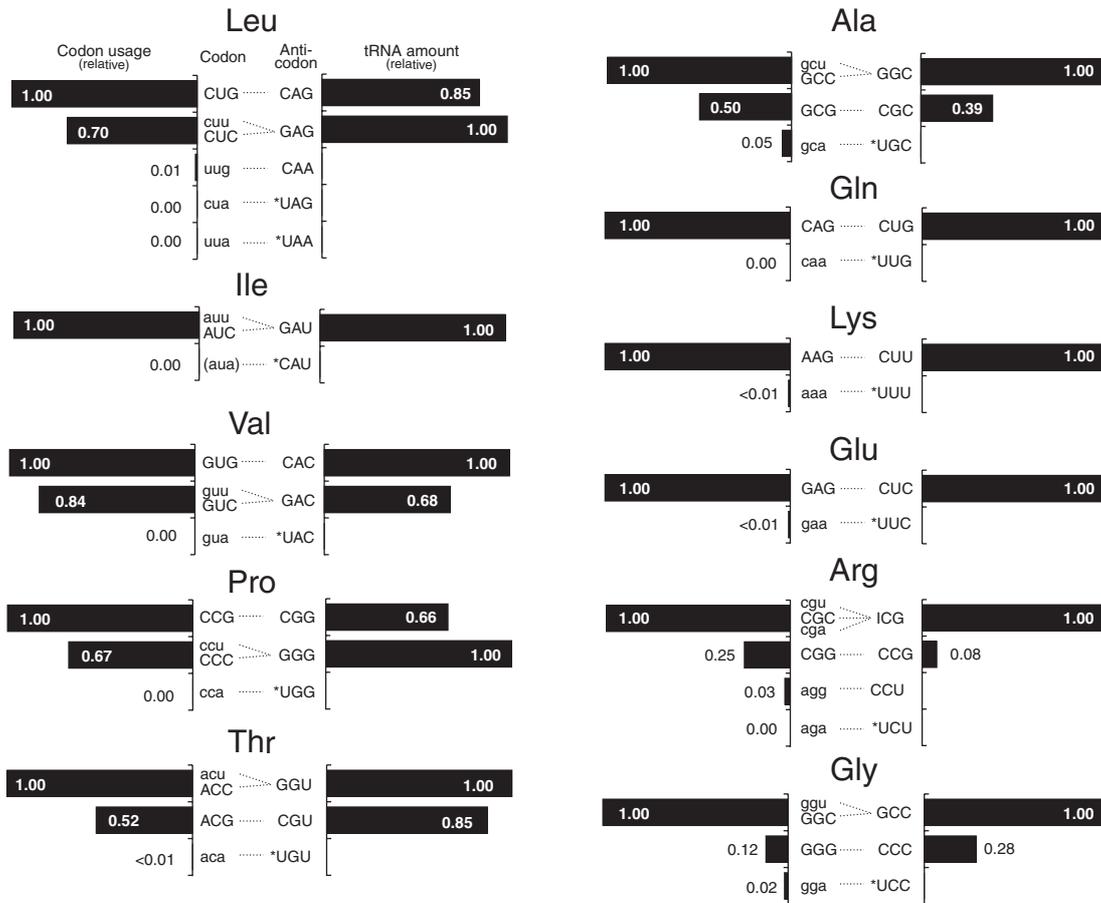


Fig. 3. Correlation between codon usage (the total of 5,516 codons) and the relative amount of isoacceptor tRNAs in *Micrococcus luteus*.

For explanation, see the legend of Fig. 1.

Table 1. Changes in codons while amino acid sequences remains constant

Original sequence	Amino acid	Asn ... Lys ... Lys ... Asn ... Lys ... Asn ... Lys ... Asn
	Codon	AAC ... AAG ... AAA ... AAC ... AAG ... AAU ... AAA ... AAU
	Anticodon	GUU ... CUU ... *UUU ... GUU ... CUU ... GUU ... *UUU ... GUU
Disappearance of AAA codon and *UUU anticodon	Amino acid	Asn ... Lys ... Lys ... Asn ... Lys ... Asn ... Lys ... Asn
	Codon	AAC ... AAG ... AAG ... AAC ... AAG ... AAU ... AAG ... AAU
	Anticodon	GUU ... CUU ... CUU ... GUU ... CUU ... GUU ... CUU ... GUU
Anticodon change and reappearance of AAA codons	Amino acid	Asn ... Lys ... Lys ... Asn ... Lys ... Asn ... Lys ... Asn
	Codon	AAA ... AAG ... AAG ... AAC ... AAG ... AAA ... AAG ... AAU
	Anticodon	G ψ U ... CUU ... CUU ... G ψ U ... CUU ... G ψ U ... CUU ... G ψ U

codon GUU cannot read codon AAA. The newly appeared anticodon was shown to be G ψ U(ψ = pseudouridine) which can pair with AAA.²⁹⁾ Here, AAA is at the site of Asn and not at the site of Lys, indicating this code change does not result in

alteration of amino acid sequence of the protein (Table 1). The scheme in Table 1 was verified by Castresana *et al.*²⁵⁾ and showed that the AAA lysine codon becomes unassigned in the common mitochondrial ancestry of hemichordate and echinoderm

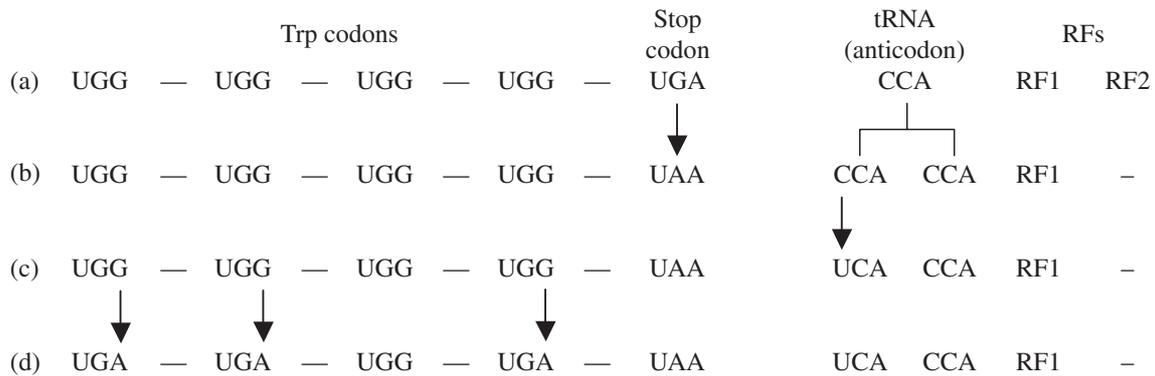


Fig. 4. Evolution of the UGA Trp codon in *Mycoplasma capricolum*.

Stages (a) and (d) are represented by *Acholeplasma laidlawi* and *Mycoplasma capricolum*, respectively. Trp UCA anticodon in the figure is *UCA = cmm³Um.³³⁾

and then reassigned as Asn in the echinoderm, leaving AAA as an assigned codon in the hemichordate. They concluded that the mitochondrial genome of *Balanoglossus carnosus* (a hemichordate species) provides a remarkable fulfillment of the predictions of the codon capture hypothesis for codon reassignment as proposed by Osawa and Jukes.⁵⁾ On the other hand, if the newly evolved tRNA anticodon is evolutionarily equivalent to the original one, AAA is revived from AAG as a lysine codon at the Lys site. A similar event may be observed in the case for CGG in *Mycoplasma* spp. The genomic G+C content of *Mycoplasma capricolum* is 24% and the tRNA translating the CGG codon is lacking (see section "Unassigned codon"), while those of *Mycoplasma genitalium* and *Mycoplasma pneumoniae* are 32 and 41%, respectively, and both have the tRNA for the CGG codon.³⁰⁾ This suggests that under weakened AT-pressure the CGG codon once unassigned may be reassigned by the appearance of the tRNA translating CGG.

The bacterial class Mollicutes includes each several species of *Acholeplasma*, *Mycoplasma*, and *Spiroplasma*. *Acholeplasma*, which shares the common ancestry with *Mycoplasma* and *Spiroplasma*, uses the standard genetic code in which only the Trp codon is UGG, while UAA, UAG and UGA are stop codons.³¹⁾ Since RF-1 interacts with UAA or UAG, and RF-2 with UAA or UGA to release the synthesized peptides from the ribosomes, it may be assumed that the *Acholeplasma* Trp codon UGG is translated with a tRNA containing the anticodon CCA, whereas UAG and/or UAA are used as stop codons, which are recognized by RF-1 (Fig. 4a).

In *Mycoplasma capricolum*, there exist a number of in-frame UGA codons at the Trp sites, but much fewer numbers of UGG codons were found. UGA does not appear at the termination site, and the only stop codons are UAA and UAG, indicating that UGA is a Trp codon in this bacterium. Indeed, the in-frame UGA (and UGG) codons in the synthetic mRNA were translated in the cell-free system of *Mycoplasma capricolum*, whereas only UGG was translated as Trp in a similar system of *Escherichia coli*.²⁰⁾ The sequence of events leading to UGA becoming a Trp codon may have proceeded as shown in Fig. 4. Presumably, under strong AT-pressure, in the ancestor of *Mycoplasma capricolum* (Fig. 4b), the UGA stop codon was totally converted to UAA and at about the same time, RF-2, which recognizes the UGA stop codon, disappeared. Thus UGA became an unassigned codon. At this stage, however, UGG was still the only one Trp codon, because the one species of Trp-tRNA with anticodon CCA could not translate UGA. On the genome of *Mycoplasma capricolum*, there exist two tRNA genes that arranged in tandem, one with anticodon TCA and another with CCA³²⁾ (Fig. 5). The transcripts of these genes (tRNAs with anticodons *UCA and CCA) were also detected.³³⁾ These facts suggest that the tRNA with anticodon CCA was duplicated (Fig. 4b) and one copy mutated to a Trp-tRNA with anticodon TCA (Fig. 4c).

Also, the deletion of RF-2 must have occurred between the stage (a) and (b) in Fig. 4. If RF-2 activity remains intact, the in-frame UGA would be recognized both by the RF-2 and the Trp-tRNA

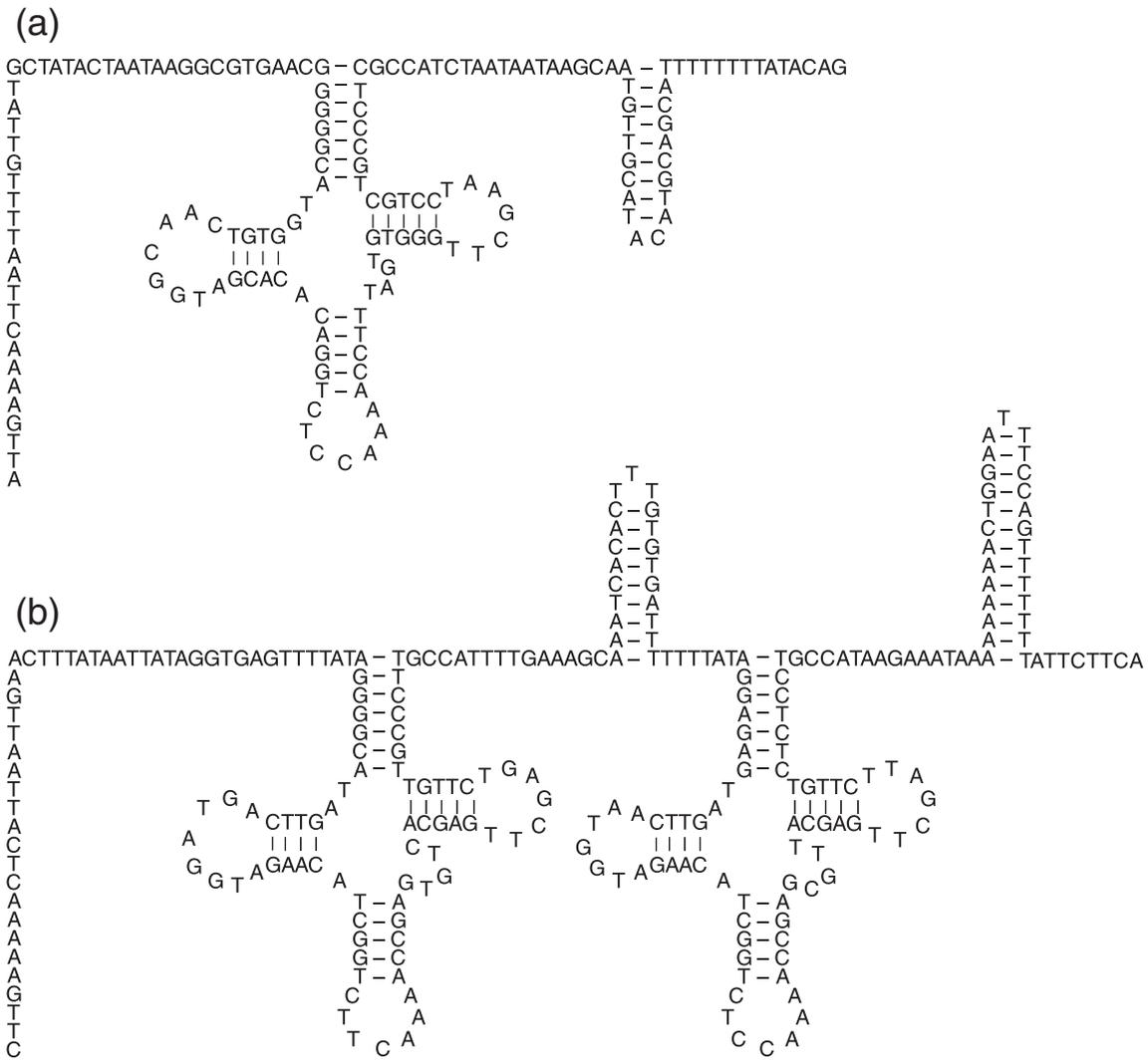


Fig. 5. tRNA Trp genes from *Acholeplasma laidlawii* (a) and *Mycoplasma capricolum* (b).

*UCA. This would be disadvantageous, because it would result in production of both truncated and complete peptides. To prove that the RF-2 is absent, Inagaki *et al.*³⁴⁾ constructed a cell-free translation system using synthetic mRNA (mRNA-UGA in Fig. 6a) in conjunction with the dialyzed S30 or the S100 fraction. The synthesized peptide, in the presence of unlabelled Met and [³H]Ile and in the absence of Trp, was not released from ribosome (Fig. 6d), because of the absence of tryptophanyl-tRNA and RF (presumably RF-2) for UGA. In contrast, when mRNA-UAA or mRNA-UAG plus the *Mycoplasma capricolum* S100 fraction was used, the synthesized peptide

was released from the ribosome (Fig. 6b and c). These experiments indicate that in *Mycoplasma capricolum*, there exists RF-1, whereas RF-2 is lacking or inactive. In fact, when the S100 fraction from *Escherichia coli* or *Bacillus subtilis*, which contains RF-2, is added to the above mRNA-UGA system, the synthesized peptide is released from the ribosomes (Fig. 6e). The lack of RF-2 was recently confirmed by total genome analysis (Glass *et al.*, 2007 GenBank acc. no. NC_007633.1). The gene for RF-1 in *Mycoplasma capricolum*, which recognizes stop codon UAA, was cloned and sequenced.³⁵⁾

This type of codon reassignment is called “stop codon capture”, because the former stop codon has

ent-day organisms, so far examined, use so-called “universal” genetic code. It should however be pointed out that these are mainly composed of what are called “model” organisms, comprising only a fraction of more than thirty million known species. There exists no evidence that the “universal code” was used in a single progenote population before diversification of the present organismic lines.

The theories to explain the early evolution of the genetic code are numerous, all of which include speculations that the coding system arose with one or a limited number of amino acids, and that others were added until a total of 20 was reached. Most of these theories are aesthetically pleasing but cannot be verified. We are going to discuss those with some experimental support, starting at the time where for protein synthesis the progenote used 20 amino acids (except selenocysteine and pyrrolysine; see below). It is reasonable to assume that in this stage, translation of codons to 20 amino acids was performed more simply than in the present highly evolved code, using a minimum number of codons and tRNA species. Nevertheless, since amino acid sequences of many of the essential and well-refined protein molecules (e.g., aminoacyl tRNA synthetases, ribosomal proteins, DNA- and RNA-polymerases, etc.) would have been required for establishment of the progenote, introduction of any new amino acids would not be allowed and the amino acid assignment of codons would have been established in a way not to affect the functionally essential sequences of proteins. In other words, the code was frozen in the respect that the same 20 amino acids are in all codes. Briefly, what were frozen were not the 64 codons of the “universal” code. The number of codons and the isoacceptor tRNAs have increased during evolution with increasing complexity of the genome, so that the fidelity, efficiency and other regulation by codon-anticodon pairing have improved in various ways.

To assume the most ancient code, the following two conditions may be taken into account. First, the early code contained a minimum number of codons for 20 amino acids. Second, there was also a minimum number of tRNA species responsible for the translation of the codons. Anticodons of these tRNAs would most probably be unmodified, because the modification of the first (sometimes second) anticodon position differs to a considerable extent in various lines of organisms, suggesting

later development of the modification systems. The code table, which fulfills these requirements, is shown in Table 2, from which it may be seen that there exists just 20 species of amino acid codons plus 1 stop codon. In this code, tRNA anticodon starting from C or G (unmodified), and codon ending with G or C were used. Thus, it may be assumed that the genome of the progenote with this ancient genetic code was rich in G+C content.

The codon usage and tRNA composition of the *Micrococcus luteus* genome, which has among the highest known G+C content, provide an important hint for deducing the ancient code. As briefly discussed in the section “Unassigned codon,” the use of codons NNA (AUA Ile, CUA Leu, UUA Leu, GUA Val, ACA Thr, GCA Ala, CAA Gln, AAA Lys, GAA Glu, CGA Arg, AGA Arg or GGA Gly) in this bacterium is null or less than 1% among the synonymous codons, and the tRNA responsible for decoding these codons could not be detected. Also, in contrast to the abundant use of NNC codons, a much lesser use of NNU codons is observed, because anticodon GNN pairs mainly NNC codon and, with a lesser affinity, NNU codons. If the *Micrococcus luteus* genetic code proceeds to its extreme, then the ancient genetic code discussed above will result. This genetic code is non-degenerate, so that there is no flexibility against mutations. A single mutation in a gene produces either a change in amino acid assignment, or more seriously a nonsense codon that inactivates the whole gene in most cases, so that the mutational load is quite high, and therefore evolution of the genetic code is virtually impossible. However, flexibility of the code is observed in different organisms. For the most part, they use the degenerate code in which many mutations on codons are tolerated by converting them to their synonymous codons without changing the amino acid sequence of the proteins. Therefore, from an evolutionary point of view, later development of a number of synonymous codons would have been advantageous (see the next section).

The above discussion does not mean that *Micrococcus luteus* is the most ancient organism lacking flexibility against mutations. Rather, the genetic code of this bacterium would largely represent retrogression to the ancient code, but has flexibility for the conversion of GC-rich codons to synonymous AT-rich codons when GC-pressure is weakened.

Table 2. Ancient genetic code

Amino acid (codon)	Anticodon						
Phe (UUC)	GAA			Tyr (UAC)	GUA	Cys (UGC)	GCA
		Ser (UCG)	CGA	Stop (UAG)	—	Trp (UGG)	CCA
				His (CAC)	GUG		
Leu (CUG)	CAG	Pro (CCG)	CGG	Gln (CAG)	CUG	Arg (CGG)	CCG
Ile (AUC)	GAU			Asn (AAC)	GUU		
Met (AUG)	CAU	Thr (ACG)	CGU	Lys (AAG)	CUU		
				Asp (GAC)	GUC		
Val (GUG)	CAC	Ala (GCG)	CGC	Glu (GAG)	CUC	Gly (GGG)	CCC

Anticodon GNN and the corresponding codons might exist in family boxes, which are however omitted from the table for simplicity.

Table 3. Archetypal genetic code of Jukes⁴¹⁾

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

Another view of the ancient code (archetypal code; Table 3) was proposed by Jukes^{39–41)} and is mainly based on the fact that unmodified U at the first position of the anticodon pairs with U, C, A and G of the 3rd position of the codon, as observed in the mammalian mitochondrial code (this is also the case in *Mycoplasma capricolum*). The code of Jukes consists of 15 family boxes, which are occupied by a single amino acid translated by a single tRNA with a UNN anticodon. In addition to

these family boxes, there are two 2-codon sets; one is for two stop codons UAR (R=A or G), and the other is for UAY (Tyr; Y=U or C), which was read by the anticodon GUA. Altogether only 16 amino acids could be used for protein synthesis according to this code. Therefore, subsequent expansion is required until a code for the 20 amino acids is reached. This scheme allows changes in the amino acid sequences of pre-existing proteins (what might be called “primitive” proteins) upon addition of new

Table 4. Early genetic code

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

amino acids. Jukes⁴¹⁾ noted, “If the organism could survive this change, acquisition of the new amino acids in the genetic code might provide for an evolutionary advantage”. However, this interesting idea presumes the presence of the “primitive” proteins. Supporting evidence does not exist to determine whether the organisms at that time could survive with such proteins.

From the ancient genetic code to the early code

Here, the early genetic code is defined as the code that existed in the common ancestry shortly before the “universal” code was established. The “universal” code, which is used by many present-day organisms, will be called hereafter the “standard code”, instead of the “universal” code, because the presently used code is not universal. The structure of the early code is similar to the standard code (compare Table 4 with Table 5). In the code of many mitochondria AUA and AUG are codons for Met and UGA and UGG are both Trp codons. The situation for Trp codons is similar to that observed in *Mycoplasma capricolum*. In both mitochondria from many organisms and *Mycoplasma* spp., four codons in each family box are read by a single anticodon UNN by four-way wobbling. If we postulate these facts as representing partial retro-

gression to the common ancestor (the progenote), evolution to the early code from the ancient code should have proceeded, under AT-pressure, to develop the A and U-ending codons as well as the tRNAs with anticodons enabling the translation of these codons. In a given family box, a tRNA with anticodon CNN duplicated and one copy mutated to UNN thereby enabling all codons in family box to be read. This was followed by the disappearance of CNN (and GNN if it existed). In a two-codon set, the CNN anticodon duplicated and one copy mutated to UNN with simultaneous appearance of the U-modification system, e.g., to modify U to *U (e.g., derivatives of 5-methyl-2-thiouridine; see above). In Table 4, the CNN anticodon is omitted from all the two-codon sets for simplicity; some of them might have existed. This early code is the same as what Jukes proposed in 1983.⁴¹⁾ This early code has considerable flexibility and much less genetic load against mutation as compared with the ancient code.

Standard genetic code

As compared with the early code, the standard code consists of 13 two-codon sets and 8 family boxes, in addition to 3 codons for Ile, and only 1 codon for Trp and Met (Table 5). Thus, the main problems to be solved in the route from the early

Table 5. Standard genetic code

Amino acid (codon)	Anticodon Euk.	<i>E. coli</i>	Amino acid (codon)	Anticodon Euk.	<i>E. coli</i>	Amino acid (codon)	Anticodon Euk.	<i>E. coli</i>	Amino acid (codon)	Anticodon Euk.	<i>E. coli</i>
Phe (UUU)	GAA	GAA	Ser (UCU)		GGA	Tyr (UAU)	GUA	GUA	Cys (UGU)	GCA	GCA
Phe (UUC)			Ser (UCC)	IGA		Tyr (UAC)			Cys (UGC)		
Leu (UUA)	*UAA	*UAA	Ser (UCA)			Stop (UAA)	—	—	Stop (UGA)	—	—
Leu (UUG)	CAA	CAA	Ser (UCG)	CGA	CGA	Stop (UAG)	—	—	Trp (UGG)	CCA	CCA
Leu (CUU)		GAG	Pro (CCU)		GGG	His (CAU)	GUG	GUG	Arg (CGU)		
Leu (CUC)	IAG		Pro (CCC)	IGG		His (CAC)			Arg (CGC)	ICG	ICG
Leu (CUA)	*UAG	*UAG	Pro (CCA)		*UGG	Gln (CAA)	*UUG	*UUG	Arg (CGA)		
Leu (CUG)	CAG	CAG	Pro (CCG)	CGG	CGG	Gln (CAG)	CUG	CUG	Arg (CGG)	CCG	CCG
Ile (AUU)		GAU	Thr (ACU)		GGU	Asn (AAU)	GUU	GUU	Ser (AGU)	GCU	GCU
Ile (AUC)	IAU		Thr (ACC)	IGU		Asn (AAC)			Ser (AGC)		
Ile (AUA)	*UAU	*CAU	Thr (ACA)		*UGU	Lys (AAA)	*UUU	*UUU	Arg (AGA)	*UCU	*UCU
Met (AUG)	CAU	CAU	Thr (ACG)	CGU	CGU	Lys (AAG)	CUU		Arg (AGG)	CCU	CCU
Val (GUU)		GAC	Ala (GCU)		GGC	Asp (GAU)	GUC	GUC	Gly (GGU)	GCC	GCC
Val (GUC)	IAC		Ala (GCC)	IGC		Asp (GAC)			Gly (GGC)		
Val (GUA)		*UAC	Ala (GCA)		*UGC	Glu (GAA)	*UUC	*UUC	Gly (GGA)	*UCC	*UCC
Val (GUG)	CAC		Ala (GCG)	CGC		Glu (GAG)	CUC		Gly (GGG)	CCC	CCC

Euk. = eukaryotes (representative). *E. coli* = *Escherichia coli*. In some eubacteria and eukaryotes, anticodons *UNN and *GNN (*G = queosine) are present in family boxes and two-codon sets, respectively. These are not shown in the table for simplicity.

code to the standard code are: (1) reduction of UGA+UGG Trp codons to a single UGG, (2) simplification of AUA+AUG Met codons to a single AUG codon, and (3) reassignment of AUA to an Ile codon and of UGA to a stop codon. These changes must have occurred in the progenote before diversification of the present-day organismic lines, because the codon composition of the standard code is basically the same in various lines of organisms with the occasional appearance of changed codons in mitochondria, *Mycoplasma* and some other organisms. The transition described (1) to (2) may be solved by assuming that UGA and AUA were unassigned presumably by strong GC-pressure just in the case of *Micrococcus luteus*. The reassignment of AUA would have been occurred, presumably under AT-pressure, by mutation of the AUY Ile codon to AUA upon appearance of the Ile-tRNA *CAU (*C = 2-lysyl C; lysidine^{41a}) in eubacteria and plant mitochondria, and of Ile-tRNA IAU in eukaryotes. The genes for the Met-tRNA CAU and the Ile-tRNA *CAU are adjacent on the chromosomes of three species of bacteria so far examined. This fact favors the ancient duplication of the gene for Met-tRNA CAU and the subsequent conversion of one of them to Ile-tRNA *CAU. The reassign-

ment of AUA in bacteria and eukaryotes may have occurred independently, because the mechanism for codon capture is different between them as mentioned above. Introduction of the UGA stop codon may also be explained by assuming that the common progenote was under GC-pressure; Trp-tRNA *UCA in the early code may have become unable to accommodate the increasing number of Trp UGG codons. Thus, the tRNA *UCA duplicated and one of them mutated to CCA which only translates UGG. The persistent GC-pressure would have removed the tRNA *UCA and Trp UGA codons. After replacement of anticodon UCA by CCA, some UAA stop codons would have mutated to UGA, which for the first time became a stop codon. At this stage, RF-2 would have emerged so as to recognize both UAA and UGA as stop codons. In closing this section, it should be emphasized that the standard genetic code is used in many organisms, and yet the codon-anticodon pairing patterns differ considerably among various groups of organisms, mitochondria and chloroplasts. For example, codons in most family boxes are translated by anticodons INN and CNN in many eukaryotes, while those are read by anticodons GNN, *UNN and/or CNN in eubacteria such as in *Escherichia*

coli (see Table 5). A single anticodon UNN is responsible for reading most of the family box codons in mitochondria and some in chloroplasts and *Mycoplasma capricolum*. Thus, evolution of the genetic code has proceeded not only in the amino acid assignment of codon, but also in the codon-anticodon pairing pattern.

Flexibility of the standard genetic code

As described in the preceding section, the standard genetic code is utilized in many organisms, although not a few organisms and mitochondria use non-standard code. The deviated code may be roughly divided into two categories, whereas all of them may be explained by the codon capture theory. One may be considered as a partial retrogression to the early or ancient code as exemplified by the code change of UGA from stop to Trp codon in many mitochondria and *Mycoplasma* spp., and AUA from Ile to Met in most of the mitochondrial species. Generation of unassigned codons, such as in *Micrococcus luteus* and others would also belong to this category. Another change would have happened by chance as exemplified by the transition of AAA from Lys to Asn in echinoderm mitochondria, UAR from stop to Gln in some ciliated protozoans and *Acetabularia*, and UGA from stop to Cys in *Euplotes*, and so on (see above sections).

Still another category of apparent code change may be noteworthy. Until now, two examples have been reported, in which stop codons (UGA and UAG) are utilized as alternate codes in the genome of the same organism, one as a stop and the second as another amino acid. Codon UGA is read as selenocysteine (SeCys) when a special hairpin-loop exists next to the in-frame UGA.⁴²⁾ Alternately, UGA at the termination site is used as a stop codon. Such a double use of the same codon UGA as SeCys and stop is widely observed in a broad range of prokaryotes and eukaryotes.⁴³⁾ It follows that this system would have been formed shortly after the standard code was established. This system differs from the post-translational modification of an amino acid after translation, because the SeCys tRNA inserts SeCys during translation (see Osawa, pp. 116–125).⁴⁾ A similar double use of the UAG codon within special context is described in meta-bacteria (= archaeobacteria), in which the UAG codon is used for pyrrolysine and stop.⁴⁴⁾ Such a special system for cooping the codon for a highly

specialized function would have emerged for the production of a very limited species of enzymes. The genetic code system has a capacity to assign more than 20 amino acids, and yet in principle only 20 amino acids are utilized for protein synthesis. However, when a certain need arises that requires the use less common amino acids, organisms can develop an optional system to use a single codon for two different purposes. It should be stressed however that there are no organisms which use the genetic code system for more than, or less than, 20 amino acids. What were frozen are 20 amino acids (magic 20!) and not the genetic code that assigns them. Thus the genetic code is still in the state of evolution.

Acknowledgements

We express our wholehearted appreciation to all our coworkers, whose names are given in the text and in "References". Cordial thanks are also due to Drs. Susumu Nishimura, Shigeyuki Yokoyama and Yoshiyuki Kuchino for their valuable suggestions and help during the course of our study.

References

- 1) Barrell, B.G., Bankier, A.T. and Drouin, J. (1979) A different genetic code in human mitochondria. *Nature* **282**, 189–194.
- 2) Yamao, F., Muto, A., Kawauchi, Y., Iwami, M., Iwagami, S., Azumi, Y. and Osawa, S. (1985) UGA is read as tryptophan in *Mycoplasma capricolum*. *Proc. Natl. Acad. Sci. USA* **82**, 2306–2309.
- 3) Osawa, S., Jukes, T.H., Watanabe, K. and Muto, A. (1992) Recent evidence for evolution of the genetic code. *Microbiol. Rev.* **59**, 229–264.
- 4) Osawa, S. (1995) *Evolution of the Genetic Code*. Oxford Univ. Press, Oxford.
- 5) Osawa, S. and Jukes, T.H. (1989) Codon reassignment (codon capture) in evolution. *J. Mol. Evol.* **28**, 271–278.
- 5a) Osawa, S. and Jukes, T.H. (1988) Evolution of the genetic code as affected by anticodon content. *Trends Genet.* **4**, 271–278.
- 6) Osawa, S., Muto, A., Ohama, T., Andachi, R., Tanaka, R. and Yamao, F. (1990) Prokaryotic genetic code. *Experientia* **46**, 1097–1106.
- 7) Osawa, S., Muto, A., Jukes, T.H. and Ohama, T. (1990) Evolutionary changes in the genetic code. *Proc. R. Soc. London, Ser. B* **241**, 19–28.
- 8) Crick, F.H.C. (1968) The origin of the genetic code. *J. Mol. Biol.* **38**, 367–379.
- 9) Yokobori, S., Suzuki, T. and Watanabe, K. (2001) Genetic code: variation in mitochondria: tRNA as a major determinant of genetic code plasticity.

- J. Mol. Evol. **53**, 314–326.
- 10) Inagaki, Y., Ehara, M., Watanabe, K.I., Hayashi-Ishimaru, Y. and Ohama T. (1998) Directionally evolving genetic code: The UGA codon from stop to tryptophan in mitochondria. *J. Mol. Evol.* **47**, 378–384.
 - 11) Lozupone, C.A., Knight, R.D. and Landweber, L.F. (2001) The molecular basis of nuclear genetic code change in ciliates. *Curr. Biol.* **11**, 65–74.
 - 12) Sánchez-Silva, R., Villalobo, E., Morin, L. and Torres, A. (2003) A new noncanonical nuclear genetic code: Translation of UAA into glutamate. *Curr. Biol.* **13**, 442–447.
 - 13) Keeling, P.J. and Doolittle, W.F. (1996) A non-canonical genetic code in an early diverging eukaryotic lineage. *EMBO J.* **15**, 2285–2290.
 - 14) Keeling, P.J. and Leander, B.S. (2003) Characterization of a non-canonical genetic code in the oxymonad *Streblomastix strix*. *J. Mol. Biol.* **326**, 1337–1349.
 - 15) Sueoka, N. (1988) Directional mutation pressure and neutral molecular evolution. *Proc. Natl. Acad. Sci. USA* **85**, 2653–2657.
 - 16) Muto, A. and Osawa, S. (1987) Guanine and cytosine content of genomic DNA and bacterial evolution. *Proc. Natl. Acad. Sci. USA* **84**, 166–169.
 - 17) Yamao, F., Andachi, Y., Muto, A., Ikemura, T. and Osawa, S. (1991) Levels of tRNAs in bacterial cells as affected by amino acid usage in proteins. *Nucleic Acids Res.* **19**, 6119–6122.
 - 18) Andachi, Y., Yamao, F., Iwami, M., Muto, A. and Osawa, S. (1987) Occurrence of unmodified adenine and uracil at the first position of anticodon in threonine tRNAs in *Mycoplasma capricolum*. *Proc. Natl. Acad. Sci. USA* **84**, 7398–7402.
 - 19) Inagaki, Y., Kojima, A., Bessho, Y., Hori, H., Ohama, T. and Osawa, S. (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.
 - 20) Oba, T., Andachi, Y., Muto, A. and Osawa, S. (1991) CGG, unassigned or nonsense codon: Occurrence in *Mycoplasma capricolum*. *Proc. Natl. Acad. Sci. USA* **88**, 921–925.
 - 21) Ohama, T., Yamao, F., Muto, A. and Osawa, S. (1987) Organization and codon usage of the streptomycin operon in *Micrococcus luteus*, a bacterium with a high genomic G+C content. *J. Bacteriol.* **169**, 4770–4777.
 - 22) Kano, A., Andachi, Y., Ohama, T. and Osawa, S. (1991) Novel anticodon composition of transfer RNAs in *Micrococcus luteus*, a bacterium with a high genomic G+C-content: correlation with codon usage. *J. Mol. Biol.* **221**, 387–401.
 - 22a) Watanabe, K. and Osawa, S. (1995) tRNA sequences and variations in the genetic code. *In* tRNA: Structure, Biosynthesis, and Function (eds. Söll, D. and RajiBandary, U.). American Society for Microbiology, Washington, D.C., pp. 215–250.
 - 23) Kano, A., Ohama, T., Abe, R. and Osawa, S. (1993) Unassigned or nonsense codons in *Micrococcus luteus*. *J. Mol. Biol.* **230**, 51–56.
 - 24) Clark-Walker, G.D., McArthur, C.R. and Spriprakash, K. (1985) Location of transcriptional control signals and transfer RNA sequence in *Torulopsis glabrata* mitochondrial DNA. *EMBO J.* **4**, 465–473.
 - 25) Castresana, J., Feldmaier-Fuchs, G. and Pääbo, S. (1988) Codon reassignment and amino acid composition in hemichordate mitochondria. *Proc. Natl. Acad. Sci. USA* **95**, 3703–3707.
 - 26) Jukes, T.H. and Osawa, S. (1991) Recent evidence for evolution of the genetic code. *In* Evolution of Life: Fossils, Molecules and Culture (eds. Osawa, S. and Honjo, T.). Springer-Verlag, Tokyo, pp. 79–95.
 - 27) Himeno, H., Masaki, H., Ohta, T., Kumagai, I., Miura, K.-I. and Watanabe, K. (1987) Unusual genetic codes and a novel genome structure for tRNA^{Ser}_{AGY} in starfish mitochondrial DNA. *Gene* **56**, 219–230.
 - 28) Ohama, T., Osawa, S., Watanabe, K. and Jukes, T.H. (1990) Evolution of the mitochondrial genetic code IV. AAA as an asparagine codon in some animal mitochondria. *J. Mol. Evol.* **30**, 329–332.
 - 29) Tomita, K., Ueda, T., Ishiwa, S., Crain, P.F., McCloskey, J.A. and Watanabe, K. (1999) 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.
 - 30) de Crécy-Lagard, V., Marck, C., Brochier-Armanet, C. and Grosjean, H. (2007) Comparative RNomics and Modomics in Mollicutes: Prediction of gene function and evolutionary implications. *IUBMB Life* **59**, 634–658.
 - 31) Tanaka, R., Muto, A. and Osawa, S. (1989) Nucleotide sequence of tryptophan tRNA gene in *Acholeplasma laidwairi*. *Nucleic Acids Res.* **17**, 5842.
 - 32) Yamao, F., Iwagami, S., Azumi, Y., Muto, A., Osawa, S., Fujita, N. and Ishihama, A. (1988) Evolutionary dynamics of tryptophan tRNAs in *Mycoplasma capricolum*. *Mol. Gen. Genet.* **212**, 364–369.
 - 33) Andachi, Y., Yamao, F., Muto, A. and Osawa, S. (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.
 - 34) Inagaki, Y., Bessho, Y. and Osawa, S. (1993) Lack of peptide-release activity responding to codon UGA in *Mycoplasma capricolum*. *Nucleic Acids Res.* **21**, 1335–1338.
 - 35) Inagaki, Y., Bessho, Y., Hori, H. and Osawa, S. (1996) Cloning of the *Mycoplasma capricolum* gene encoding peptide-chain release factor. *Gene*

- 169, 101–103.
- 36) Ohama, T., Suzuki, T., Mori, M., Osawa, S., Ueda, T., Watanabe, K. and Nakase, T. (1993) Non-universal decoding of the leucine codon CUG in several *Candida* species. *Nucleic Acids Res.* **21**, 4039–4045.
- 37) Watanabe, K., Ueda, T., Yokogawa, T., Suzuki, T., Nishikawa, K., Mori, M., Ohama, T., Nakabayashi, H., Nakase, T. and Osawa, S. (1993) Molecular mechanism of the genetic code variations found in *Candida* species and its implications in evolution of the genetic code. *In* *The Translational Apparatus* (eds. Nierhaus, K.H., Franceschi, F., Subramanian, A.R., Erdmann, V.A. and Wittmann-Liebold, B.). Plenum Press, New York, pp. 647–656.
- 38) Hayashi-Ishimaru, Y., Ohama, T., Kawatsu, Y., Nakamura, K. and Osawa, S. (1996) UAG is a sense codon in several chlorophycean mitochondria. *Curr. Genet.* **30**, 29–33.
- 39) Jukes, T.H. (1966) *Molecules and Evolution*. Columbia University Press, New York.
- 40) Jukes, T.H. (1981) Amino acid codes in as possible clues to primitive codes. *J. Mol. Evol.* **18**, 15–17.
- 41) Jukes, T.H. (1983) Evolution of the amino acid code: inferences from mitochondrial codes. *J. Mol. Evol.* **19**, 219–225.
- 41a) Muramatsu, T., Nishikawa, K., Nemoto, F., Kuchino, Y., Nishimura, S., Miyazawa, S. and Yokoyama, S. (1988). Codon and amino acid specificities of a transfer RNA are both converted by a single post-transcriptional modification. *Nature* **336**, 179–181.
- 42) Zinoni, F., Heider, J. and Böck, A. (1990) Features of the formate-dehydrogenase mRNA necessary for decoding of the UGA codon as selenocysteine. *Proc. Natl. Acad. Sci. USA* **87**, 4660–4664.
- 43) Tormay, P., Wilting, R., Heider, J. and Böck, A. (1994) Genes coding for the selenocysteine-inserting tRNA species from *Desulfomicrobium baculatum* and *Clostridium thermoaceticum*: Structural and evolutionary implications. *J. Bacteriol.* **176**, 1268–1274.
- 44) Srinivasan, G., James, C.M. and Krzycki, J.A. (2002) Pyrrolysine encoded by UAG in Archaea: charging of a UAG-decoding specialized tRNA. *Science* **296**, 1459–1462.

(Received Nov. 16, 2007; accepted Dec. 28, 2007)

Profile

Syozo Osawa, Dr. Sci., and Professor Emeritus of Nagoya University and Hiroshima University, was born in Tokyo in 1928. He graduated from Nagoya University, Faculty of Science, Department of Biology (amphibian embryology course) in 1951, and then studied biochemistry of the cell nuclei in the laboratory of Dr. Alfred E. Mirsky at the Rockefeller Institute for Medical Research, New York from 1954 to 1955. He returned to Nagoya University and worked on molecular biology of translational apparatus at the Department of Biology and the Institute of Molecular Biology (1956–1962). In 1963, he moved to Hiroshima University as a professor of the Department of Biochemistry and Biophysics of the Institute of Nuclear Medicine and Biology, where he continued the studies on molecular biology of the translational apparatus, especially of the biosynthesis, structure and genetics of ribosomes. Meanwhile, he began to study molecular phylogeny of the ribosomal components. In 1979, Hori and Osawa succeeded in constructing a phylogenetic tree of 5S ribosomal RNAs from 54 eukaryotes and prokaryotes. One of the main conclusions was that *Halobacterium* (one of what Woese named “archaebacteria” and then renamed “Archaea”), is phylogenetically closer to eukaryotes than eubacteria. Osawa was then appointed to a professor of Molecular Genetics of the Department of Biology, Nagoya University in 1981, where he and his collaborators performed several lines of work until Osawa’s retirement in 1992. Two of them may be considered as representatives during the above period. (1) Hori and Osawa constructed a phylogenetic tree of 352 5S rRNA sequences from major groups of organisms when the DNA sequencing technique had not been developed yet (1987). The tree supports the idea that eubacteria diverged during the early stages of evolution, followed by separation of metabacteria (named by Hori & Osawa, 1982) and eukaryotes. As Cavalier-Smith emphasized (2002), it is a pity that the name metabacteria did not catch on for archaebacteria, since they are undoubtedly the most derived and recent of all bacterial phyla. Aarchaebacteria and Archaea are surely the misleading names (Mayr, 1998). (2) Osawa and his colleagues (inclusive of the co-authors of this review article) conducted an extensive investigation on evolution of the genetic code, and published a monograph “Evolution of the



Genetic Code” from Oxford University Press in 1995. After retirement from Nagoya University in 1992, he was appointed to an advisor of JT Biohistory Research Hall, Takatsuki, Osaka, where he and his associates studied the molecular phylogeny of the carabid ground beetles from 1992 to 2000. A monograph of this work, entitled “Molecular Evolution and Phylogeny of Carabid Ground Beetles”, was published from Springer Verlag in 2004, and the subsequent progress on this subject was published in PJA Ser B: 82(7), 2006. He was awarded the Promotion Prize from the Japanese Biochemical Society (1966), the Chunichi Culture Award (1985), the Kihara Award of the Genetics Society of Japan (1987), the Promotion Prize from the Japan Genetics Organization (1989), the Japan Academy Prize (1992), and Motoo Kimura Memorial Prize of Science (2001). He is the honorary member of the Genetics Society of Japan and the honorary member of Society of Evolutionary Studies, Japan. He is also an amateur entomologist and belongs to several entomological societies in Japan. There are numerous new species of beetles found by him. Among them, some were described by himself, and many others were named *osawai* or *syozoi* by professional entomologists.