

Recent Evidence for Evolution of the Genetic Code

SYOZO OSAWA,^{1*} THOMAS H. JUKES,^{2*} KIMITSUNA WATANABE,^{3†} AND AKIRA MUTO¹

Department of Biology, Nagoya University, Chikusa-ku, Nagoya 464,¹ and Department of Biological Sciences, Tokyo Institute of Technology, Nagatsuta, Yokohama 227,³ Japan, and Space Sciences Laboratory and Department of Integrative Biology, University of California, Berkeley, California 94720²

INTRODUCTION	230
Wobble Pairing and the Code	230
History	230
Codes without commas	231
RNY Hypothesis	232
Sequencing of Genes—the Ultimate Solution to the Code	232
Frozen-Accident and Stereochemical Theories	232
NATURE OF THE UNIVERSAL GENETIC CODE.....	233
Codon-Anticodon Pairing Rules.....	233
U	233
G	234
I	234
A	234
C	234
Four-way wobbling of U versus the two-out-of-three mechanism	234
Role of Nucleotide 37 in Codon-Anticodon Pairing	234
DIRECTIONAL MUTATION PRESSURE AND CODON USAGE	235
Directional Mutation Pressure.....	235
Genesis of Directional Mutation Pressure	235
Directional Mutation Pressure in Bacteria	236
Directional Mutation Pressure in Vertebrate DNA.....	236
Directional Mutation Pressure in Mitochondria	236
Directional Mutation Pressure and Codon Usage.....	236
Amino acid codons	236
Initiation and stop codons	237
Directional Mutation Pressure and Amino Acid Composition of Proteins.....	237
Codon Selection by Other Factors.....	238
Codon selection by tRNA	238
Stop Codon Selection by Release Factors.....	239
VARIABILITY OF tRNAs	240
Anticodon Composition of tRNA in Organisms and Organelles	240
Bacteria	240
Chloroplasts and mitochondria.....	240
Resemblance between <i>Mycoplasma</i> spp. and mitochondria	240
Eukaryotes	242
DISTRIBUTION OF CODE CHANGES IN NUCLEAR AND MITOCHONDRIAL SYSTEMS.....	242
Nuclear Code	242
Mitochondrial Code	243
Time Required for Code Change	245
CODON CAPTURE	246
Unassigned Codons	247
Possible Mechanisms of Code Changes.....	249
Codon capture.....	249
UGA from stop to Trp	249
UAR from stop to Gln.....	250
UGA from stop to Cys	250
CUG from Leu to Ser	250
Mitochondrial Code.....	251
UGA from stop to Trp	251
AAA from Lys to Asn; AUA from Met to Ile as a reversal of Ile to Met; UAA from stop to Tyr.....	251

* Corresponding authors.

† Present address: Department of Industrial Chemistry, Faculty of Engineering, University of Tokyo, Hongo, Bunkyo-Ku, Tokyo, Japan.

CUN from Leu to Thr.....	251
AUA from Ile to Met.....	252
Yeast mitochondria.....	252
Metazoan mitochondria.....	252
AGR from Arg to Ser or stop.....	252
UGA AS A SELENOCYSTEINE CODON.....	254
Modifications of Secys-Incorporating Systems.....	255
Secys in Evolution.....	255
EARLY HISTORY OF THE CODE.....	256
Evolutionary Origin of tRNA.....	256
Codons in Primitive Systems.....	257
EPILOGUE.....	257
ACKNOWLEDGMENTS.....	258
REFERENCES.....	258

INTRODUCTION

The genetic code, or amino acid code, is a name that refers to the system for passing information from DNA and RNA into the synthesis of proteins. The code is usually written as a set of all the 64 possible arrangements, the codons, of the four RNA nucleotides (Table 1). Each of 61 codons is assigned to a specific amino acid, and the other 3 are stop codons that terminate the synthesis of a protein molecule.

The code in Table 1 was established around 1966 by using mainly *Escherichia coli* systems. This code was thought to be common to all organisms, i.e., universal, because it applied to certain other organisms such as tobacco mosaic virus and vertebrates.

Such apparent universality of the code led Crick to propose the frozen-accident hypothesis (42), which is discussed below. It states that the proteins had become so sophisticated in a single pool of progenote cells that any changes in codon meaning would disrupt proteins by making unacceptable amino acid substitutions throughout their sequences, so that evolution of the code stopped. In 1979, less than 15 years since the code was first deciphered, it was found that the code in vertebrate mitochondria differed from the universal code by using codons AUA for Met and UGA for Trp (14). This shattered the myth of a frozen universal genetic code. The changes leading to the use of UGA for Trp and AUA for Met in the vertebrate mitochondrial code are discussed in a later section. To account for these changes, it was proposed that because the genome of the mitochondria is small, the mitochondria can probably tolerate changes in the code that would not be acceptable to a larger and more complex nuclear genome (109). This proposal (although it is still being reiterated [see below]) became obsolete and unlikely in 1985, when it was discovered that the code had changed in intact organisms. In *Mycoplasma capricolum*, UGA codes for Trp. In certain ciliated protozoans, UAR codes for Gln. Some code changes in other nuclear genomes have also been reported. As a result, it is now realized that the code is still evolving in both mitochondrial and nuclear genomes, presumably by essentially the same mechanisms, which can take place without deleterious changes in amino acid sequences of proteins.

Wobble Pairing and the Code

The general pattern of the genetic code results from biochemical properties of nucleotides rather than from any mathematical formula. The first two bases in the codon pair with the second two anticodon bases in tRNA molecules, according to Watson-Crick rules (236). The first anticodon base pairs with the third codon base according to the wobble rules (41). The triplet nature of the code is a result of these rules. There is no

such thing as a two-letter code or a two-and-a-half-letter code. The fact that G pairs with U as well as with C is responsible for the existence of the genetic code as we know it. The pairing of G and U takes place between the first anticodon positions and the third codon position.

All pairs of codons ending in a pyrimidine, such as UUU and UUC, are codes for a single amino acid. This is because anticodon GAA, for example, pairs with both UUU and UUC. An A in the first anticodon positions almost never exists, and Uhlenbeck et al. (229) have shown that A in ANN anticodons, obtained by site-specific base substitution, does not pair efficiently with U.

Most pairs of codons ending in a purine, such as CAA and CAG, also code for a single amino acid, because U in the first anticodon position pairs with both A and G in the third codon position. There are two exceptions: AUA and AUG code for two different amino acids and UGA is a stop codon, but UGG codes for tryptophan. The latter is explained by another wobble rule, i.e., that C in the first codon position pairs only with G, so that anticodon CCA for tryptophan pairs only with UGG. Anticodon UCA is excluded because it would pair with UGA (stop) as well as with UGG (Trp).

Anticodon-codon pairing with A and G in the third codon position has some special features. U in the first anticodon position is often modified into a thiouridine derivative or its equivalent, which pairs strongly with A and only weakly or not at all with G; in such cases, anticodons with C in the first position supply the pairing for G.

An obvious feature of the code is that there are eight cases in which a single amino acid has four codons. These groups of four are called family boxes. Wobble pairing in eukaryotic family boxes usually includes pairing of I (inosine) with U, C, and A.

Family boxes in the mitochondrial codes, and also in a few other cases, have another type of wobble pairing: U in the first anticodon position pairs with all four of the third anticodon bases, U, C, A, and G.

History

Students who have been introduced to molecular biology in the past 20 years usually take the genetic code for granted. Yet the early history of the code is still quite interesting, because it shows how variable the early guesses were before any experimental evidence was available.

Dounce (51) was the first to propose a model for protein synthesis based on a relationship between RNA and amino acids. He postulated that "the specific arrangement of amino acid residues in a given peptide chain" is derived from the "specific arrangement of nucleotide residues in a corre-

TABLE 1. The universal amino acid code^a

Codon	Amino acid	Codon	Amino acid	Codon	Amino acid	Codon	Amino acid
UUU	Phenylalanine	UCU	Serine	UAU	Tyrosine	UGU	Cysteine
UUC	Phenylalanine	UCC	Serine	UAC	Tyrosine	UGC	Cysteine
UUA	Leucine	UCA	Serine	UAA	Stop	UGA	Stop
UUG	Leucine	UCG	Serine	UAG	Stop	UGG	Tryptophan
CUU	Leucine	CCU	Proline	CAU	Histidine	CGU	Arginine
CUC	Leucine	CCC	Proline	CAC	Histidine	CGC	Arginine
CUA	Leucine	CCA	Proline	CAA	Glutamine	CGA	Arginine
CUG	Leucine	CCG	Proline	CAG	Glutamine	CGG	Arginine
AUU	Isoleucine	ACU	Threonine	AAU	Asparagine	AGU	Serine
AUC	Isoleucine	ACC	Threonine	AAC	Asparagine	AGC	Serine
AUA	Isoleucine	ACA	Threonine	AAA	Lysine	AGA	Arginine
AUG	Methionine	ACG	Threonine	AAG	Lysine	AGG	Arginine
GUU	Valine	GCU	Alanine	GAU	Aspartic acid	GGU	Glycine
GUC	Valine	GCC	Alanine	GAC	Aspartic acid	GGC	Glycine
GUA	Valine	GCA	Alanine	GAA	Glutamic acid	GGA	Glycine
GUG	Valine	GCG	Alanine	GAG	Glutamic acid	GGG	Glycine

^a The following abbreviations are used in the text: N—A, C, G, or U (T); R—A or G; Y—C or U (T).

sponding specific nucleic acid molecule” and that the amino acids were specified by the “immediate surroundings” of the nucleic acid bases, so that there were 10 possible such surroundings for adenine (AAA, GAA, CAG, etc.) and 10 more for each of the other three bases to give a total of 40 such triplets, even if direction were disregarded, so that AAG would be equivalent to GAA, for example. If direction were important, the possibilities would be increased, but 40 triplets would be enough to code for all 20 amino acids. He also suggested that the information in the RNA template might be stored in DNA: “it could conceivably happen that the deoxyribonucleic acid gene molecules would act as templates for ribonucleic acid synthesis, and that the ribonucleic acids synthesized on the gene templates would then in turn become templates for protein synthesis in the nucleus or cytoplasm or both.” Dounce’s proposal provided for the “sequence hypothesis,” for the triplet genetic code, and for mRNA, which was not even conceived of in 1952. Indeed, the fact that amino acids in proteins were present in a definite order was just coming to light, and the double helix was a year away.

The publication by Watson and Crick (236) contained the statement that the pairing of bases in DNA “suggests a possible copying mechanism for the genetic material.” This immediately thrust the genetic code into prominence as the most important unsolved question in biology. Several erroneous attempts were made to guess what it was. George Gamow, a cosmologist, proposed a “diamond code,” with each amino acid specified by a quartet of bases in the DNA double helix (62). Each amino acid fitted into a hole formed by a specific quartet. This proposal was doomed from the start, because proteins are synthesized in the cytoplasm and not in the nucleus of eukaryotic cells, as had been shown by Borsook et al. in 1949 (24). The proposal was also unsupported because it allowed only certain amino acids to adjoin each other in proteins, and no such restrictions on amino acid sequences exist.

Codes without commas. The term “codes without commas” was coined by Crick and colleagues (45) in a theoretical discussion of how coding sequences of nucleotides could be read. At that time there was no idea that coding sequences could be placed in the correct reading frame by an initiator codon. The authors listed two difficulties in the

“coding problem.” First, why are there not 64 kinds of amino acids? Second, how does one know how to choose the groups of three nucleotides?

The sequence could be written as

...U C A, C G G, A U A, U G C..., or
 ...U, C A C, G G A, U A U, G C..., or
 ...U C, A C G, G A U, A U G, C...

in which the commas divide the letters into groups of three, each representing one amino acid. “The problem is how to read the code if the commas are [removed] i.e., a comma-less code.” (45)

The authors offered a solution by proposing that “certain triplets (codons) make sense and some make nonsense,” and that a code could be devised in which the maximum number of amino acids coded for was not greater than 20. A solution for 20 was given by first excluding the four triplets, AAA, CCC, GGG, and UUU, because if AAA is placed next to AAA, the sequence can be misinterpreted by reading it in the wrong frame. The remaining 60 triplets were grouped into 20 sets of three, the sets being cyclic permutations of one another. One such set could be the actual code, and the other two would be excluded.

The point was to reduce the 64 codons to a total of 20, one for each amino acid (although we cannot see why more than one codon per amino acid would not have been a better choice. Indeed, Dounce [51] had this idea.) The comma-free code had no evolutionary flexibility among organisms with different G+C contents in DNA.

The flexibility is seen when homologous genes in different organisms are compared (114, 160). The silent sites of such genes can become higher in G+C content, without changing the amino acid sequence, when the G+C content of DNA is increased (106). This would not be possible in the commaless code, because its codons do not contain silent sites. Any nucleotide replacement in its codons produces either a change in amino acid assignment or a nonsense codon.

Another objection, which was not mentioned, is that 70% of possible single-nucleotide substitutions produced by point mutations will give rise to untranslatable codons, so that evolution would be virtually impossible. In the universal code, only 4.2% of possible point mutations produce stop

codons. This objection to the commaless code, the blocking of evolution, is perhaps its most serious defect.

This elegant hypothesis of 1957 was destroyed by an ugly fact when Nirenberg and Matthaei discovered (164) that poly(U), acting as an mRNA, coded for polyphenylalanine; thus an "excluded" triplet, UUU, was the breakthrough clue to solving the coding problem. Later, it was shown that coding sequences were placed in the correct reading frame by a start codon, usually AUG. This automatically inserted the commas (see below) and drove another nail into the coffin of the commaless code.

However, in 1988 Crick said of the 1957 proposal: ". . . the correct genetic code. . . has proved decisively that the whole idea is quite erroneous. However, it is just conceivable that it may have played a role near the origin of life, when the code first began to evolve, but this is speculation." (43)

Theories die hard, especially when they have been published in a scientific journal, even though Crick himself (43) has criticized theorists who have an idea "that they can never abandon." If the comma-free code was completely wrong today, why should it have been any more likely 3 billion years ago? Indeed, it would have impeded evolution so much that it would have been unacceptable at any stage. Crick revived the comma-free idea when he and his coworkers (44) proposed that the primitive mRNA was a "repeating family of sequences. . . RRY,RRY,RRY. . . where the commas are written to show the correct phase of reading, and for the anticodon [loop] the family 3' UGYYRUU with the triplet in italics". . . "this restricted base sequence is comma-free in the sense of Crick, Griffith and Orgel" and "the codons allowed are GGY(Gly) GAY(Asp) AGY(Ser) and AAY(Asn)."

The complete anticodon loop in tRNA was postulated by Crick and colleagues (44) as participating in the translation process. This loop, tRNA positions 32 to 38, consists, in all tRNAs, of the sequence 5'-Y-U-anticodon-R-N-3' (219). The model proposed by Crick and coworkers for the origin of protein synthesis consisted of (i) an mRNA molecule made up of alternating RRY sequences and (ii) four different archetypal tRNA molecules, each with an anticodon loop carrying the sequence UGYYRY. The anticodon in this sequence is YYR, read from the 3'-OH end to the 5' end of the tRNA. The anticodons can therefore be UUR, UCR, CUR, or CCR. These would pair with codons AAY, AGY, GAY, and GGY, respectively. These are codons for asparagine, serine, aspartic acid, and glycine, respectively. The other details, such as translocation, were described by Crick and colleagues (44).

This proposal would apparently limit the composition of proteins to the four amino acids that are encoded by RRY, because Y (C and U) in the third codon position always pairs only with G.

RNY Hypothesis

The RNY hypothesis is a variation of the RRY proposal for coding sequences. Both proposals say that a commaless code was the starting point, which seems unlikely because of the constraints placed on evolution as discussed above. The RNY model allows for twice as many codons as RRY and hence provides for eight amino acids (Ala, Asn, Asp, Gly, Ser, Thr, Tyr, and Val). Shepherd (211) has claimed a residuum of (RNY)_n in coding sequences, but Wong and

Cedergren (246) pointed out that analyses based on rates of silent substitutions, frequencies of nucleotide doublets, and synonymous codon usage rule out support for a primitive (RNY)_n gene structure. RNN predominates over YNN in coding sequences, first because glutamic acid, aspartic acid, and lysine are all RNN and are present in proteins on average at levels higher than their representation in the genetic code table (115), and second because the stop codons are YNN and occur only once in each gene for a protein. This predominance of RNN over YNN is therefore related to protein function and chain termination rather than to a residue of a primitive code.

Sequencing of Genes—the Ultimate Solution to the Code

Then a new event happened. A method was developed for sequencing RNA molecules, such as occur in RNA viruses, by Sanger and coworkers (204). A viral protein was sequenced by Fiers and colleagues (56) and aligned with the sequence of the viral RNA. The correspondence between the RNA and the amino acid sequences could be placed in three different reading frames. One of these gave rise to the genetic code corresponding to the viral protein sequence, whereas the other two did not. Furthermore, the other two reading frames led to one stop codon on an average of about 22 codons, and these were therefore not "open reading frames." Therefore, all the codons that had been deduced from the earlier experiments with synthetic RNAs were verified by a system occurring naturally that was not comma free.

If we had waited a little longer, the code would have been discovered by sequencing a protein and an RNA molecule obtained from a virus!

Frozen-Accident and Stereochemical Theories

The frozen-accident theory was stated by Crick (42), 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 accompanied by many simultaneous mutations 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." [emphasis in original]

Freezing of the code, according to this theory, took place when life had evolved to a certain level of complexity. Before this point had been reached, changes in protein sequences that resulted from changes in the code could be tolerated. Crick proposed (42) that the final steps of evolution of the code would "involve an increase in the precision of recognition and the introduction of new amino acids." It has been pointed out that increasing the code to its present total of 20 amino acids provides for the great diversity of

TABLE 2. Anticodon-codon pairing; present status of wobble usage

First anticodon nucleoside	Third codon nucleoside	Usage	Systems
U	U, C, A, G	Family boxes	Mitochondria, <i>Mycoplasma</i> spp., chloroplasts
xo ⁵ U	U, A, G	Family boxes (Ser UCN, Val, Thr, Ala)	Eubacteria
cmnm ⁵ U, mcm ⁵ U, Um	A, G	Two-codon sets	Mitochondria, bacteria, eukaryotes
xm ⁵ s ² U	A, (G)	Two-codon sets	Eubacteria, eukaryotes
G	U, C	Two-codon sets	All
G	U, C	Family boxes	Bacteria
Q	U, C	Two-codon sets	Eubacteria, eukaryotes
I	U, C, A	Arg CGN	Eubacteria
I	U, C, A	All family boxes except Gly GGN	Eukaryotes
A (rare)	U?	Thr ACU, Arg CGN	<i>Mycoplasma</i> spp., yeast mitochondria
C	G	All	All
L (2-lysyl C)	A	Ile AUA	Eubacteria, plant mitochondria

organisms in the biosphere (110). There is some evidence for a freezing of the code in one respect: the same 20 amino acids are in all codes, including mitochondrial codes. However, the known changes in the meaning of codons are difficult to reconcile with the frozen-accident theory. This will be discussed below.

The stereochemical theory says that the code is universal because each amino acid fits its own anticodon or codon in some way. There is evidence for a specific relationship between the anticodon and the amino acid in some, but not all, cases. The most outstanding example is that when anticodon *CAU (*C is lysidine) in a eubacterial tRNA for isoleucine is changed to CUA, the tRNA becomes a methionine tRNA (154). In contrast, for many suppressor tRNAs a nucleotide substitution takes place in the anticodon without changing the amino acid that charges the tRNA.

Representation of amino acids in the genetic code does not coincide with their average usage in proteins. Arginine, with 6 of the 61 codons that code for amino acids, might be expected to be present in proteins at a level of about 9.8%, but the average arginine content of proteins is only about 4% (115, 125). The reverse is true for lysine and glutamic acid. These findings might be the result of evolutionary divergence from the proteins that existed in earlier organisms at the time the universal code first appeared.

In its extreme form, the stereochemical theory is said to liken the genetic code to a "periodic table" (120) in which the "polarity and bulkiness of amino acid side chains can be used to predict the anticodon with considerable confidence." Such a postulation does not provide for changes in the genetic code in mitochondria and in some intact organisms. For example, "universal" lysine codon AAA is a codon for asparagine in mitochondria of flatworms and echinoderms (20, 116, 173) and "universal" leucine anticodon CAG is an anticodon for serine in *Candida cylindracea*, a yeast (249). Woese et al. (242) were strongly inclined toward stereochemical explanations of the code.

In general, it may be best to keep theories of evolution of the code as close as possible to existing experimental findings and flexible enough to accommodate future discoveries. It is unlikely that some mathematical solution will be found to explain, in a single tour de force, the nature of the code, although efforts to achieve this are numerous. These efforts usually ignore the relation of the wobble rules to the structure of the code.

NATURE OF THE UNIVERSAL GENETIC CODE

Codon-Anticodon Pairing Rules

Codon-anticodon pairing takes place according to the wobble rules (41). These have expanded since 1966, and they differ in various organisms and systems. Several modifications of the first anticodon base participate in pairing. The latest version of the rules is given in Table 2.

An obvious feature of the universal genetic code is that there are eight cases in which a single amino acid has four codons. These groups of four are called family boxes. These eight family boxes are each potentially translatable by a single anticodon with the general formula UNN (see below). In most systems, however, there is more than one anticodon per family box.

Twelve two-codon sets are not in family boxes. These sets end either in a pyrimidine (U or C) or in a purine (A or G). The seven pyrimidine-terminated sets are each translated by a single GNN anticodon, as discussed above. The five sets that end with purines are UUR, CAR, AAR, GAR, andAGR, and they can each pair in translation with an anticodon *UNN (*U is modified U [see below]) or they can be translated by two anticodons: the G-terminated codon by anticodon CNN and the A-terminated codon by anticodon *UNN.

Most pairs of codons ending in a purine, such as CAA and CAG, code for a single amino acid, because U in the first anticodon position pairs with both A and G in the third codon position. There are two exceptions: AUA and AUG code for two different amino acids, Ile and Met, respectively, and UGA is a stop codon but UGG codes for Trp (see above).

U. An unmodified U in the first position of the anticodon pairs with all four bases, U, C, A, and G, in the third position of codons (four-way wobble). This form of translation takes place in some of the mitochondria, in *Mycoplasma* spp., and in two family boxes of the chloroplast code.

There are various modifications of U that restrict its pairing. One modification of U, ⁺U, is a derivative of uridine-5-oxyacetic acid (xo⁵U) that has been found only in eubacteria. The xo⁵U takes the c^{3'}-endo form as well as the c^{2'}-endo form and is thus flexible enough to recognize U in addition to A and G at the third position of codons (250). Anticodons UNN and ⁺UNN occur only in family boxes, and never in two-codon sets, because NNY and NNR in two-codon sets code for different amino acids.

Other modifications of U, *U, are 2-thiolated uridine (xm⁵s²U), 5-carboxymethylaminomethyluridine (cmnm⁵U),

5-methylcarbonylmethyluridine (mcm⁵U), and 2-*O*-methyluridine (Um) (see below). Since codons NNY and NNR in 12 two-codon sets code for different amino acids, a distinction must be made between NNY codons pairing with anticodon GNN and NNR codons pairing with anticodon UNN, the first anticodon nucleotide U of which must be modified to prevent it from mispairing with NNY codons (75). In consequence, the U in all UNN anticodons in two-codon sets is modified to *U.

In tRNA species with xm⁵s²U in the first anticodon position, the steric effect between the 2-thiocarbonyl group and the 2'-hydroxy group stabilizes the c³-endo form by preventing it from taking the c²-endo form and enhances the rigidity of the anticodon moiety to allow a stable pairing with A at the third position of the codon (250). The modified forms cmnm⁵U, mcm⁵U, and Um, although lacking the 2-thiocarbonyl group, take the c³'-endo form and pair more efficiently with A than with G, so that the misrecognition of codons terminating with uridine is prohibited (145, 250). Because of this, NNR two-codon sets are often translated by two anticodons: the G-terminated codon mainly by anticodon CNN, and the A-terminated codon by anticodon *UNN. The presence of anticodon *UNN in a family box is not deleterious, if anticodon GNN is present. In fact, tRNA with anticodon cmnm⁵UCC is present for Gly in *Bacillus subtilis* (156).

G. In many bacteria and eukaryotes, G at the first anticodon position of tRNAs that recognize NAY in two-codon sets is modified to queosine (Q) or its derivative (165, 166). Tyr tRNA with G, instead of Q, at the first anticodon position often mistakenly pairs with stop codon UAG (166). Possibly Q correctly recognizes codons NAY, preventing mispairing with codons NAR for another amino acid (or stop codon). The first anticodon nucleoside G is unmodified in tRNAs for codons in family boxes. Thus, in most cases, misreading of NAY codons by UUN anticodons seems to be prevented by modification of U in the first anticodon position to *U so that only NAR codons are recognized, while misreading of NAR codons by GUN anticodons is prevented by modification of G to Q. Such double insurance to prevent misreading in two-codon sets is not always the case, since the first anticodon nucleoside G for NAY codons in *Micrococcus luteus* (122), *Mycoplasma capricolum* (4), and some mitochondria (219) is not modified. The misreading of codon NAR by anticodon GUN might occur at such a low frequency that appreciable functional disturbance does not result.

I. I at the first anticodon position occurs in eight anticodons of eukaryotic family boxes and in the Arg (CGN) family box in eubacteria with the exclusion of GNN anticodons. In eukaryotes, the enzyme hypoxanthine ribosyltransferase is considered to replace adenosine by hypoxanthine riboside (inosine) (55) in the first position of anticodons. In making those replacements possible, ANN anticodons may have been formed by mutation of GNN. INN anticodons would produce intolerable ambiguity in pairs of two-codon sets by simultaneously recognizing codons for different amino acids. Therefore, INN anticodons occur only in family boxes and never in two-codon sets.

A. Lewin (138) states that "there seems to be an absolute ban on the employment of A" (at the first anticodon position). This ban is not absolute, because an unmodified A at the first anticodon position has been reported for yeast mitochondrial tRNA^{Arg}_{ACG} (217) and for *Mycoplasma* spp. tRNA^{Thr}_{ASU} (4, 203). *Aspergillus* mitochondrial tRNA^{Gly}_{ACC} (126) and chloroplast tRNA^{Arg}_{ACG} (determined by DNA sequencing) (213, 230) might also have an unmodified A.

In *Mycoplasma* spp. tRNA^{Thr}_{AGU} occurs together with

tRNA^{Thr}_{UGU} in the Thr family box (3, 4, 203). Andachi and coworkers (3) concluded that anticodon AGU has been developed for the translation of the most abundantly used Thr codon, ACU. The tRNA^{Thr}_{AGU} gene has resulted from a mutation in the anticodon of the tRNA^{Thr}_{UGU} gene that is similar in sequence to the *B. subtilis* tRNA^{Thr}_{UGU} gene, and not to the gene for tRNA^{Thr}_{GGU} (3). The anticodon AGU in *Mycoplasma capricolum* translates codons ACU, ACC, and ACG in vitro and, probably to a lesser extent, ACA (126a).

C. Ile codon AUA in eubacteria and plant mitochondria pairs only with anticodon *CAU, in which *C is 2-lysylcytidine (L) (154, 155, 239). The *Escherichia coli* tRNA^{Ile}_{LAU} is charged exclusively with Ile by isoleucyl tRNA synthetase and recognizes only codon AUA. When L is replaced by unmodified C, the identity of this tRNA changes from Ile to Met and translates only codon AUG as Met (154). No examples having anticodon LNN other than tRNA^{Ile}_{LAU} have been reported.

Four-way wobbling of U versus the two-out-of-three mechanism. As noted above, a single anticodon, UNN, U unmodified, exists in most or all family boxes of mitochondria and *Mycoplasma* spp. In contrast to four-way wobbling, Lagerkvist (130, 131) proposed that these codons are translated by the "two-out-of-three" mechanism without discrimination between the nucleosides in the third codon position and the first anticodon position, so that only the second and third positions of the anticodon pair with the codon. Four-way wobbling is a more consistent explanation. In yeast mitochondria, only one species of tRNA, tRNA^{Arg}_{ACG}, exists for CGN codons and yet codon CGG is used. The same is true for tobacco (233) and rice chloroplasts, whereas liverwort chloroplasts have both tRNA^{Arg}_{ACG} (modification unknown) and tRNA^{Arg}_{CCG} (233).

Sibler and colleagues (217) and Shinozaki and colleagues (213) state that anticodon ACG, like anticodon UCG, would translate all the CGN codons by a two-out-of-three mechanism. However, in a cell-free translation system of *Mycoplasma capricolum*, anticodon ICG (Arg) translates codons CGU, CGC, and CGA following the standard wobble rules, and yet codon CGG is not translated (168). The two-out-of-three mechanism does not seem to occur in this case. It is more likely that the first nucleoside U of anticodons normally pairs by four-way wobble with U, C, A, or G.

Alternative explanations for translation of codon CGG in yeast mitochondria or chloroplasts in the apparent absence of anticodon CCG could be that tRNA^{Arg}_{CCG} for codon CGG is imported into yeast mitochondria or chloroplasts or that tRNA^{Arg}_{ACG} can read all the CGN codons by four-way wobbling to some extent. According to the wobble rules, unmodified A pairs only with U, but unconventional pairing of A with C, G, and A at a low frequency is possible (see above; 126a). In this connection, it might be significant that tRNA^{Arg}_{ACG} occurs only when the usage of CGN Arg codons is low compared with that of AGR Arg codons.

Role of Nucleotide 37 in Codon-Anticodon Pairing

Nucleotide 37, adjoining the third anticodon position in tRNA, is often extensively modified, and the modifications may be concerned with strengthening ANU and UNA pairing between the third anticodon nucleotide (position 36) and the first codon nucleotide. In contrast, the tRNA nucleotide at position 33 is an invariant unmodified U. Presumably an unmodified U would be the least likely to interfere with wobble pairing by the adjacent first anticodon nucleotide (position 34) (22, 108, 165).

In the case of pairing between U at position 36 (anticodon position 3) and A at codon position 1 (3-1 pairing), the modification of nucleotide 37 is usually a bulky threonyl-containing side chain on position 6 of adenine [*N*-((9- β -D-ribofuranosyl purine-6-yl)carbamoyl)threonine] (t^6A) (22, 165). Such side chains may prevent, by steric hindrance, wobble-pairing between U in the third anticodon position and G in the first codon position (108). Such prevention of U-G pairing would be important to the fidelity of translation of all codons starting with A except initiator methionine, which is sometimes encoded in bacteria by GUG, and it is significant that initiator methionine tRNA in bacteria has an unmodified A at position 37 in all 16 sequences plus three in chloroplasts that have been determined (219). Unmodified A at 37 would presumably facilitate U-G pairing at position 36 (108). In this connection, it is interesting that in *Mycoplasma capricolum*, position A-37 of elongator tRNA^{Met}_{CAU} is not modified, whereas A-37 of this tRNA species in other organisms is always modified to t^6A , like other tRNAs for ANN codons (4). However, because of AT pressure, codon GUG (valine) is only rarely used (0.2% among all codons) in *Mycoplasma capricolum* (175), so that occasional U · G mispairing between anticodon CAU (Met) and codon GUG (Val) may not be serious.

In 3-1 pairing between A and U, the most common modification at position 37 is a large isopentenyl side chain attached to adenine (*N*₆-isopentenyladenosine) (i^6A) or its derivative. An even more bulky group, the Y-base, occupies position 37 in eukaryotic phenylalanine tRNAs for 12 of the 17 species of this tRNA in which this nucleotide has been identified; the other 5 have 1-methylguanosine (m^1G) or 6-methyladenosine (m^6A) at this position (22, 108, 122, 165).

Codons starting with U must not pair with anticodons ending with G. These are anticodons for Leu (CUN), Pro, His, Gln, and Arg (CGN). Most nucleotides in position 37 in tRNAs for these amino acids are methyl-G (22, 108, 122, 165). Evidently there is no 3-1 pairing between G and U, but only between G and C.

Position 37 is often unmodified following an anticodon ending in C (for codons starting with G), but in other cases it is methylated (22, 108, 122). For example, this position is occupied by m^6A , 2-methylthio-*N*-6-isopentenyladenosine (ms^2i^6A), or 2-methyladenosine (m^2A) in *E. coli* (165), *B. subtilis* (165), and *Mycoplasma capricolum* (4), whereas it is always unmodified A in *Micrococcus luteus* (tRNA^{Val}_{GAC}, tRNA^{Val}_{CAC}, tRNA^{Ala}_{GGC}, tRNA^{Ala}_{CGC}, tRNA^{Asp}_{GUC}, tRNA^{Glu}_{CUC}, tRNA^{Gly}_{GCC}, and tRNA^{Gly}_{CCC}) (122). Presumably, 3-1 pairing between C and G is strong enough to prevent mispairing, so that modification of A-37 would not be essential.

As noted above, modification profiles of position 37 are similar but not exactly the same among different organisms. It would be of interest if these variations reflect species-specific fine adjustment of codon-anticodon pairing in the translation system.

DIRECTIONAL MUTATION PRESSURE AND CODON USAGE

Directional Mutation Pressure

The genomic DNA of each organism has its characteristic G+C content, often varying in eukaryotes on different chromosomes or in different regions of a chromosome. In eubacteria, the mean genomic G+C content varies from 25 to 75% and is related to phylogeny (Fig. 1), suggesting that

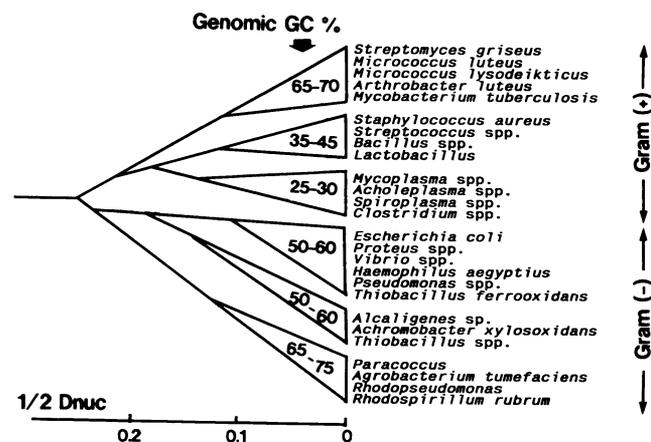


FIG. 1. Phylogenetic tree of eubacteria constructed by 5S rRNA sequences. Abbreviation: Dnuc, evolutionary distance. The G+C content (%) is given for representative bacteria. Modified from reference 89 with permission.

directional evolutionary pressure (constraint) has determined the specific G+C content in each phylogenetic line (88, 89).

A high G+C content is sometimes claimed to be advantageous, e.g., for thermophilicity or halophilicity (15-17, 121). Actually, the genomic G+C content is related to phylogeny and not to such phenotypic characters. It is more likely that directional mutation pressure results from mutations and that the magnitude of mutation pressure is biased with respect to the G+C content and varies among phylogenetic lines, as first proposed by Sueoka (225). According to his theory, the G+C content of genomic DNA is determined by the effective base conversion rate u ($G \cdot C$ to $A \cdot T$) and v ($A \cdot T$ to $G \cdot C$); the G+C content at equilibrium (p) is $v/(u + v)$. It follows that directional mutation pressure has been exerted on DNA toward $A \cdot T$ predominating over $G \cdot C$ (AT pressure) or towards $G \cdot C$ over $A \cdot T$ (GC pressure) (159, 160, 181-184, 226).

Genesis of Directional Mutation Pressure

Mutations can be induced by external factors such as chemical mutagens and UV irradiation, or X rays. However, as directional mutation pressure is exerted for a very long time during evolution, its cause may be internal rather than external. There are examples showing that the G+C content of DNA differs even within the same cell when DNA replication systems differ, such as in chromosomes of higher vertebrates (see below) and in nuclear and mitochondrial DNAs in eukaryotic cells (232). In the *E. coli* system, the G+C content of DNA and codon usage pattern of lysogenic bacteriophages (lambda phage, etc.), whose DNAs are integrated and replicated together with the host chromosome, are similar to those of the host DNA. On the other hand, virulent phages (T4, T7, etc.) that use their own DNA replication system have a genomic G+C content and codon usage different from those of the host (232). These facts suggest that the DNA replication system is involved in determining the genomic G+C content.

The most probable cause of such differences is copy errors during DNA replication. The copy errors occurring predominantly from $A \cdot T$ to $G \cdot C$ or from $G \cdot C$ to $A \cdot T$ would result in a different genomic G+C content among phylogenetic lines, even if the error frequency is very low.

Among mutator genes (*mut*) of *E. coli*, a mutation of the *mutT* gene specifically induces transversions from A · T to C · G pairs at a high rate (40) and that of *mutY* does the same from G · C to T · A pairs (163). The *mut* gene products are mainly components functioning in DNA replication or repair. The *mutT* product has been shown to suppress transversional copy errors from A · T to C · G pairs in DNA replication in vitro (2).

Directional Mutation Pressure in Bacteria

There exists a weak positive correlation of the G+C content of both rRNA and tRNA genes with genomic G+C content. However, the G+C content of spacers and protein genes reveals a strong correlation to genomic G+C content: among various bacteria the G+C content of spacers ranges from about 20 to 80% and that of protein genes from 30 to 75%, as the genomic DNA G+C content varies from 25% (*Mycoplasma capricolum*) to 74% (*Micrococcus luteus*) (160). Thus for a given bacterial species, the G+C contents of spacers, protein genes, and stable RNA genes are all biased in the same direction as the G+C content of the total genome. Such mutation pressure seems to have been exerted on the entire genome. The differential levels of G+C content in the different components of the genome in a given organism can be understood as the consequence of selective constraints that have been exerted to eliminate functionally deleterious mutants (124). Since most parts of spacers are functionally the least important components of the genome, most of the mutations in these regions are selectively neutral and therefore the evolutionary rate of spacers is higher than that of other regions. The rRNA and tRNA genes are less variable because their transcripts are nontranslatable, and most of their sequences are important for biological functions. Protein genes are more variable than these stable RNA genes because many synonymous codon changes and conservative amino acid changes occur without deleterious effects.

Directional Mutation Pressure in Vertebrate DNA

The chromosomes of higher vertebrates reveal mosaic structures consisting of G+C-rich and G+C-poor DNA segments, which correspond to R bands and G bands, respectively (8, 15–17, 86, 100). In accordance with this, the silent positions of codons are high in G and C in the G+C-rich segments and low in the G+C-poor segments (8, 17, 86). DNA replication in eukaryotes takes place in two phases; the R-band and G-band DNAs replicate early and late in the cell cycle, respectively (57, 86). GC pressure might be higher in the early phase, whereas AT pressure might predominate in the late phase, because of the possible use of two different replication systems. For further discussions of directional mutation pressure and the origin of chromosomal compartments, see the recent review by Filipowski (57).

Directional Mutation Pressure in Mitochondria

Generally, one strand of DNA has, on the average, about the same base composition as the other strand, which suggests that directional mutations and their fixations occur evenly on both the strands. This is seen in genomes of many organisms and in certain mitochondria such as those of green plants, yeasts, *Trypanosoma brucei*, and *Drosophila* spp. so that T (or C) contents are nearly equal to A (or G) contents

in the silent sites of codons, regardless of which strand codes for the gene. In vertebrate mitochondria the genomic G+C content of codon silent sites increases in species from 28% (*Xenopus laevis*) to 51% (human) (114). This may be due to selective mutations and/or fixations of one of the two DNA strands. The H strand seems to be exposed as a single strand much longer than the L strand during replication, so that observable mutations accumulate more in the H strand than in the L strand because there is less chance for repair. The selective mutations on the H strand may be A to G, leading to accumulation of C in the major sense (mRNA-like) L strand by replacement of T by C (6). This accumulation of C occurs in the silent sites of codons and also in the D loop, tRNA-like, and rRNA-like sequences in the L strand (6). Thus, GC pressure, which in this case is somewhat different from the usual symmetrical GC pressure, acts on the whole genome. Interestingly, the silent sites of NADH dehydrogenase subunit 6 sequence in the H strand are rich in G because of loss of A, in contrast to the richness of C in other mRNA-like sequences on the L strand (6).

The pattern of such asymmetrical directional mutations seems to be specific in mitochondria of certain lineages. In echinoderm mitochondria, the order of choice in the L strand is A > T = C > G (10, 30, 84, 105), whereas in mitochondria of *Fasciola hepatica* the order is T > G ≫ A > C (64). These asymmetrical directional mutation pressures could be a result of the difference in the degree of repair and proofreading between the L and H strands (10).

Directional Mutation Pressure and Codon Usage

Amino acid codons. The third position of codons is the most variable, because transitions in this position do not result in amino acid replacement, except for codons AUR and UGR, and all substitutions in this position in family boxes are silent. The first position is the next most variable, because the synonymous codons for UUR and CUR (Leu) and for CGR and AGR (Arg) differ in the first position and also many conservative amino acid changes occur by changes in the first position. Changes in the second codon position always result in amino acid replacements, so that functional constraint is the strongest in this position.

Such relationships may be seen in Fig. 2, in which the G+C contents of the first, second, and third codon positions, respectively, from various bacterial species are plotted against the G+C content of their genome. The G+C content of all three positions has a linear positive relationship with genome G+C content, although slopes differ, the order of steepness being the third, first, and second codon positions (160). This order agrees with that of differences in variability or the relative occurrence of neutral mutations at the three codon positions, implying that most of the directional mutations are neutral.

Similar observations were made by Bernardi and Bernardi (16) for prokaryotic, viral, and vertebrate genes. These authors gave an entirely different explanation, i.e., that the bias is the result of environmental pressures (15–17). Our analyses (160) and those by Sueoka (226) do not support their idea (57).

Comparisons of the codon usage of *spc* operons, one of the ribosomal protein gene clusters for *Micrococcus luteus* (genomic G+C content, 74%) (171), *E. coli* (50%) (33), and *Mycoplasma capricolum* (25%) (175) also confirm our view (Table 3). The G+C content of the third codon position is more than 95% in *Micrococcus luteus*, 53% in *E. coli*, and 10% in *Mycoplasma capricolum*. The same bias is seen at

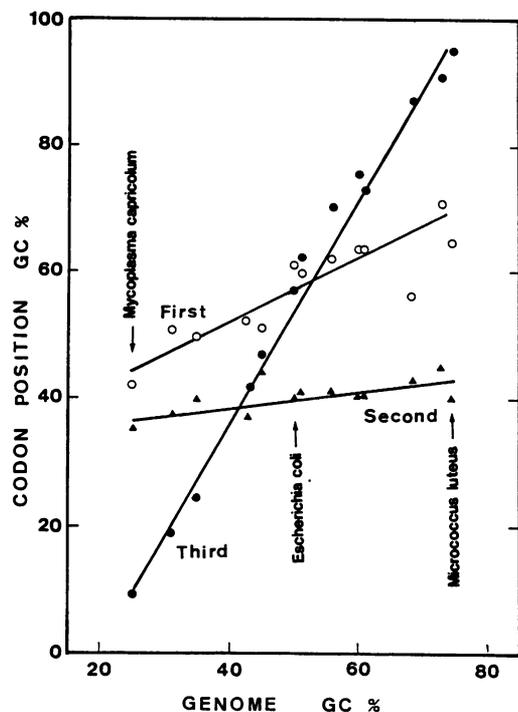


FIG. 2. Correlation of the G+C content between total genomic DNA and the first, second, and third codon positions. Modified from reference 160 with permission.

the first positions. Among six synonymous codons of Leu, *Micrococcus luteus* uses almost exclusively CUC and CUG codons, whereas *Mycoplasma capricolum* prefers UUA codons. A similar situation exists for the Arg synonymous codons CGC and CGG (*Micrococcus luteus*) and AGA (*Mycoplasma capricolum*).

Systematic analyses of the relation between the genomic G+C content and codon usage in other systems are fragmen-

tary. Chloroplasts from tobacco (213) and liverwort (177), mitochondria from yeasts (81, 99), liverwort (178), and maize (178), and some lower eukaryotes such as *Paramecium* spp. (193, 194) and slime mold (235) have genomes with a very low G+C content and reveal codon usage patterns very similar to that of *Mycoplasma capricolum* (3, 158, 175). In animal mitochondria there is also close correlation of G+C content of silent sites of codons, as well as that of the noncoding D loop, with the G+C content of the genomes (114). The codon usage pattern of the *Acanthamoeba castellanii* actin gene (161) closely resembles that of *Micrococcus luteus* (171, 172, 174). In the high-G+C alga *Chlamydomonas reinhardtii*, there is a strong GC influence on silent sites of codons in 14 genes examined (cited in reference 26). Also, in mammals, codon usage is related to local chromosomal G+C content (8, 15-17, 100). These facts suggest that directional mutation pressure similar to that in eubacteria is also exerted in organelles and eukaryotes, although other selection pressures such as that by tRNA seem to modulate the codon usage to various extents.

Initiation and stop codons. In *Micrococcus luteus*, GUG is used as an initiation codon in 10 of 18 protein genes examined (171), whereas AUG greatly predominates over GUG in *E. coli* (232) and is exclusively used in *Mycoplasma capricolum* (175). Most termination codons in *Mycoplasma capricolum* (175) and *E. coli* are UAA (25, 232), whereas *Micrococcus luteus* uses 15 UGA, 2 UAA, and 1 UAG (171). Brown and coworkers (25, 26) compared the usage of three stop codons in prokaryotes with the genomic G+C content and found that the higher the G+C content, the more usage of UGA, and the higher the A+T content, the more usage of UAA. These observations are in accordance with the positive relationship between the G+C content of genomic DNA and that of codons shown in Fig. 2.

Directional Mutation Pressure and Amino Acid Composition of Proteins

Sueoka (224) explored the relationship between the G+C content of DNA and the amino acid composition of total

TABLE 3. Codon usage in *spc* operations from *Micrococcus luteus*, *E. coli*, and *Mycoplasma capricolum*^a

Codon (amino acid)	Codon usage in:			Codon (amino acid)	Codon usage in:			Codon (amino acid)	Codon usage in:			Codon (amino acid)	Codon usage in:		
	L	E	M		L	E	M		L	E	M		L	E	M
UUU (Phe)	1	23	75	UCU (Ser)	5	35	31	UAU (Tyr)	1	16	34	UGU (Cys)	1	5	8
UUC (Phe)	57	53	9	UCC (Ser)	55	24	0	UAC (Tyr)	41	28	9	UGC (Cys)	3	7	3
UUA (Leu)	0	9	137	UCA (Ser)	1	5	70	UAA (stop)	- ^b	-	-	UGA (stop)	-	-	10 ^c
UUG (Leu)	1	14	5	UCG (Ser)	19	5	1	UAG (stop)	-	-	-	UGG (Trp)	7	7	1
CUU (Leu)	0	11	13	CCU (Pro)	2	16	23	CAU (His)	2	14	13	CGU (Arg)	25	93	17
CUC (Leu)	62	13	0	CCC (Pro)	31	2	0	CAC (His)	16	11	7	CGC (Arg)	97	45	1
CUA (Leu)	0	1	12	CCA (Pro)	0	8	42	CAA (Gln)	0	28	76	CGA (Arg)	2	6	2
CUG (Leu)	93	106	0	CCG (Pro)	48	49	2	CAG (Gln)	85	45	1	CGG (Arg)	17	1	0
AUU (Ile)	0	56	156	ACU (Thr)	2	38	74	AAU (Asn)	1	19	89	AGU (Ser)	2	7	22
AUC (Ile)	111	91	13	ACC (Thr)	89	43	2	AAC (Asn)	48	43	16	AGC (Ser)	10	11	4
AUA (Ile)	0	1	28	ACA (Thr)	0	9	48	AAA (Lys)	0	110	213	AGA (Arg)	0	2	66
AUG (Met)	44	57	39	ACG (Thr)	40	11	0	AAG (Lys)	117	53	22	AGG (Arg)	5	0	1
GUU (Val)	0	77	87	GCU (Ala)	3	71	70	GAU (Asp)	7	39	57	GGU (Gly)	28	110	68
GUC (Val)	94	31	2	GCC (Ala)	111	27	0	GAC (Asp)	81	40	8	GGC (Gly)	142	65	0
GUA (Val)	0	44	49	GCA (Ala)	8	53	52	GAA (Glu)	2	73	98	GGA (Gly)	7	5	77
GUG (Val)	108	27	8	GCG (Ala)	50	50	1	GAG (Glu)	98	35	7	GGG (Gly)	15	15	7

^a The numbers show the codon usage per operon in the protein genes in the *spc*-operon from *Micrococcus luteus* (172) (L), *E. coli* (33) (E), and *Mycoplasma capricolum* (175) (M).

^b -, Stop codons not included.

^c UGA is the Trp codon in *M. capricolum*.

TABLE 4. Composition of amino acids assigned by AT and GC codons^a

Amino acid ^a	Molar content (%) in ^b :		
	<i>M. luteus</i> (G+C, 74%)	<i>E. coli</i> (G+C, 50%)	<i>M. capricolum</i> (G+C, 25%)
AT codons			
Phe	3.0	3.8	4.2
Ile	5.8	7.4	10.3
Met	2.3	2.8	2.0
Tyr	2.2	2.2	3.2
Asn	2.6	3.1	5.3
Lys	6.1	8.2	11.8
Total (a)	22.0	27.5	35.8
GC codons			
Pro	4.2	3.7	3.4
Ala	9.0	10.0	6.2
Arg (CGN)	7.4	7.2	1.0
Gly	10.0	9.7	7.6
Total (b)	30.6	30.6	18.2
(b)/(a)	1.39	1.15	0.51

^a AT and GC refer to codons with AA, AT, TA, and TT, or GG, GC, CG, and CC, respectively, in the first two nucleotide positions.

^b The values are molar content (%) of each amino acid in the protein genes in the *spc* operon of *Micrococcus luteus* (172), *E. coli* (33), and *Mycoplasma capricolum* (175).

protein in a wide variety of bacterial species. Ala and Gly levels were higher and Ile and Tyr levels were lower in bacteria with a high G+C content than in their counterparts with a low G+C content. Table 4 compares the average compositions of amino acids assigned by codons which have A or U at both the first and second positions (AT codons) and those assigned by codons having G or C at both positions (GC codons), in eight homologous ribosomal protein genes compared in *Micrococcus luteus*, *E. coli*, and *Mycoplasma capricolum*. The total content of amino acids assigned by AT codons, Phe, Ile, Met, Tyr, Asn, and Lys, increases significantly with increasing genomic A+T content. The total content of amino acids assigned by GC codons, Pro, Ala, Arg (CGN), and Gly, tends to increase with increasing genomic G+C content from *Mycoplasma capricolum* to either *E. coli* or *Micrococcus luteus*, showing that directional mutation pressure affects the amino acid composition of proteins (172, 175, 184). Its effect is presumably exerted mostly on amino acids in near-neutral sites in the protein molecules. A slight but significantly positive correlation between the G+C content of the codon second position and the nonsilent fraction of the first position with the genomic G+C content is the result of such directional amino acid substitutions. The *b/a* values are the most marked findings in Table 4.

Differences of amino acid composition with response to the genomic G+C content were reported in comparisons of two homologous sets of genes (*trpA* and *trpB*), with different G+C contents, in *B. subtilis*, *E. coli*, *Salmonella typhimurium*, and *Pseudomonas aeruginosa*, and 13 homologous sets of genes among various animal mitochondria with different genomic G+C contents (114). The most consistent effects of AT pressure were on Phe, Tyr, and Asn and those of GC pressure were on Ala.

Codon Selection by Other Factors

Codon selection by tRNA. Since the biased use of synonymous codons is strongly related to the bias of genomic G+C content, the species-specific codon usage pattern in eubacteria is determined primarily by AT/GC mutation pressure (Table 3; Fig. 2). However, the synonymous codon usage for Leu, for example, in *E. coli* (50% G+C in the genome), is biased to a considerable extent by other factors. For example, CUG is preferentially used by *E. coli* among six synonymous Leu codons (Table 3).

Ikemura (97-99) found a close correlation between the intracellular abundance of tRNAs and the occurrence of various codons in *E. coli* and *Saccharomyces cerevisiae*, as follows. (i) When synonymous codons for an amino acid are translated by more than two tRNAs having different anticodons (isoacceptor tRNAs), the codons recognized by the abundant tRNA species are used more frequently than those read by less abundant tRNA(s). The abundance of the Leu codon CUG in *E. coli* is accompanied by the largest intracellular amount of the corresponding tRNA^{Leu}_{CAG} among the isoacceptors. The optimal codons are different among organisms, because populations of isoacceptor tRNAs differ.

(ii) When one species of tRNA anticodon reads more than one codon by wobbling, the higher the affinity of the codon to the anticodon (forming more stable base pairs with the anticodon), the higher the usage. For example, two synonymous codons, NNU and NNC, in NNY-type two-codon sets are read by a single anticodon, GNN. Ikemura (98) contends that the NNC codon is more frequently used than NNU, because anticodon GNN has a higher affinity for codon NNC than for NNU. Ikemura (98) concluded that bias of synonymous codon usage is determined by the relative amounts of isoacceptor tRNAs and/or by the nature of the anticodons, acting as a selection pressure.

Inefficient translation of codons by rare anticodons is subject to negative selection. Therefore, the codons used in the highly expressed genes, such as ribosomal protein genes (Table 3), are more efficiently selected against. In effect, optimal codons are strongly selected by tRNAs, as in the case of the CUG Leu codon in *E. coli*. In weakly expressed genes, the major factor determining codon usage pattern is mutation pressure, because the effect of tRNAs is weak owing to a relaxation of selection by tRNA. Figure 3a shows the correlation between the relative usage of NNC and NNU codons in NNY-type two-codon sets and the expression levels of more than 200 different genes in *E. coli* (171, 184, 210). Since the two codons in the sets are each translated by a single anticodon, GNN, the selection by tRNAs is exclusively due to the affinity between codon and anticodon. The higher the expression level of the genes, the higher the usage of NNC codons that have a higher affinity to anticodon GNN; the bias decreases with decreasing expression level and reaches a bottom value of about 40% in *E. coli* and 34% in *B. subtilis*. The bottom value may be taken as an approximate measure of the directional mutation pressure in these bacteria (172, 184).

The above description does not mean that all the silent positions in weakly expressed genes are free from such selection pressure. For example, in both *E. coli* and *B. subtilis*, contents of NNG codons in NNR-type two-codon sets (AAR Lys and GAR Glu) are lower than 35% when translated by a single anticodon (UNN), regardless of the expression levels (Fig. 3b). The NNG content in *E. coli* is much less than that of the G+C content of the whole genome (50%) or that of spacer regions (47%), suggesting the pres-

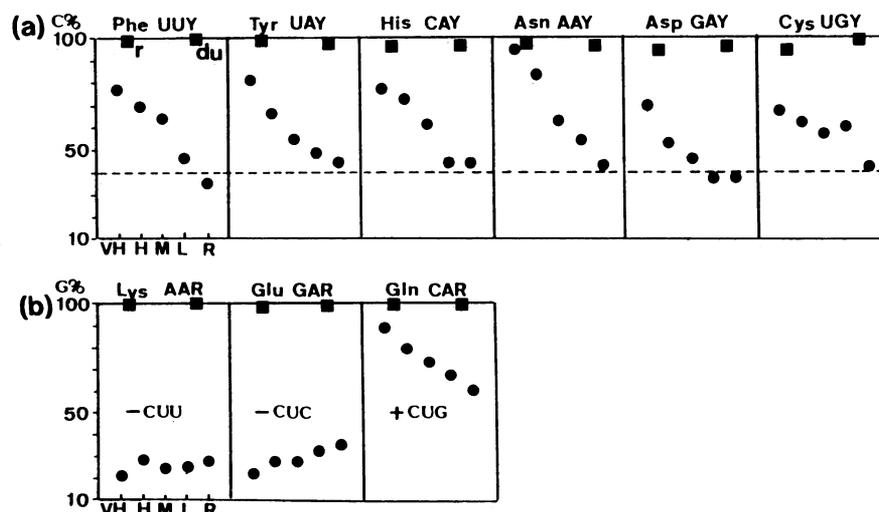


FIG. 3. Codon usage in two-codon sets of various classes of genes in *E. coli* (●) and *Micrococcus luteus* (■). The data for *E. coli* were taken mainly from reference 210, with the addition of *uvrA* and *uvrB* to class L. Abbreviations: VH, very highly expressed genes; H, highly expressed genes; M, moderately expressed genes; L, poorly expressed genes; R, regulatory genes which are expressed poorly. (a) Usage of NNC codons (%) in NNY-type two-codon sets. The dashed line represents the average NNC content of the class R genes of all the NNY-type two-codon sets. (b) Usage of NNG codons in NAR-type two-codon sets. +/- CUN means the presence/absence of CUN anticodon in translation of the respective two-codon sets. In *Micrococcus luteus*, r represents ribosomal and its related protein genes, and du represents *dnaA* and *uvrB* genes. Modified from reference 172 with permission of Oxford University Press.

ence of negative selection by *UNN anticodons to use fewer NNG codons than NNA throughout the genes. This would be because anticodon *UNN pairs mainly with codon NNA but very poorly with NNG, so that even in weakly expressed genes, NNG is not well utilized and cannot be used extensively. On the other hand, the presence of the CNN anticodon (e.g., CUG for Gln), which translates exclusively codon NNG, strongly enhances the NNG usage with a gradient from the highly to the weakly expressed genes (Fig. 3b) (184, 187). The NNG content in the weakly expressed gene is higher (60 to 70%) than the genomic G+C content, suggesting that positive selection by CNN anticodons to use more NNG codons has been exerted even in this class of genes. The development of CNN anticodons would respond to a tendency to use more NNG codons to avoid negative selection (181). In fact, bacteria with high G+C contents have CNN anticodons in NNR sets (181–184).

In the case of synonymous codon selection in family boxes, its correlation with the gene expression level is more difficult to analyze, because of complicated quantitative and qualitative interrelations between two to three isoacceptor tRNAs that usually translate codons in family boxes. Nevertheless, there is a clear tendency that bias by tRNA is stronger in highly expressed genes than in weakly expressed genes (210).

The average G+C content of the third codon position in *Micrococcus luteus* and *Mycoplasma capricolum* is 95 and 9%, respectively. In *Micrococcus luteus*, both highly expressed genes such as ribosomal protein genes and weakly expressed genes such as *dnaA* and *uvrB* have an average G+C content in the third codon position of about 95% (172, 184). The noncoding spacer regions are also strongly biased toward G·C (in *Micrococcus luteus*) or A·T (in *Mycoplasma capricolum*), but the G+C-contents of the spacer are about 20% lower (*Micrococcus luteus*) or higher (*Mycoplasma capricolum*) than those of the third codon positions. This may be explained in one of two ways. (i) In the extremely high- or low-G+C bacteria, the G+C contents of third codon positions are exclusively determined by directional mutation

pressure without much participation of tRNAs, while the spacers are constrained by other factors. (ii) G+C contents of the third positions of both highly and weakly expressed genes are biased by tRNAs to some extent, in addition to directional mutation pressure, while the spacers are almost constraint free. Analyses of the effect of directional mutation pressure and tRNAs on codon usages in two-codon sets from various bacteria favor the latter view (187).

Stop Codon Selection by Release Factors

As described above, the usage of UAA and UGA is correlated with genomic G+C content, whereas UAG is used only rarely throughout the eubacteria. UAA and UAG are recognized by release factor 1 (RF1), and UAA and UGA are recognized by RF2 (see reference 32 for a review). These facts, together with a much lower usage of UAG than of UAA or UGA, suggest that RF2 has a higher affinity than RF1 for UAA, so that most UAAs (and all UGAs) would be recognized by RF2 (25). Thus, much higher usage of UAA and UGA over UAG could be brought about by selection pressure because of the apparently higher efficiency of RF2 over RF1 for translational termination. In *E. coli*, UAA predominates over UGA in highly expressed genes, and the use of UGA increases with decrease of expression level, reaching almost the same usage of UGA and UAA in poorly expressed genes (25). This suggests that in the *E. coli* lineage, a positive selection pressure by RF2 to use UAA has been exerted. In *Micrococcus luteus*, UGA greatly predominates over UAA (171), suggesting that GC pressure is the main force for this usage.

Brown and coworkers (25) suggested that the nucleotide following the stop codon is important for efficient translational termination and that the stop signal may be a tetranucleotide, rather than simply a trinucleotide codon. U is the most highly represented in the nucleotide position following all three stop codons in *E. coli*, whereas A and C are less frequent. The situation is similar in other bacteria having a

moderate or lower genomic G+C content, such as *S. typhimurium*, *B. subtilis*, *Mycoplasma capricolum*, some methanogenic metabacteria, and bacteriophages. This pattern is accentuated in highly expressed genes, but is not as marked in either weakly expressed genes or those that terminate in UAG, the codon recognized only by RF1. This also suggests that RF2 selects for stop signal choice in these bacteria, in addition to directional mutation pressure. However, in *Micrococcus luteus* and *Streptomyces* spp. with high genomic G+C contents, UGA is the most abundant stop codon (UAA is much less abundant [see above]), and 70 to 80% of the nucleotides following stop codons are C or G. In *Micrococcus luteus*, C follows 11 of 18 stop codons examined, with 2 Gs, 1 A, and 4 Us (171, 174). This pattern is completely different from that of other bacteria. This is probably due to the predominance of high GC pressure over the selection by RF, so that not only the stop codon but also the adjacent nucleotide is influenced mainly by mutation pressure.

In eukaryotes, a single RF protein recognizes three stop codons (32). Here again, certain stop codons and nucleotides following stop codons are preferentially used, and the pattern is accentuated in highly expressed genes. The pattern is variable among organisms, but generally the signals UAA followed by A or G and UGA followed by A or G are preferred (26).

VARIABILITY OF tRNAs

Anticodon Composition of tRNA in Organisms and Organelles

The anticodon composition lists vary among organisms and organelles. The smallest (and minimum possible) number of anticodons is 23, but vertebrate mitochondria have only 22 because they have discarded UCU (Arg), and the largest predictable is 45 (eukaryotes). These variabilities result sometimes from codon usage that has been determined primarily by directional mutation pressure and sometimes from the tendency to reduce the genome size (genomic economization). In some cases, certain codons and the corresponding tRNAs disappear from the genome.

Bacteria. Table 5 shows the anticodon species of several eubacteria and fungal mitochondria. A higher genomic G+C content is accompanied by an increase in G or C in silent positions of codons, and since CNN anticodons efficiently translate NNG codons, one would expect more CNN anticodons in eubacteria with a higher G+C content. This is the case. *Micrococcus luteus* (G+C content, 74%) has 13 CNN anticodons (122). Two more (CGA and CCU) are presumed to exist. *E. coli* (G+C content, 50%) has 11 CNN anticodons (127), and *Thermus thermophilus* (121) (G+C content, 69%) has three more anticodons (CAC, CUU, and CUC) (71) that are not found in *E. coli*. *B. subtilis* (G+C content, 43%) has at least six CNN anticodons (231), and *Mycoplasma capricolum* (G+C content, 25%) has only five (4).

Under extremely high GC pressure, such as in *Micrococcus luteus*, UNN codons (U: modified) or LAU (L is lysidine) become only a trace or nonexistent.

The eight family boxes in *Micrococcus luteus* (122), *E. coli* (127), and many other bacteria have seven GNN anticodons, but these are absent from *Mycoplasma capricolum*, and a single UNN anticodon, U, unmodified, pairing with four codons, is present in six of the eight family boxes (4). Table 6 summarizes this pattern. The higher the G+C content of the genome (or the higher the GC pressure), the higher the CNN anticodon content and the lower the UNN anticodon content.

In contrast, the lower the G+C content (or the higher the AT pressure), the lower the CNN and GNN anticodon content.

Since directional mutation pressure is exerted on individual nucleotide sites of DNA, the tRNA content is unlikely to be the direct target of the mutation pressure. We have seen that the synonymous codon choice in extremely high or low G+C bacteria is determined to a considerable extent by directional mutation pressure. Nonetheless, there is a strong correlation between the relative amount of isoacceptor tRNA for an amino acid and the usage of the corresponding synonymous codons in these bacteria (122) (Table 7). It follows that an increase or decrease in the amount of certain tRNA species is, in principle, an adaptive phenomenon affected by codon usage.

Along with an increase in AT pressure, for example, codon NNG in a two-codon set is converted to another synonymous codon NNA, resulting in the predominance of the latter. Since anticodon *UNN pairs mainly with codon NNA but can also pair with codon NNG to some extent, anticodon CNN for codon NNG is no longer needed and disappears from the genome. In contrast, GC pressure converts codon NNA to NNG, accompanied by a decrease in the amount of *UNN and the appearance of anticodon CNN, but anticodon *UNN cannot disappear completely until codon NNA is retired from use, because anticodon CNN does not pair with codon NNA. A similar situation may be seen for anticodons LAU (under GC pressure) and CCG (under AT pressure) (Table 6). Anticodon GNN is necessarily retained, because of its reading both NNU and NNC codons, unless anticodon UNN pairs with all four codons in family boxes (e.g., in *Mycoplasma* spp.).

The anticodon compositions described above are consequences of evolution, but are not tightly linked to the specific phylogenetic lines. They result from directional mutation pressure that has been exerted independently in various lines. For example, in the metabacterial group, which is phylogenetically more related to eukaryotes than to eubacteria (88, 89, 103, 150), anticodon usage follows the "CNN rules"; 14 CNN anticodons have been reported from *Halobacterium* spp. (G+C content, 63%), but only 1 has been found in methanogens (G+C content, 30%) although two obligatory CNN anticodons (CAU and CCA) are presumed to exist in methanogens (181, 184) (Table 8). According to Gupta (68), no INN anticodons are present in halobacteria. The halobacterial code is the same as the eubacterial code except for the absence of ICG (Arg) and the presence of GCG, UCG, and CCG.

Chloroplasts and mitochondria. Chloroplasts have DNA high in A+T (28 to 39% G+C in DNA) and follow the same rule as eubacteria (181); there are only five CNN anticodons, and GNN anticodons are absent from two family boxes in the chloroplast code, which is identical with the universal code. Two family boxes use single UNN anticodons (Table 9).

Most mitochondria, except for those in green plants, have one or two CNN anticodons, one of which is CAU, which apparently translates both Met codons, AUA and AUG. GNN anticodons are absent from all eight family boxes, and these use single UNN anticodons (181) (Table 10).

Resemblance between *Mycoplasma* spp. and mitochondria. The *Mycoplasma* genomes are the smallest for all known free-living organisms (159). They are regarded as degenerate forms of gram-positive bacteria (159). The *Mycoplasma capricolum* genome contains 30 tRNA genes for 29 tRNA species (157), a much smaller number than in the genome of *E. coli*, which has 78 genes for 45 tRNA species (or 41 anticodon species) (127), or than in *B. subtilis*, with at least 51 genes for 31 different tRNA species (231). Only tRNA_{UUU}^{Lys}

TABLE 5. Anticodons in eubacteria and fungal mitochondria^{a,b}

Amino acid (codon)	Anticodon in:	Amino acid (codon)	Anticodon in:	Amino acid (codon)	Anticodon in:	Amino acid (codon)	Anticodon in:
	L E M mt		L E M mt		L E M mt		L E M mt
Phe (UUU)	GAA GAA GAA GAA	Ser (UCU)	GGA GGA	Tyr (UAU)	GUA ⁸ QUA GUA GUA	Cys (UGU)	GCA GCA GCA GCA
Phe (UUC)		Ser (UCC)		Tyr (UAC)		Cys (UGC)	
Leu (UUA)	¹ UAA ¹ UAA ² UAA	Ser (UCA)	UGA UGA	Stop (UAA)		Stop (UGA)	
Leu (UUG)	³ CAA ³ CAA	Ser (UCG)	⁷ UGA CGA	Stop (UAG)		Trp (UGG)	³ CCA ³ CCA ¹ UCA ² UCA ^c
Leu (CUU)	GAG GAG	Pro (CCU)	GGG GGG	His (CAU)	GUG ⁸ QUG GUG GUG	Arg (CGU)	
Leu (CUC)		Pro (CCC)	GGG GGG	His (CAC)		Arg (CGC)	¹¹ HCG ¹¹ HCG ¹¹ HCG
Leu (CUA)	UAG UAG	Pro (CCA)	UGG UGG	Gln (CAA)	⁹ TUG ² UUG ⁴ UUG	Arg (CGA)	U/ACC ^d
Leu (CUG)	⁴ UAG CAG CAG	Pro (CCG)	⁴ TGG CCG CCG	Gln (CAG)	CUG CUG	Arg (CGG)	CCG CCG -
Ile (AUU)	GAU GAU GAU	Thr (ACU)	AGU	Asn (AAU)	GUU ⁸ QUU GUU GUU	Ser (AGU)	GCU GCU GCU GCU
Ile (AUC)		Thr (ACC)	GGU GGU	Asn (AAC)		Ser (AGC)	
Ile (AUA)	- ⁵ LAU ⁵ LAU	Thr (ACA)	UGU UGU	Lys (AAA)	¹⁰ UUU ² UUU ² UUU	Arg (AGA)	-
Met (AUG)	⁶ CAU CAU	Thr (ACG)	UGU ^e ⁴ TGU CGU CGU	Lys (AAG)	CUU CUU CUU ^f	Arg (AGG)	¹² UCU ² UCU ² UCU
Val (GUU)	GAC GAC	Ala (GCU)	GGC GGC	Asp (GAU)	GUC ⁸ QUC GUC GUC	Gly (GGU)	GCC GCC ACC ^g
Val (GUC)		Ala (GCC)	GGC GGC	Asp (GAC)		Gly (GGC)	GCC GCC UCC UCC
Val (GUA)	⁷ UAC	Ala (GCA)	UGC UGC	Glu (GAA)	¹⁰ UUC ² UUC ⁴ UUC	Gly (GGA)	
Val (GUG)	CAC	Ala (GCG)	CGC	Glu (GAG)	CUC	Gly (GGG)	CCC ¹³ UCC CCC

^a Reprinted from reference 4 with permission.
^b Abbreviations: L, *Micrococcus luteus*; E, *E. coli*; M, *Mycoplasma capricolum*; mt, fungal mitochondria; -, probably unassigned codon (most of the anticodons were determined in *Micrococcus luteus* (95, 96, 122). Some more may exist; the nucleoside and its modification at the first anticodon position N for tRNA^{Ser} are unknown. In yeast (*S. cerevisiae* and *T. glabrata*) mitochondria, both AUA and AUG code for Met and CUN codes for Thr. In *T. glabrata* mitochondria, CGN are probable unassigned codons (36, 180). *Mycoplasma capricolum* and yeast mitochondria, UGA codes for Trp. Superscript numbers represent modifications of anticodon first nucleosides as follows: 1, 5-carboxymethylaminomethyl-2-O-methyluridine (cm^mU^m); 2, 5-carboxymethylaminomethyluridine (cm^mU); 3, 2'-O-methylcytidine (Cm); 4, probably modified; 5, 4-amino-2-(N⁶-lysino-1-β-D-ribofuranosyl)pyrimidinium (L, lysidine); 6, N⁴-acetylcytidine (ac^cC); 7, uridine-5-oxyacetic acid (o⁵U); 8, quanosine (Q); 9, probably 2-thiouridine (s²U); 10, 5-methylaminomethyl-2-thiouridine (mm^ms²U); 11, inosine (I); 12, 5-methoxycarbonylmethyluridine (mc^mU); 13, unidentified modification. *E. coli* anticodons, including those from DNA sequences, are from references 127 and 219. Fungal mitochondrial anticodons (*S. cerevisiae*, *N. crassa*, *Aspergillus nidulans*, and *Schizosaccharomyces pombe*), including those from DNA sequences, are from references 29, 50, 69, 75, 126, and 217.
^c *Schizosaccharomyces pombe* mitochondria have one species of Trp tRNA with anticodon CCA (not UCA) that translates codon UGA at a low efficiency in vitro (50).
^d ACCG was reported only from *S. cerevisiae* mitochondria (217). UCG is present in mitochondria of *Schizosaccharomyces pombe* (50) and all animals.
^e From DNA sequence (96).
^f Two different genes for tRNAs with anticodon CAU were reported from mitochondria of *A. nidulans* and *Schizosaccharomyces pombe*. Whether one of them has anticodon *LAU for codon AUA (Ile) is not known.
^g Host nuclear origin (146).
^h ACC (modification unknown) only from *A. nidulans* mitochondria (126).

TABLE 6. Effect of genomic G+C content on eubacterial tRNA anticodon compositions and codon-anticodon pairing patterns^{a,b}

Bacterium ^c (G+C content)	Family box		Family box (Arg CGN)		NNY two-codon set		NNR two-codon set		Ile NNY/A	
	Codon	Anticodon	Codon	Anticodon	Codon	Anticodon	Codon	Anticodon	Codon	Anticodon
<i>M. luteus</i> (74%)	NNU	GNN	CGU	ICG	NNU	GNN	NNA	-	AUU	GAU
	NNC		CGC		NNC				AUC	
	NNA	CGA	NNG		CNN	AUA			-	
	NNG	CGG	CNN							
<i>E. coli</i> (50%)	NNU	GNN	CGU	ICG	NNU	GNN	NNA	UNN	AUU	GAU
	NNC		CGC		NNC				AUC	
	NNA	CGA	NNG		CNN	AUA			LAU	
	NNG	CGG	CNN							
<i>B. subtilis</i> (43%)	NNU	GNN	CGU	ICG	NNU	GNN	NNA	UNN	AUU	GAU
	NNC		CGC		NNC				AUC	
	NNA	CGA	NNG		CNN	AUA			LAU	
	NNG	CGG	CNN							
<i>M. capricolum</i> (25%)	NNU	UNN	CGU	ICG	NNU	GNN	NNA	UNN	AUU	GAU
	NNC		CGC		NNC				AUC	
	NNA		CGA		NNG	CNN			AUA	LAU
	NNG		CGG	-						

^a Reprinted from reference 122 with permission.

^b The table shows the general tendency of anticodon compositions in eubacteria with various genomic G+C contents and does not necessarily indicate that all the boxes and sets actually have these compositions.

^c The bacteria are arranged in descending order of G+C content.

^d -, Deletion. Modification of U is not indicated (see the text).

in *Mycoplasma capricolum* is encoded by two tRNA genes with identical sequences (157). The *Mycoplasma* tRNAs are characterized by a low content of modified nucleosides. Thus, the genomes of *Mycoplasma capricolum* and mitochondria seem to have discarded the genes for many redundant and nonobligate tRNAs and those for many enzymes for tRNA nucleoside modifications (4) as a consequence of genomic economization and AT pressure, resulting in similarity of tRNA usage and the anticodon lists.

Eukaryotes. Forty-four anticodons have been discovered in eukaryotes, and one more, UAA, is presumed to exist (181, 219). The list of anticodons (Table 11) has conspicuous differences from the anticodons of other organisms because it contains eight anticodons with inosine in the first position as noted above. The eukaryotic code carries a full complement of 15 CNN anticodons, because UNN anticodons are modified, presumably to pair primarily with NNA and only weakly or not at all with NNG, as suggested by Guthrie and Abelson (69) for yeast tRNAs, so that NNG anticodons CNN may be required for NNG codons. Indeed, *S. cerevisiae* contains two species of genes for Gln tRNAs with anticodons UUG and CUG, and deletion of the gene for tRNA^{Gln}_{CUG} is lethal (240). This would indicate that the Gln

CAG codon is normally translated by tRNA^{Gln}_{CUG} and not by tRNA^{Gln}_{UUG}. Note that in some eubacteria the NNG codon is translated in the absence of CNN and in the presence of *UNN anticodon, even when U is thiolated uridine (Table 5; Fig. 3b). The CNN rule for bacteria may not be applicable to eukaryotes for this reason. Eukaryotes have larger and more complex ribosomes than prokaryotes do, and this difference may lead to undescribed conformational differences in codon-anticodon pairing.

DISTRIBUTION OF CODE CHANGES IN NUCLEAR AND MITOCHONDRIAL SYSTEMS

Nuclear Code

The occurrence of universal and nonuniversal codes in the nuclear systems of organisms is shown in Fig. 4. *Mycoplasma capricolum* (248) and its related species (35, 53, 101, 102, 128, 198) use UGA as a Trp codon. No altered codes have been found in metabacteria (archaeobacteria).

In the ciliated protozoans *Tetrahymena* spp. (11, 60, 70, 90, 129, 167) and *Paramecium* spp. (31, 60, 193, 194), two universal stop codons, UAA and UAG, code for Gln, in

TABLE 7. Correlation between codon usage and relative amount of isoacceptor tRNAs in *Micrococcus luteus* and *Mycoplasma capricolum*^{a,b}

Amino acid	<i>M. luteus</i>				<i>M. capricolum</i>			
	Codon	Codon usage	Amt of tRNA	Anticodon	Codon	Codon usage	Amt of tRNA	Anticodon
Leu	CUG	100	89	CAG	UUA	100	100	*UAA
	CUY	70	100	GAG	CUN	23	21	UAG
	UUG	1	- ^c	CAA	UUG	5	31	CAA
	CUA	0	-	*UAG				
	UUA	0	-	*UAA				
Ile	AUY	100	100	GAU	AUY	100	100	GAU
	AUA	0	-	*LAU	AUA	21	48	*LAU
Val	GUG	100	100	CAC				
	GUY	84	68	GAC				
	GUA	0	-	*UAC				
Pro	CCG	100	66	CGG				
	CCY	67	100	GGG				
	CCA	<1	-	*UGG				
Thr	ACY	100	100	GGU				
	ACG	52	85	CGU				
	ACA	<1	-	*UGU				
Ala	GCY	100	100	GGC				
	GCG	50	39	CGC				
	GCA	5	-	*UGC				
Lys	AAG	100	100	CUU	AAA	100	100	*UUU
	AAA	<1	-	*UUU	AAG	10	21	CUU
Trp					UGA	100	100	*UCA
					UGG	17	20	CCA
Arg	CGY/Y	100	100	ICG	AGR	100	100	*UCU
	CGG	25	8	CCG	CGY/A	31	33	ICG
	AGG	3	-	CCU	(CGG)	0	0	(CCG)
	AGA	0	-	*UCU				
Gly	GGY	100	100	GCC				
	GGG	12	28	CCC				
	GGA	2	-	*UCC				

^a Data from reference 122 for *Micrococcus luteus* and 246a for *Mycoplasma capricolum*.

^b The relative amount of each codon for an amino acid and the corresponding tRNA (as anticodon) are expressed as the ratio of the most abundant one as 100. All the codons and anticodons are shown regardless of their usage. Blank areas indicate no isoacceptors. The pairing between codon NNU and anticodon *UNN in family boxes and that between codon NNG and anticodon *UNN in two-codon sets were neglected to avoid complexity (* indicates a modified nucleoside). CCG is an unassigned codon in *Mycoplasma capricolum*.

^c -, Not detected.

addition to CAA and CAG. This has also been reported for two other ciliates, *Oxytricha* (82) and *Stylonichia* (77, 78), whereas another ciliate, *Euplotes*, uses UAA as a stop codon (72) and UGA as a Cys codon (148). In protozoology, *Tetrahymena* and *Paramecium* spp. are classified as holotrichous ciliates, whereas *Euplotes*, *Oxytricha*, and *Stylonichia* spp. belong to the hypotrichous ciliates. The molecular phylogenetic tree (79) supports this view. *Euplotes* branched off earlier than *Oxytricha/Stylonichia*. In view of the reported changes, one must suppose that the ancestor leading to these two ciliate groups had used UAR as stop codons and that the change in UAR stop to Gln occurred twice independently, once in the holotrichous line and once in the ancestor of *Oxytricha/Stylonichia* after *Euplotes* branched off. Another change, UGA to Cys, took place in *Euplotes*. However, to establish the assignment of UAR codons in *Oxytricha/Stylonichia*, further investigations, such as identification of the

responsible tRNA, are required. Quite recently, UAA and, probably, UAG were reported to code for Gln in *Oxytricha fallax* (241). Interestingly, two *Acetabularia* species (unicellular green algae which are far from ciliates phylogenetically) also use UAR for Gln codons (205). It seems probable that the same change in the code has occurred independently in at least two distinct lineages.

CUG, a universal Leu codon, is assigned to Ser in the yeast *Candida cylindracea* (123) and its related species (unpublished data).

Mitochondrial Code

Mitochondria can be divided into two categories, plants and nonplants, with respect to code usage. Plant mitochondria use the universal code, whereas nonplant mitochondria

TABLE 8. Known anticodons in metabacteria^{a,b}

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

^a Data from references 184 and 219.

^b Modifications of the anticodon first nucleoside are not indicated. Abbreviations: H, *Halobacterium* spp; M, methanogens. All methanogen anticodons are from *Methanococcus* spp. except GCC (glycine), which was reported only from *Methanobacterium* spp.

use codon UGA for Trp and have many other code changes (Table 12).

In *Schizosaccharomyces pombe* mitochondria, most of the Trp sites in the genes are coded for by TGG, and only three are coded for by TGA in intronic open reading frames (50). The tRNA Trp has anticodon CCA (not UCA) that translates the UGA codon to Trp in vitro at a low frequency. Note that some CNN anticodons apparently translate codon NNA in addition to NNG (see below).

Codon AUA is used for Met instead of Ile (13) in yeast and most animal mitochondria, whereas mitochondria of euascomycetes, protozoa, platyhelminths, and echinoderms use AUA for Ile. Yeast mitochondria are the only mitochondria which are known to use codon CUN for Thr rather than Leu (139).

The universal lysine codon AAA was found to code for Asn in mitochondria of starfish (*Asterina pectinifera*) (84) and *Asterias amurensis* (9), in three species of sea urchins (30, 105, 218), and in mitochondria of platyhelminths, *Fasciola hepatica*, and planaria, as shown by comparing AAA sites in several mitochondrial genes of platyhelminths with the corresponding sites in other mitochondria including echinoderms, yeasts, protozoa, and vertebrates (20, 64, 173).

The amino acid assignments of AGR codons differ in mitochondria of various organisms. Yeasts, euascomycetes, protozoa, and coelenterate mitochondria use these codons for Arg, whereas most invertebrate mitochondria use AGR for Ser.

Both AGR codons are used for Ser in mitochondria of echinoderms (starfish and sea urchins) (9, 30, 84, 105, 218). The same assignment of Ser to AGR was also found in mitochondria of two species of nematodes (243, 244), *Meloidogyne javanica* (244), and squid (*Doryteuthi* species) (212). In *Drosophila melanogaster* and *D. yakuba* mitochondria, AGA codes for Ser rather than Arg as shown by sequence comparisons of *Drosophila* mitochondrial genes with those of other organisms (37, 47, 63). Codon AGG has not been found in any gene of *Drosophila* mitochondria. Mitochondria of vertebrates use AGR codons for stop (human [5], bovine [6], mouse [21], and *Xenopus laevis* [199]).

Codon UAA seems to be used for Tyr rather than stop in planarian mitochondria (20). Planarian mitochondria would have only one stop codon, UAG, because UGA is a Trp codon. In the reported mitochondrial genes of *Fasciola* species, all stop codons are UAG, which is the most rarely used stop codon throughout mitochondrial genes in other species (64). It is possible that UAA is a Tyr codon in *Fasciola* mitochondria as well as in planarian mitochondria.

The changes in the mitochondrial genetic code may be summarized in the form of a phylogenetic tree (Fig. 5) (116). The first change, codon UGA from stop to Trp, occurred in nonplant mitochondria after the separation from plant mitochondria. In the animal line, AGR Arg codons became Ser starting with platyhelminths and continuing as Ser in nematodes, echinoderms, molluscs, and insects until AGR became stop in vertebrates. AAA changed from Lys to Asn in the branch leading to platyhelminths and, as a separate

TABLE 9. Anticodons in chloroplasts^{a,b}

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

^a From references 181 and 219.

^b Modifications of the anticodon first nucleoside are not indicated.

^c -, Stop codons have no anticodons.

^d Reported only for liverwort chloroplasts (176, 177, 188, 230). Not present in tobacco (213, 233) and rice (84a) chloroplast genomes.

event, in echinoderms. UAA stop probably codes for Tyr in platyhelminths. Starting with nematodes, AUA changed from Ile to Met. This change was reversed in platyhelminths and in echinoderms independently.

In the fungal line, changes in the yeast mitochondrial code are AUA from Ile to Met (which occurred independently from that in the animal mitochondrial code) and CUN from Leu to Thr. No changes from the universal code other than UGA to Trp have taken place in euscomycetes.

Time Required for Code Change

When unusual genetic codes were discovered in *Mycoplasma capricolum* and in ciliated protozoans, Grivell (66) proposed that these codes may be "relics of a primitive genetic code that possibly preceded the emergence of the so-called universal code," and that "the ciliates branched off from the primitive eukaryotic ancestor very early in evolution, quite possibly at a time when protein translation was still tolerant of changes in the genetic code." Lewin (138) gave a similar explanation: "At this time (when ciliates branched) the code was still quite error-prone, with termination in particular being an erratic process. Perhaps in fact termination was so leaky that many proteins ended (if not at random sites) at any one of many possible sites. In the major branch of evolution, termination became effective at all the termination codons as the precision of protein synthesis

increased, but in *Mycoplasma* and ciliates only some of the termination codons were retained."

Meyer and coworkers (148) gave a similar explanation for the change of codon UGA from stop to Cys in *Euplotes* spp. The proposals that "translation was tolerant of changes" and that translation was "an erratic process" have no basis in fact. We suggest that an entirely different explanation is more probable and that these altered codes may be of recent origin and derived from the universal code.

Figure 6 shows a phylogenetic tree of the gram-positive eubacterial group constructed from the 5S rRNA sequence comparisons. *Acholeplasma laidlawii*, a member of the class *Mollicutes*, in which the genera *Mycoplasma* and *Spiroplasma* are also classified, does not use UGA for Trp (228). The code change UGA for Trp evidently occurred in the *Mycoplasma/Spiroplasma* lineage after its separation from *Acholeplasma laidlawii* and from their common ancestor (159, 183) and occurred not more than 0.6 billion years ago as estimated from the tree.

The phylogenetic position of ciliates has been placed near the animal kingdom both in the 5S rRNA tree (89) and in the classical trees, whereas some other molecular trees show ciliates as branching off somewhat earlier. However, with any branching point, it must be after prokaryotic emergence. In the prokaryotic genetic code, UAR is used as stop and UGA is used as stop or, rarely, as Trp. Furthermore, the change UAR to Gln from stop also occurred independently in a green

TABLE 10. Anticodons in vertebrate mitochondria^a

Amino acid (codon)	Anticodon	Amino acid (codon)	Anticodon	Amino acid (codon)	Anticodon	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)		Stop (UAA)	^b	Trp (UGA)	UCA
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)		Lys (AAA)	UUU	Stop (AGA)	-
Met (AUG)		Thr (ACG)		Lys (AAG)		Stop (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)	

^a From references 13, 14, 21, and 181.

^b -, Stop codons have no anticodons.

alga, *Acetabularia* spp. (see above). Differences in nucleotide sequences between the new tRNAs that translate UAR codons and the other Gln tRNA, with anticodon UmUG, in *Tetrahymena* spp. are 13 and 11 nucleotide substitutions in tRNA^{Gln}_{UmUG} when compared with tRNA^{Gln}_{UmUA} and tRNA^{Gln}_{CUA}, respectively, and 4 substitutions in tRNA^{Gln}_{UmUA} compared with tRNA^{Gln}_{CUA} (70). These differences may be a rough clue to the time elapsed since evolutionary divergence. Differences between yeast and rat or wheat tRNAs with the same anticodon are 13 to 15 substitutions, and differences between *Drosophila* spp. and vertebrates are 5 to 12 substitutions. From these facts, and from our discussion above, one must conclude that this code change, stop to Gln, took place well within the time of existence of eukaryotes (118).

The change of CUG, a universal Leu codon, to Ser in *C. cylindracea* is even more striking. *C. cylindracea* branched off quite recently from other species of hemiascomycetes such as *Saccharomyces* and *Torulopsis* spp., in all of which CUG is a regular Leu codon. Since codon CUG is for Leu and not for Ser in the species belonging to the euascomycetes (*Aspergillus*, *Penicillium*, and *Neurospora* spp.), sharing a common ancestry with hemiascomycetes (89), the change of CUG to Ser from Leu is most probably a recent event, nearly 0.15 billion years ago as calculated from the 5S rRNA phylogenetic tree (89).

It is well established that a eubacterial species entered the cytoplasm of eukaryotic cells and took up endosymbiotic residence to become an organelle, the mitochondrion. Dur-

ing mitochondrial evolution, a number of codon reassignments occurred. The oldest is probably the change of codon UGA for Trp from stop (Fig. 5), since all mitochondria except those of green plants use UGA for Trp. Other code changes took place in different lineages at different points after the emergence of eukaryotes (20, 116, 117, 173, 180, 185, 186). It is thus evident that all nonuniversal codes in mitochondria are derived from the universal code.

CODON CAPTURE

The term codon capture refers to a change in the meaning of a codon. It takes place by the following steps: first, a codon disappears from translated sequences; second, it reappears with a changed meaning. We discuss below the evolutionary events that produce such disappearances and reappearances.

If the change in meaning of a codon were to occur directly, the tRNA that translated the codon to amino acid A would have to change its aminoacylation site to react with aminoacyl-tRNA synthetase for another amino acid, B. B would be inserted instead of A at all sites previously occupied by the codon for amino acid A. As a result, the nucleotide sequences would not change, but amino acid sequences in proteins would. Such direct replacements of one amino acid by another throughout proteins would be disruptive. Even in mitochondria where only a small number of proteins are

TABLE 11. Anticodons in eukaryotes^{a,b}

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

^a From references 181 and 219.

^b Modifications of the anticodon first nucleoside are not indicated except for I (inosine). UUA and CUA are anticodons for Gln in ciliated protozoa and presumably in *Acetabularia* spp. Anticodons in this table are those found in various eukaryotes; this does not mean that all of them exist in a single species.

^c -, Stop codons have no anticodons.

^d *, Anticodons presumed to exist.

synthesized, such replacements could be lethal at a site essential for function of a protein.

A more plausible possibility is that the codon reassignment follows the codon capture rule, in which a series of changes involved is nondisruptive, i.e., neutral. This theory proposes a temporary disappearance of an amino acid codon (or stop codon) from coding frames by conversion to another synonymous codon and a loss of the corresponding tRNA that translates the codon. This produces an unassigned codon. For a stop codon, the release factor must simultaneously change

so as not to recognize the stop codon. The codon reappears later by conversion of another codon and emergence of a tRNA that translates the reappeared codon with a different assignment. As a result, the nucleotide sequences change while the amino acid sequences of proteins do not change.

Unassigned Codons

We propose that the production of an unassigned codon is an intermediate step of the codon reassignment during

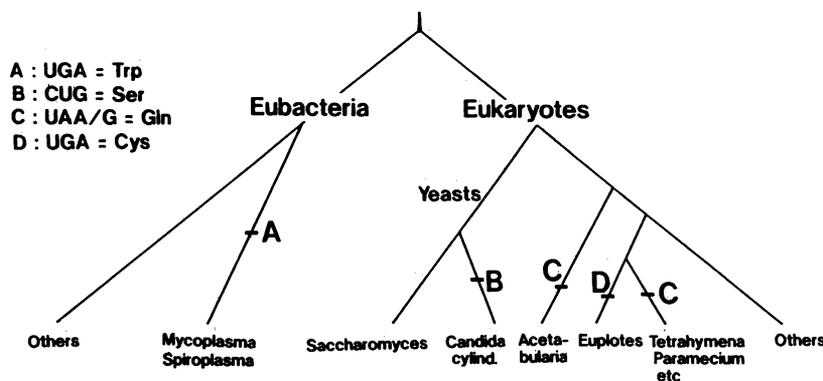


FIG. 4. Evolution of the genetic code in different organisms.

TABLE 12. Genetic code variations found in nonplant mitochondria

Organism	UGA Stop	AUA Ile	AAA Lys	AGR Arg	CUN Leu	UAA Stop	Examples (references)
Vertebrates	Trp	Met	Lys	Stop	Leu	Stop	Human (5), bovine (6), rat (61), mouse (21), chicken (48), toad (199)
Arthropods	Trp	Met	Lys	Ser (AGA only)	Leu	Stop	<i>Drosophila</i> spp. (37, 63), mosquito (92, 93)
Echinoderms	Trp	Ile	Asn	Ser	Leu	Stop	Sea urchin (30, 105), starfish (9, 84, 104, 218)
Molluscs	Trp	Met	Lys	Ser	Leu	Stop	Squid (212)
Nematodes	Trp	Met	- ^a	Ser	-	-	<i>Ascaris suum</i> (243, 244), <i>Caenorhabditis elegans</i> (243, 244)
Platyhelminths	Trp	Ile	Asn	Ser	Leu	Tyr?	<i>Fasciola</i> spp. (64), planaria (20)
Coelenterates	Trp ^b	-	-	Arg	-	-	Hydra (244), <i>Metridium senile</i> (244)
Yeasts	Trp	Met	Lys	Arg	Thr	Stop	<i>Saccharomyces cerevisiae</i> (81, 94, 141, 142, 208), <i>Torulopsis glabrata</i> (1)
Eucomycetes	Trp	Ile	Lys	Arg	Leu	Stop	<i>Aspergillus nidulans</i> (65, 162, 234), <i>Neurospora crassa</i> (27, 28)
Protozoa	Trp	Ile	Lys	Arg	Leu	Stop	<i>Trypanosoma brucei</i> (80), <i>Paramecium</i> spp. (195, 196, 209)

^a -, Not determined.

^b Reference 242a.

evolution of the genetic code. It can occur for NNA codons and for UUR/CUN Leu and AGR/CGN Arg, because the disappearance of Leu and Arg codons is not harmful if they are converted to synonymous codons. In family boxes, when there are two anticodons INN and CNN (and sometimes +UNN in eukaryotes), codon NNG (read by anticodon CNN) can become unassigned. In eubacteria, only CGG may become unassigned in this way, because I at the first anticodon position occurs only in the Arg CGN family box.

In family boxes where anticodon GNN, +UNN, and sometimes CNN are involved, codon NNG would not become unassigned, because anticodon CNN can disappear, but +UNN is usually converted to UNN pairing with all four codons in the family box (see above). Also, NNG codons in two-codon sets would not become unassigned. Codon NNG can disappear together with anticodon CNN by conversion to codon NNA (read by anticodon *UNN), but it will reappear by back-mutation with the same meaning because of the presence of anticodon *UNN that translates the reappeared NNG codon in addition to codon NNA, at least in bacteria. Codons NNG for Leu and Arg are exceptions, for the reasons mentioned above. In eukaryotes, however, codons NNG might become unassigned, if eukaryotic anticodons *UNN and CNN specifically pair with codons NNA and NNG, respectively (69). Codon NNG may mutate to NNA with deletion of anticodon CNN, and codon NNG, when it appears by back-mutations, will be selected against

because of the absence of anticodon CNN and the presumed inability of anticodon *UNN to pair with codon NNG.

Directional mutation pressure, in its extreme, completely removes certain codons (183) and the corresponding tRNA, so that the codon will become unassigned. Note, however, that the composition and amount of tRNAs more or less shift the usage of codons up or down. This effect, in addition to mutation pressure, would play a role in the disappearance of certain codons (187). Some examples are given below.

(i) In *Mycoplasma capricolum*, tRNA^{Arg}_{CCG}, which is needed for translation of the CGG codon, is absent (4), and no CGG codon has been found in the genes examined (4, 175). In accordance with these findings, translation of synthetic mRNA containing in-frame CGG codons in a *Mycoplasma* cell-free system ceases just before CGG is reached. The synthesized peptide stays in the P site of ribosomes in the form of peptidyl-tRNA, leaving the A site empty, and is not released from ribosomes (168). Perhaps strong AT pressure has converted all CGG codons to the synonymous codons CGU, CGA, or AGR by silent mutation with disappearance of tRNA^{Arg}_{CCG}, and CGG has become an unassigned codon.

(ii) In mitochondria of the yeast *Torulopsis glabrata*, neither Arg CGN codons nor a tRNA Arg able to translate them has been found (36). All Arg codons are AGR.

In mitochondria of *S. cerevisiae*, a close relative of *T. glabrata*, mitochondrial genes are A+T rich, averaging 33%

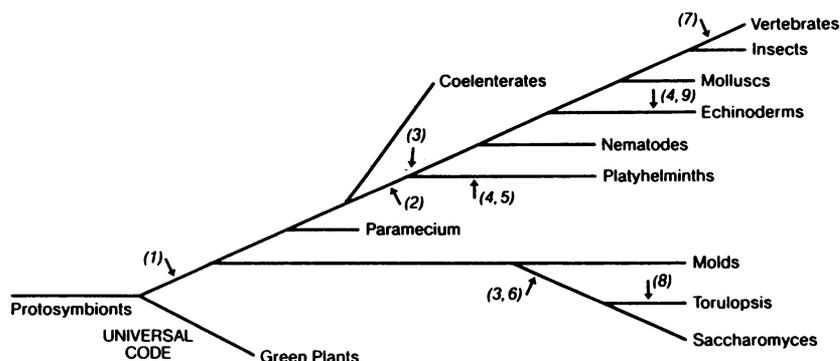


FIG. 5. Evolution of the genetic code in nonplant mitochondria. (1) UGA, Trp; (2) AGR, Ser; (3) AUA, Met; (4) AAA, Asn; (5) UAA, Tyr?; (6) CUN, Thr; (7) AGR, stop; (8) CGN, noncoding; (9) AUA, to Ile from Met. The point of change (3) is not definite. Modified from reference 116 with permission.

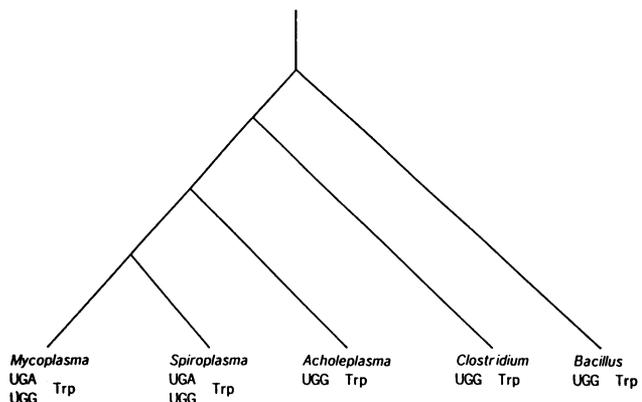


FIG. 6. Branching order of several gram-positive eubacteria and their Trp codon usage. Reprinted from references 159 and 183 with permission.

G+C, and Arg is encoded almost entirely by AGR. Most Arg CGN codons appear to have mutated silently to Arg AGR. The tRNA^{Arg}_{ACG} responsible for translation of rare CGN codons has an exceptionally high A+T content (18% G+C) (180). It is possible that AT pressure causes codons CGN to fall into almost complete disuse and that the functional constraints on the corresponding tRNA gene are thereby reduced, allowing it to accumulate mutations that would otherwise be deleterious. Because of AT pressure, most of these mutations are towards A or T, resulting in unusually A+T-rich base composition in *S. cerevisiae* tRNA^{Arg}_{ACG}. *T. glabrata* mitochondrial tRNA Arg would have been lost in a similar way, so that CGN codons become unassigned (180).

(iii) In *Micrococcus luteus*, six codons, UUA (Leu), CUA (Leu), AUA (Ile), GUA (Val), CAA (Gln), and AGA (Arg), all ending with A, are completely absent in 5,516 codons examined (171, 172, 174). The tRNAs with anticodon UNN (U: modified) or LAU responsible for reading these codons are also not detected (122) (Table 7), suggesting that along with almost complete conversion of NNA codons to NNC or NNG by GC pressure assisted by positive selection of these codons by anticodons CNN and GNN (Fig. 3a) (172, 187), UNN anticodons become very scarce or nonexistent and abundant NNG or NNC codons are read by CNN or GNN anticodons. A few NNA codons that appear at this stage via mutations will be immediately selected against, because of the lack of sufficient UNN (or LAU) anticodons to translate them. In vitro translation with the *Micrococcus luteus* extract of synthetic mRNA containing various Arg codons in frame showed that all codons except AGA were translated. Even codon AGG, whose usage is only 2% among all Arg codons, was translated to some extent, while translation of AGA was completely absent (unpublished data). It is possible that at least some of the undetected NNA codons in *Micrococcus luteus* are unassigned codons.

The occurrence of unassigned codons implies that the codon tables for some life-forms have fewer than 64 codons. Other mechanisms that include directional mutation pressure are also possible.

Possible Mechanisms of Code Changes

Genomic economization reduces the genome size by discarding many nonessential genes. This process, often accompanied by AT pressure, is prominent in mitochondrial

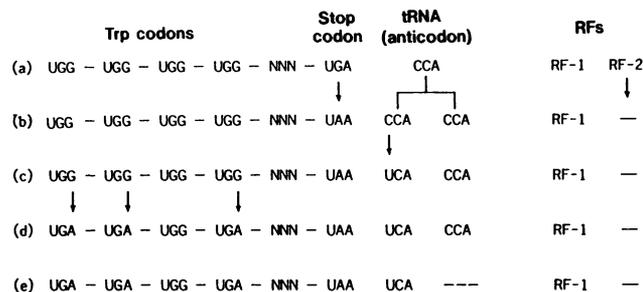


FIG. 7. Evolution of the UGA Trp codon in *Mycoplasma* spp. Stages (d) and (e) are represented by *Mycoplasma capricolum* and *Mycoplasma pneumoniae/genitalium*, respectively. Modified from references 159 and 183 with permission.

evolution. A gene for a tRNA may disappear gradually from the genome owing to progress of genomic economization. Because of this disappearance, the corresponding codon is converted to other synonymous codons by negative selection and becomes unassigned.

Changes of RF may also be responsible for one or two stop codons becoming unassigned by converting to another stop codon(s), if the corresponding RF is deleted or its activity is abolished. This process seems, in some cases (e.g., in *Mycoplasma* spp. [see below]), to be impelled by directional mutation pressure, resulting in disappearance of a stop codon followed by disappearance of the corresponding RF activity, but it is also possible that mutations accumulated in the RF so that it gradually loses its activity for the corresponding stop codon(s) (e.g., in ciliated protozoans or in planarian mitochondria [see below]). As long as an RF recognizes a stop codon, the stop codon cannot become unassigned and therefore cannot be used for an amino acid, unless there is a special mechanism, as with selenocysteine (see below).

Unassigned (nonsense) codons, such as CGG in *Mycoplasma capricolum*, fail to bring about release of the polypeptides from ribosomes, presumably because of their lack of the ability to interact with RF, and therefore they cannot function as stop codons.

Codon capture. The unassigned codons detected here or those presumed to be produced could conceivably be reassigned to the same amino acid (or stop signal) or to another amino acid depending on the appearance of a new tRNA (and RF).

Other examples of code changes in various systems will be discussed in terms of the codon capture theory, assuming that an unassigned codon was produced in one of the three ways discussed above. The unassigned codon capture may be implemented in one of the following procedures: (i) by a change in the anticodon; (ii) by a change to a different amino acid in aminoacylation of a tRNA molecule, while retaining the same anticodon; or (iii) in mitochondria, by a change in codon-anticodon pairing (117).

The first procedure occurred in changing codon UGA from stop to Trp; codon CUG from Leu to Ser; codons UAR and UGA from stop to Gln, Cys, or Tyr; and codon AAA from Lys to Asn. The second procedure occurred in changing codon CUN from Leu to Thr. The third procedure was in changing codon AGR from Arg to Ser, AGR from Ser to stop, and anticodon CAU Met from pairing only with codon AUG to pairing with AUA and AUG.

UGA from stop to Trp. Evidence clearly indicates that

65% with an average of about 40%. This suggests that switching of directional mutation pressure (from A · T to G · C or G · C to A · T) has occurred within a short time. It is thus possible that although *C. cylindracea* is now very high in genomic G+C content (61%), its ancestor was under AT pressure, so that codons CUG (and CUC) would have disappeared, having been converted to the synonymous Leu codon CUA, CUU, or UUR. Indeed, CUG is a rare Leu codon in low-G+C yeasts. The tRNA_{CAG}^{Leu} also would have disappeared from the genome by relaxation of functional constraints because of the disappearance of CUG codons, and so CUG would have become unassigned. The tRNA_{CAG}^{Leu} was not found in *C. cylindracea*, whereas the Leu tRNA with anticodon IAG must remain to pair with codons CUY and CUA, and was actually found (249). Next, in the lineage of *C. cylindracea*, GC pressure would have replaced AT pressure, and some sequence changes took place in one of the Ser tRNAs. Indeed, a Ser tRNA species with anticodon CAG exists in *C. cylindracea*. This tRNA can be directly derived from what previously was tRNA_{IAG}^{Ser}, because there is a C insertion in the intron-containing tRNA gene to create anticodon CAG by splicing (249). The tRNA can translate codon CUG as Ser in a cell-free system by using synthetic mRNA. Thus, codon CUG could reappear as Ser, pairing with the new tRNA_{CAG}^{Ser}. Some of the present CUG Ser codons in *C. cylindracea* and its related species could have been derived from UCN or AGY Ser codons. However, these Ser codons cannot be converted to CUG by a single mutation. Presumably, some (e.g., UCG) codons first mutated to Pro CCG or Leu UUG and thence to Ser CUG. Other CUG Ser codons could have been derived from various codons for other amino acids. Indispensable Ser sites (e.g., the active sites in an indispensable protein gene, only a single copy of which exists in an organism) would not have been directly involved in this reassignment. We infer that most or all present-day *C. cylindracea* Ser CUG codons arose after the code change through individual mutations of various other codons rather than through a large-scale direct replacement of Leu by Ser.

Mitochondrial Code

As a result of genomic economization, nonplant mitochondria contain a limited set of tRNAs (22 species in metazoan mitochondria and 24 in yeast mitochondria) and most metazoan mitochondrial tRNAs have unusual structures with lack of the D-loop-T-loop interaction (5, 6). Cytoplasmic tRNAs are not imported into metazoan mitochondria (200). Certain features characteristic of mitochondrial tRNAs are largely responsible for genetic code changes in mitochondria.

The tRNAs of plant mitochondria, in which the universal code is used, have the usual cloverleaf secondary and L-shaped tertiary structures in common with bacterial and eukaryotic cytosolic tRNAs. Some of the tRNAs are the descendants of those in the original symbiotic eubacteria that are ancestors of the present mitochondria. Some others have features in common with those of chloroplast tRNAs, and still others are of nuclear origin transported from the cytoplasm (116, 143, 144).

In the following changes in the mitochondrial genetic code, production of unassigned codons was brought about either by directional mutation pressure or by genomic economization. Unassigned codon capture took place in one of the ways outlined above.

UGA from stop to Trp. The change of UGA from stop to

Trp in *Mycoplasma capricolum* was described above. The same or similar sequence of events probably occurred in the evolution of nonplant mitochondria.

AAA from Lys to Asn; AUA from Met to Ile as a reversal of Ile to Met; UAA from stop to Tyr. The deduced amino acid assignments, AUA for Ile and AAA for Asn in mitochondria of echinoderms (84, 104) and platyhelminths (20, 173) and UAA probably for Tyr in planarian mitochondria (20), could have all resulted from similar codon captures. It is reasonable to speculate that in mitochondria of the ancestor of these animals, codons AUR Met and AAR Lys were translated by tRNA_{CAU}^{Met} and tRNA_{UUU}^{Lys} (or CUU; see below), respectively, and UAA was a stop codon recognized by RF1. The three codons AUA, AAA, and UAA would have disappeared from coding sequences following a loss of function of the corresponding tRNAs and a loss of the UAA-recognizing function of RF1. The following two possibilities are suggested for loss of assignment of these three codons.

Ancestral mitochondrial tRNA Lys in echinoderms and platyhelminths could formerly have had anticodon CUU capable of translating both AAA and AAG, because some animal mitochondrial tRNAs with anticodon CNN, such as tRNA_{CAU}^{Met} in mammalian and *Aedes* mitochondria and tRNA_{CUU}^{Lys} in *Aedes* mitochondria, seem to recognize both codons NNA and NNG (52, 91, 219). Along with an accumulation of mutations in the Lys or Met tRNA, their functions gradually changed so as to translate only codon NNG. In planarian mitochondria, RF became specific for stop codon UAG. As a result, the NNA codons AAA, AUA, and UAA would have been replaced in coding frames by synonymous NNG codons and become unassigned.

The second possibility is that GC pressure in the genomes of ancestors of these animal mitochondria (see above) resulted in mutation of the three NNA codons to NNG. Indeed, the G content of the codon third positions is markedly higher for echinoderms and *Fasciola* spp. than for yeasts or *Drosophila* spp. Also, tRNA_{UUU}^{Lys} for AAA and AAG codons mutated to tRNA_{CUU}^{Lys}, which would translate only codon AAG. tRNA_{CAU}^{Met} underwent structural changes to become specific for codon AUG and RF became specific for anticodon UAG in planarian mitochondria (20).

In either case, the NNA codon that disappeared was later captured by an amino acid that had previously been assigned only to an NNY two-codon set. In echinoderm mitochondria, there is only one species of Ile tRNA (anticodon GAU in DNA) and Asn tRNA (anticodon GUU in DNA). These tRNAs must be responsible for reading NNA codons in addition to NNY. This could have followed a conversion of anticodon GNN to INN. INN would pair with all three codons NNU, NNC, and NNA for the same amino acid, AUY/A for Ile (echinoderms and planaria), AAY/A for Asn (echinoderms, *Fasciola* spp., and planaria), and UAY/A for Tyr (planaria).

CUN from Leu to Thr. The reassignment of CUN codons from Leu to Thr in yeasts may have followed the steps described below (180).

There are two tRNA Thr species in mitochondria of yeasts (139); one is tRNA_{UGU}^{Thr}, translating universal ACN Thr codons, and the second is tRNA_{UAG}^{Thr}, translating CUN codons which are present only in yeast mitochondria. The latter tRNA probably originated by mutation of tRNA_{UAG}^{Leu} rather than by mutation of tRNA_{UGU}^{Thr} (215).

There are also two different threonyl-tRNA synthetases in the yeast nuclear genome which charge these two threonyl-tRNAs separately. One of them, MST-1, aminoacylating

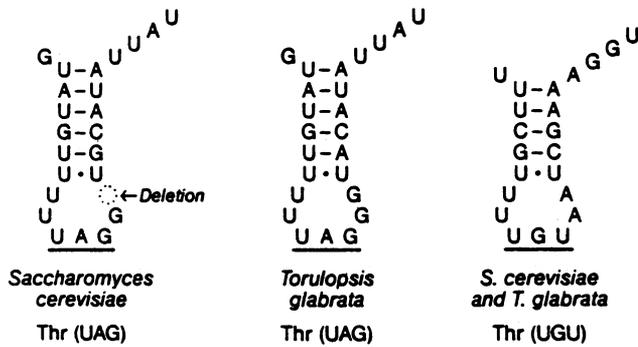


FIG. 9. Similarity in the anticodon arm structures of Thr-tRNA (UAG) and Thr-tRNA (UGU) from two yeast species, *S. cerevisiae* and *T. glabrata*. Reprinted from reference 180 with permission.

only tRNA^{Thr}_{UAG}, has a high sequence similarity to threonyl-tRNA synthetases for tRNA^{Thr}_{UGU} (189).

The G+C content of the mitochondrial genome of *S. cerevisiae* is 18%, including less than 10% G+C in silent positions of the codon, so that most Leu codons are UUR. With the disappearance of CUN Leu codons by conversion to UUR, a relaxation of selective constraints on tRNA^{Thr}_{UAG} occurred. Indeed, tRNA^{Thr}_{UAG} is quite high in A+T (79%), suggesting accumulation of A·T by functional relaxation under AT pressure. This tRNA lost its ability to interact with leucyl-tRNA synthetase and acquired the ability to interact with MST-1 (181), so that its gene was not removed from the genome. When CUN codons subsequently appeared in reading frames from mutations of various codons, they were translated as Thr by the new tRNA^{Thr}_{UAG} that evolved from tRNA^{Leu}_{UAG}.

Some of the present-day CUN Thr codons in yeast mitochondria could have been originally derived from ACN Thr. However, ACN cannot be converted to CUN by a single mutation, and it presumably mutated to AUN or UUR via UCN and thence to CUR. Other CUN Thr codons in yeast mitochondria could have been derived from various codons for other amino acids. Indispensable ACN Thr sites would not have been directly involved in this reassignment.

Because the tRNA^{Thr}_{UAG} of *S. cerevisiae* has eight bases in its anticodon loop, it was once considered that the extra base insertion into the loop would change the amino acid acceptor specificity from Leu to Thr (139). However, tRNA^{Thr}_{UAG} of *T. glabrata* is able to form a normal 7-base anticodon loop by adding an extra U·U pair in the anticodon stem, as shown in Fig. 9 (36). Accordingly, it was proposed that *S. cerevisiae* tRNA^{Thr}_{UAG} would have an anticodon loop structure similar to that of *T. glabrata* (180) if a U·U pair is added to the anticodon stem, and a 6-base anticodon loop is permitted (1).

AUA from Ile to Met. The amino acid assignment of codon AUA differs among mitochondria of various species. The model in Fig. 5 indicates that switching of AUA Ile to Met took place twice independently: in the ancestor leading to yeasts, and in the ancestor to metazoa (116). Reading of both AUA and AUG codons as Met seems to be performed by an unusual codon-anticodon pairing with tRNA^{Met}_{CAU}.

Yeast mitochondria. *Aspergillus* and *Neurospora* spp. share a recent common ancestor with yeasts, and in these fungal mitochondria AUA is the most frequently used Ile codon, as might be expected from their high A+T content in DNA. Many of the AUA Ile sites in mitochondrial genes of *Aspergillus* and *Neurospora* are replaced by AUU (mostly AUU) Ile in yeast mitochondria. No Ile AUA sites in

Aspergillus and *Neurospora* mitochondria correspond to AUA Met in yeast mitochondria (186).

Because the mitochondrial code has evolved from the eubacterial code, translation of AUA as Ile and AUG as Met in *Aspergillus* and *Neurospora* mitochondria may be performed by tRNA^{Ile}_{LAU} and tRNA^{Met}_{CAU}, respectively. In fact, genes for three species of tRNAs with CAU anticodons have been found in *Aspergillus nidulans* mitochondria, the initiator Met tRNA, the primary elongator Met tRNA, and possibly tRNA^{Ile}_{LAU} specific for codon AUA. On the other hand, yeast mitochondria have only elongator and initiator Met tRNAs, both with the CAU anticodon (13). These facts suggest that tRNA^{Ile}_{LAU} disappeared before or concomitantly with reassignment of AUA from Ile to Met. One possibility is that the gene for tRNA^{Ile}_{LAU} would have been gradually deleted as a result of genomic economization, whereas that of tRNA^{Met}_{CAU} remained unchanged. During this time, AUA Ile codons all mutated to AUU Ile and other codons, because of the constraint imposed by shortage of the tRNA^{Met}_{CAU}, and codon AUA became unassigned. Elongator tRNA^{Met}_{CAU} then underwent structural changes so as to translate both AUA and AUG as Met. At this stage, AUA could appear in reading frames from mutations of various codons, including AUG Met, and would be translated by the altered tRNA^{Met}_{CAU}.

Metazoan mitochondria. Metazoan mitochondria contain only one species of tRNA with anticodon CAU, presumably functioning for both initiation and elongation. This indicates that one species of tRNA with anticodon CAU was deleted in the early stage of metazoan mitochondrial evolution. It is difficult to decide which Met tRNA was deleted, but the reassignment of AUA from Ile to Met would have occurred in a way similar to that in yeast mitochondria.

As to the unusual pairing of anticodon CAU with AUA in addition to AUG, Sibley and colleagues (216) speculated that an extra unpaired nucleotide within the base-paired T stem may influence the pairing properties of the CAU anticodon of yeast Met-m-tRNA by conferring the ability to translate both codons AUG and AUA. Unpaired nucleotides in the T stem similar to this are found in Met tRNAs of many animal mitochondrial genes (219).

AGR from Arg to Ser or stop. AGA is the most abundantly used Arg codon in mitochondria of simpler eukaryotes, presumably as a result of high AT pressure. Most of these AGR Arg sites that occur in yeasts or *Trypanosoma* spp. are occupied by CGN Arg codons (and not by Ser codons) in corresponding mitochondrial genes in metazoan species, in which AGR is used as either Ser or stop codons. In contrast, the AGR Ser sites in metazoan mitochondria are replaced by UCN Ser, AGY Ser, and others and not by Arg codons. Thus, no direct replacements of AGR Arg by AGR Ser have occurred (185).

The complete replacement of AGR mostly by CGN Arg codons and the absence of tRNA^{Arg}_{UCU} in metazoan mitochondria suggest that conversion of AGR to CGN took place in mitochondria of the metazoan ancestor, even in the presence of AT pressure, as a result of selective constraints following gradual removal of the gene for tRNA^{Arg}_{UCU} with the progress of genomic economization. By this time, all AGR codons were converted to other codons, mainly to CGN, and codon AGR became unassigned.

Following this, the structure of tRNA^{Ser}_{GCU}, pairing primarily with AGY codons, has undergone structural changes so that it translates AGR codons as Ser. AGR codons then appeared in the reading frames from mutations of AGY Ser codons and other codons. AGR codons were thus captured either by tRNA^{Ser}_{GCU} (in most invertebrates) or by tRNA^{Ser}_{UCU}

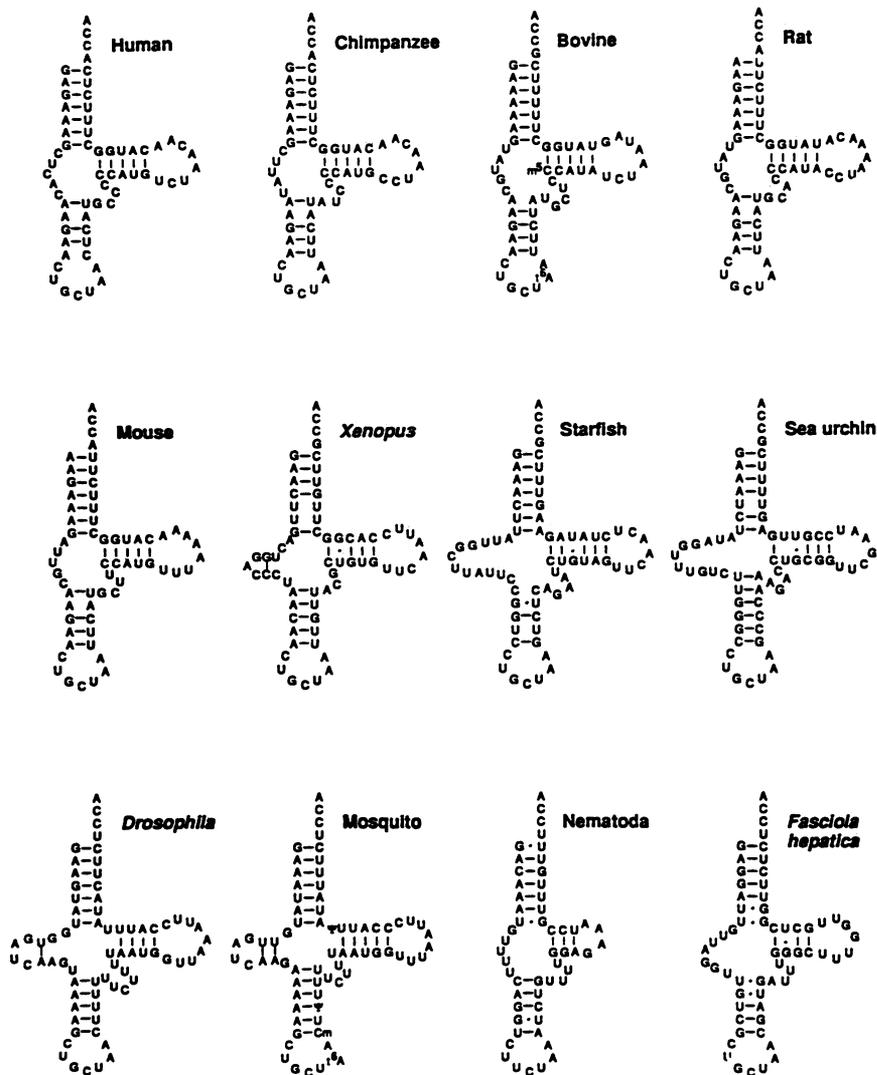


FIG. 10. Cloverleaf structures of Ser-tRNAs specific for AGY codons found in various animal mitochondria. Nucleotide sequences of the tRNAs from bovines and mosquitoes are determined at the RNA level, and those from other animals are deduced from their DNA sequences.

(in *Ascaris suum*), which seems to translate all AGN codons as Ser by four-way wobbling.

During vertebrate mitochondrial evolution, AGR stop codons were apparently created from UAG stop codons by deletion of the first nucleotide U and by the use of R as the third nucleotide that had existed next to the ancestral UAG stop in UAGR (185). RF of rat mitochondria recognizes only UAR stop codons and not AGR in vitro (134). Further studies are needed to clarify this.

How have tRNAs^{Ser}_{GCU} acquired an additional ability to translate AGR codons and then lost this ability in vertebrate mitochondria? These tRNAs have unusual secondary structures, as shown in Fig. 10, in which the D arm is lacking or incomplete. This truncated D arm could be responsible for G · R pairing, in addition to G · Y pairing, between the anticodon and codons. Because tRNAs^{Ser}_{GCU} of all animal mitochondria have identical anticodon loops, there should be some structural differences in regions other than the anticodon loop depending on which codons, AGY (vertebrates) or AGN (invertebrates), are translated as Ser. The

truncated D arm is considerably larger in mitochondria of invertebrates than mitochondria of vertebrates. *Fasciola hepatica* mitochondria use AGR as Ser in addition to AGY, and yet their tRNA^{Ser}_{GCU} has a small pseudo-D arm (244). Therefore, the size of the D arm would not be responsible for the differential translation of AGY and AGN in metazoan mitochondria. Another feature is the G · C pair at the bottom of the anticodon stem, which is present for tRNA^{Ser}_{GCU} of mitochondria of invertebrates but not in those of vertebrates, in which its place is taken by the A · U pair (Fig. 10). Whether this structure is responsible should be examined.

These hypothesized changes in nuclear and mitochondrial codes all provide for the disappearance of a codon followed by its reappearance with a different assignment. In almost every case, the amino acid sequences of the proteins involved are unchanged, so the reassignments are nondisruptive, but sometimes sequences are changed temporarily (yeast mitochondrial CUN Leu to Thr [139] and *C. cylindracea* CUG Leu to Ser [123]).

UGA AS A SELENOCYSTEINE CODON

One of the most remarkable properties of coding is the occasional incorporation of selenocysteine (Secys) in polypeptide synthesis in both prokaryotes and vertebrates. Secys has been sometimes termed the "21st amino acid." It occurs as the active center of a few enzymes, and it is inserted by a special tRNA, with anticodon UCA pairing with UGA, which is also a stop codon. Enzymes containing Secys have not been detected in green plants.

Pinsent (192) found that selenite was needed for the production of formate dehydrogenase activity by *E. coli*. Her observation on the relation of selenium to this enzyme lay dormant for 32 years.

In 1957, selenium deficiency was produced experimentally in rats and chickens (190, 207). This led to the use of selenium as a dietary supplement for farm animals and to much interest in the biochemistry of selenium. Glutathione peroxidase, an enzyme of vertebrates, was shown to contain selenium (58, 85, 170, 201) in the form of Secys (38).

Genes of glutathione peroxidase in mice and of formate dehydrogenase in *E. coli* were found to contain an internal TGA codon that evidently did not carry out its usual function of terminating protein synthesis (34, 251). Instead UGA coded for Secys, as shown by the correspondence of TGA in the gene (34) to Secys in bovine glutathione peroxidase, whose complete amino acid sequence had been determined by Günzler and coworkers (67).

Notably, Secys cannot replace cysteine in cysteine tRNA. In this respect, Secys is unlike selenomethionine, which can become aminoacylated to methionine tRNA and is then incorporated into thiolase of *Clostridium kluyveri* via codon AUG to replace some of the methionine at sites that are apparently randomized (73).

Secys occurs in specific positions in the polypeptide chains of glycine reductase, formate dehydrogenase, and hydrogenase in bacteria, including *E. coli* (76, 136, 147).

Böck and colleagues (23a; see also reference 58a) list four genes "whose products are required for the formation of selenoproteins in *E. coli*." The first is *selC*, which codes for a tRNA with anticodon UCA. This tRNA (tRNA Sec) is aminoacylated with serine by an enzyme, seryl-tRNA ligase, and, significantly, this enzyme also charges all the other *E. coli* seryl-tRNAs. The aminoacylated tRNA Sec (seryl-tRNA Sec) is the substrate for the gene product of the second gene, *selA*. The gene product, Secys synthase, contains covalently bound pyridoxal phosphate. This enzyme catalyzes a reaction which leads to the conversion of seryl-tRNA Sec into selenocysteyl-tRNA Sec by two steps. The first is the dehydration of seryl to form the enzyme-bound acrylyl-tRNA intermediate. The gene product of the third gene, *selD*, with Mg^{2+} , ATP, and selenide than catalyzes the 2,3 addition of H_2Se to the acrylyl double bond, resulting in selenocysteyl-tRNA-Sec. The Se for the reaction is provided by the product of the third gene, *selD* (136). The fourth gene, *selB*, provides a specific translation factor that inserts Secys into protein and replaces elongation factor Tu for this purpose. Elongation factor Tu is the regular elongation factor for other amino acids.

The mechanism of synthesis of selenocysteyl-tRNA in mammals is similar to that seen in *E. coli* (151).

The complexity of the procedures for aminoacylating tRNA Sec with Secys, by means of converting serine to Secys, is quite unlike the usual procedure for charging other tRNAs. We have suggested (112) that Secys, in contrast to selenomethionine and cysteine, cannot become directly at-

tached to an uncharged tRNA molecule because of the high reactivity of the HSe^- group.

Secys is present in glutathione peroxidase of mammals and chickens, in selenoprotein P, and in type 1 iodothyronine deiodinase in rats (19, 59, 132, 133, 170, 191, 227).

The structures and sequences of the tRNA from rat liver reported by Hatfield et al. (74) differ markedly from those of the tRNA Sec of *E. coli* described subsequently by Leinfelder et al. (137) in that the aminoacyl stem of the latter had 8 bp instead of the usual 7 bp. Only 29 sites, 11 of which are invariant nucleotides, are identical in the 87 sites compared in the two tRNAs, and this identity could occur by chance.

The gene for type 1 iodothyronine deiodinase (ITD), obtained from rat liver, has a TGA codon at position 382, corresponding to amino acid residue 126 (19). The gene was translated in rabbit reticulocyte lysates into a full-length protein terminating with UAG at nucleotide 778. In contrast, a wheat germ lysate translated the gene with termination at the UGA codon. The authors explained this by the absence of Secys tRNA in plants (132) so that, in the wheat germ system UGA at position 382 functioned to terminate polypeptide synthesis. There was no sequence similarity between ITD and glutathione peroxidase. The discovery of Secys in ITD explained the essential role of Se in thyroid hormone action.

Secys at codon 126 in ITD could be replaced by cysteine with the enzyme retaining 20% of its ITD activity.

Safran et al. (202) described type 2 ITD from rat brain. It did not contain Se. They propose that types 1 and 2 arose independently in evolution.

Zinoni et al. (252) found that the reading of UGA as Secys in the mRNA of formate dehydrogenase (*E. coli*) depends on the presence and identity of 40 bases 3' to the internal UGA codon. Apparently, these bases form a hairpin loop next to the UGA codon (252), and this loop is instrumental in the translation of UGA as Secys.

In contrast to prokaryotes, a specific 3' untranslated segment was needed for incorporation of Secys at UGA internal sites of proteins in eukaryotes (18). This segment was found in both rat and human mRNAs for type 1 iodothyronine 5'-deiodinase. These sequences were not required for expression of a mutant in which Cys replaced Secys. Similar stem-loop-forming sequences were predicted for glutathione peroxidase. Mutations in these sequences reduced or eliminated translation of iodothyronine 5'-deiodinase. The sequences were several hundred nucleotides downstream from the coding region of iodothyronine 5'-deiodinase.

From these findings, it is concluded that there is a difference between prokaryotes and eukaryotes in the mechanism that controls incorporation of Secys at the internal UGA site of mRNA for Secys-containing proteins. The essential secondary structure in prokaryotes is included in the coding region. This might be interpreted as showing that use of Secys is an evolutionary innovation that has been introduced separately into prokaryotic and eukaryotic lineages and therefore is of recent origin. Perhaps it was preceded by the use of Cys instead of the present Secys.

Motchnik and Tappel (152, 153) found Se in several locations within the Se-containing subunit of selenoprotein-P. They showed (153) that this subunit of selenoprotein-P contained about 9 mol of selenium; this was the first selenoprotein found to contain more than one Se per polypeptide subunit, and they proposed that it might function for selenium transport.

Amino acid analysis accounted for all the Se content, 7.5

TABLE 13. Anticodon-codon pairing and incorporation of Secys

Anticodon in tRNA Sec	Codon at Secys position in mRNA	Activity (%)	Reference
UCA	UGA	100	12
UGA	UGA	<1	12
UGA	UCA	100	12
GCA	UGA	28-32	12
UCA	UGU, UGC	5-10	252
UCA	UGC	ca. 100 ^a	252
UCA	UGU	20	19

^a *E. coli* MC4100 with wild-type *rpsL* gene.

± 1.0 atoms per molecule, of the protein as Secys (197). The authors suggest that selenoprotein-P may be “a protective agent in the plasma against free radicals or other reactive molecules.”

More recently, Hill and coworkers (83) cloned and sequenced cDNA of selenoprotein-P from rat liver. It contained 10 TGA codons for Secys in its open reading frame prior to the TAA stop codon, and 9 of the 10 were in the terminal 122 amino acids. In two cases, two TGA codons were separated by a single other codon. The authors found that “determination of the possible folding patterns” of the mRNA “predicted a stable stem-loop structure [252] following only 1 of the 10 UGAs. This does not support the stem-loop hypothesis” of Zinoni and coworkers (252), but see the discussion by Berry et al. (18) cited above. Selenoprotein-P was formed rapidly following administration of ⁷⁵Se to a rat; it may be involved in supplying Se to the tissues, including the brain.

Modifications of Secys-Incorporating Systems

Baron and coworkers (12) made changes in the anticodon of tRNA Sec and measured their activity in incorporating Secys into *E. coli* formate dehydrogenase at site 140 (FDH_N activity). Shortening the 8-bp amino acid acceptor stem of the tRNA by 1 bp produced inactivation. Replacement of the extra arm by that of a serine-inserting tRNA abolished the activity.

A change of the anticodon from UCA to UGA not surprisingly stopped the insertion of Secys, but if the target UGA codon was changed to UCA, pairing with anticodon UGA, Secys was inserted at site 140. Alteration of anticodon UCA to GCA (cysteine anticodon) enabled some activity (30%) to take place, presumably by partial pairing between anticodon GCA and codon UGA. Nonstandard wobble rules evidently are in effect.

As stated above, anticodon UGA was active when codon UGA at position 140 was changed to UCA (Table 13). None of these changes produced incorporation of Secys at “non-selenocysteine sites” in the protein.

Diamond and colleagues (49) used a bovine liver preparation and found that tRNA Ser with Trp anticodon CmCA, instead of UCA, would bind to termination codon UGA in vitro but showed “little or no binding in response to a variety of other codons including those for tryptophan [which is surprising] and serine.” Anticodon CmCA would be expected to pair only with UGG Trp. Judging from Table 13, codon-anticodon pairing can take place in vivo at position 140 if anticodon sites 2 and 3 form standard complementary pairs with the first two codon sites, even when the

first anticodon site does not complement the third codon site.

Lee and coworkers (133) state that tRNA Ser, with anticodon CmCA, “most certainly reads specialized UGA codons which designate selenocysteine in proteins.”

These results are summarized in Table 13, which lists the various anticodon-codon pairs that have been examined experimentally.

Secys in Evolution

Leinfelder and coworkers (137) proposed that UGA existed in the anaerobic world as a codon for Secys, not for stop, and that in the aerobic world codon UGA could have “acquired other functions such as its more familiar role in termination.”

In the “anaerobic world” of archaic organisms that presumably lived in the ocean, Secys could have been stabilized because of the absence of oxygen and may have been used in many enzymes. However, selenium is far less abundant (0.004 ppm) than is sulfur (884 ppm) in seawater (237), so the use of Secys may have been restricted by the rarity of selenium.

The use of Secys in proteins in terms of evolution may be either a survival or a novelty. Leinfelder and coworkers (137), as summarized above, have discussed it as a survival. The opposite viewpoint, that it is a novelty, is supported by the following arguments. Some enzymes containing Secys are found to function, although with a diminished efficiency, if Cys is substituted for Secys (251). This could indicate that Secys is a recent evolutionary improvement and that the more primitive form of the enzymes used Cys, with codons UGU and UGC and anticodon GCA. Perhaps, at this stage, UGA was a codon for Trp.

The next stage in evolution would be when UGA changed from Trp to stop (see below). To introduce Secys in protein sequences, it would be necessary to develop the elaborate system that exists today, including the use of a Ser tRNA to carry Secys, and the appearance of three enzymes. These modifications have the appearance of sophisticated evolutionary innovation rather than the survival of primitive systems.

The formate dehydrogenase subunits of *E. coli* and *Methanobacterium formicicum* (214) have the following sequences in the vicinity of their active sites, which are Secys and Cys, respectively:

136
-Cys-Ala-Arg-Val- Secys-His-Gly-Pro-Ser- Val-Ala-Gly-Leu-
128
-Cys-Ala-Arg-Leu- Cys -His-Gly-Pro-Thr- Val-Ala-Gly-Leu-

The evolutionary relationship is obviously close. Secys is encoded by UGA, and Cys is encoded by UGC. The *E. coli* enzyme has four to five times higher specific activity than the *Methanobacterium formicicum* enzyme does. Both enzymes have very similar surroundings for their active sites, yet one (*E. coli*) functions with Secys and the other (*Methanobacterium formicicum*) functions with Cys. Böck and Stadtman (23) ask: “Did the Cys-containing enzyme derive from the Secys-containing one or vice versa?”

Why was UGA not captured by Secys with elimination of UGA stop, as has occurred in several systems for UGA Trp? If this had happened, Secys would be clearly “the 21st amino acid.” As it is, Secys is introduced into proteins by modification of a Ser tRNA and by insertion at UGA sites that still retain the property of chain termination in some

TABLE 14. Molar ratios of some amino acids formed by abiotic synthesis^a

Amino acid	Molar ratio in amino acid from:	
	Electrical discharge	Murchison meteorite
Glycine	100	100, 98
Alanine	180	36, 44
α -Amino- <i>n</i> -butyric acid	61	19, 18
Norvaline	14	14, 3
Valine	4.4	19, 10
Norleucine	1.4	2
Leucine	2.6	4
Isoleucine	1.1	4
Alloisoleucine	1.2	4
Proline	0.3	22, 16
Aspartic acid	7.7	13, 5
Glutamic acid	1.7	20, 18
Serine	1.1	

^a From references 46 and 238.

circumstances (19). Codon UGA is unique as having four meanings: stop, tryptophan, Secys, and cysteine (148).

Extensive reviews of selenium biochemistry have been published by Stadtman (220–223).

EARLY HISTORY OF THE CODE

A comprehensive and detailed evaluation of amino acids as participants and nonparticipants in the genetic code was made by Weber and Miller (238). Their bluntly stated opinion was, “the basic reason that amino acids were used, was their abundance in the primitive ocean”; this possibility for each amino acid received primary attention in their evaluation. They also considered functions of amino acids in proteins, the stability of the amino acid and its peptides, and its stability to racemization when attached to tRNA.

Amino acids (racemic except, of course, glycine) are produced by electrical discharge in mixtures of inorganic compounds, such as methane, ammonia, hydrogen, carbon dioxide, nitrogen, and water, by the well-known Miller (149) procedure and modifications of it. Amino acids of extraterrestrial origin were found in the carbonaceous Murchison meteorite (46) (Table 14). Some of them are found in proteins (proteinous), whereas others are nonproteinous (Table 14), and just why the latter are not used in protein synthesis has long been debated. Weber and Miller (238) place particular emphasis on “the puzzling absence of alpha-amino-*n*-butyric acid, norvaline and norleucine” from the genetic code in view of the fact that these three amino acids are produced by abiotic synthesis in amounts greater than several of the proteinous amino acids. Weber and Miller (238) speculate that they may have been coded for and subsequently discarded. Norleucine can be incorporated into proteins (7, 39, 179), and it can charge methionyl-tRNA.

Glutamine and asparagine are unstable and are not found in abiotically synthesized mixtures, but Gln is formed by amidation of Glu-tRNA Gln, which is charged with Glu by glutamyl-tRNA synthetase in several organisms (206). Therefore, Gln and perhaps also Asn could easily have been introduced into protein synthesis by such a mechanism even though they are not formed by abiotic synthesis.

The fact that eight of the alpha-amino acids most commonly produced by abiotic synthesis (Table 14) are in the universal code is strong evidence that the primitive coding

system drew on these amino acids, as proposed by Weber and Miller (238).

It is often hypothesized that the coding process started with a few, such as 7 or 8, amino acids and that others were added by various means until the total of 20 was reached. Wong (245) has produced a mathematical argument in favor of the total stopping at 20, and, although Secys is used in a few proteins, it is, because of the difficulties for its use, “the exception that proves the rule” (see above).

Evolutionary Origin of tRNA

All known tRNAs have so many properties in common, including their function, that it is evident that they are homologous. Gene duplication is common: for example, there are two yeast serine tRNAs, anticodon IGA, that differ by only three transitional nucleotide substitutions. In contrast, evolution in sequences of tRNAs in general has proceeded so far that Holmquist and coworkers (87) found that the average divergence between pairs of tRNAs for different amino acids had reached an equilibrium between forward and backward mutation of variable sites. The secondary structures of tRNAs except for some nonplant mitochondrial tRNAs (Fig. 10) (219) have the following characteristics in common: (i) four helical regions containing, respectively 7, 4, 5, and 5 bp in which all pairs but two are of variable composition; (ii) four loops of unpaired nucleotides, of which the third loop is variable in size and may contain up to 17 nucleotides; (iii) a total of 23 invariant or semi-invariant nucleotides. The tertiary structure, as depicted on the cover of *Nucleic Acids Research*, is constant. In this respect, tRNA is reminiscent of the hemoglobins, which have a highly variable sequence and a constant tertiary structure.

Given these properties, it seems that tRNAs are descended from a single ancestral molecule. What does this imply for the genetic code? Eigen and coworkers (54) performed sequence analyses of tRNA and concluded that “about one-third of the present tRNA sequence divergence was present at urkingdom level at about the time when archaeobacteria separated from eubacteria, about 2.5 ± 0.5 BYA [billion years ago].” The remaining one-third of tRNA sequence divergence, which preceded the urkingdom divergence, would therefore place the earliest tRNA, i.e., the origin of the genetic code, at not more than 3.8 ± 0.5 BYA. Their conclusion was that “the genetic code is not older than, but almost as old as, our planet.”

This still does not solve the problem of what was the first anticodon in the first tRNA. Was it CCC for glycine? It is difficult to conceive of polyglycine having functions that are associated with proteins. Probably few evolutionists would challenge the conclusion of Eigen et al. (54) that the genetic code is not older than the Earth.

Perhaps the first tRNA gene rapidly duplicated and mutated in its anticodon and aminoacyl receptor sites so as to enable several amino acids to be used in protein synthesis. Another possibility is that aminoacylation of early tRNAs was nonspecific, so that the original tRNA could code for several different amino acids. A remnant of nonspecificity still is present; valine is incorporated instead of isoleucine at a level of about 0.03% (140). It is easy to suggest that early proteins were full of “mistakes” in sequence, but this is an evasive argument for reconstructing primitive codes by “explaining everything.” A third possibility is that a simpler adaptor molecule preceded tRNA in evolution.

Codons in Primitive Systems

There were always 64 codons, because this is a mathematical property of four variables, taken three at a time. In primitive codes they may have all been divided among a small number, less than 20, of amino acids, or there may have been a few assigned codons and many unassigned codons, as suggested by Lehman (135). Wong's coevolution hypothesis (245) proposes that at first the 64 codons were distributed among seven amino acids: Glu, 16; Asp, 14; Val, 10; Ser, 9; Phe, Ala, Gly, 4 apiece; stop, 3. These were then reassigned, except for Ala, Gly, and stop, to give the present universal code. The initial list is forbiddingly high in acidic amino acids and has no basic amino acids, which are important in the amphotericity that is so characteristic of proteins.

Wong's proposal expands the primitive list through known biosynthetic pathways that led to the formation of Tyr from Phe; Cys and Trp from Ser; Lys, Thr, and Asn from Asp; Arg, Pro, and Gln from Glu; His from Gln; Met and Ile from Thr; and Leu from Val. Wong points out that many of the new amino acids have codons with "domains contiguous with those of their precursors." However, this may also be true for some nonprecursors. Although many introductions of "new" amino acids into protein sequences would be disruptive of organisms (in the same manner that Crick [42] has pointed out in discussing the frozen-accident model), it is also true that all theories of evolution of the code that require an increase in its amino acid content must allow for the introduction of new amino acids at some stage.

Weber and Miller (238) did not offer a biosynthetic proposal for addition of various proteinous amino acids that are not listed in Table 14, such as lysine, arginine, tryptophan, and methionine. Instead they offered the possibility that such amino acids had been produced by various prebiotic syntheses.

The table of the code (Table 1) looks as if the simplest units, and perhaps the earlier units, are the family boxes, which later became subdivided. On the basis of this possibility, Jukes (107) suggested that an earlier (archetypal) form of the code consisted entirely of family boxes and that eight quartets in the present universal code had kept their original early assignments as family boxes.

He proposed that the archetypal code consisted solely of 16 family boxes. One of these contained four stop codons (UAN), and the other 15 were each occupied by a single amino acid with its four codons pairing with a single tRNA molecule. When the vertebrate mitochondrial code was discovered in 1979, it was found that there were indeed such tRNA molecules. Their anticodons have the formula UNN, in which U is the first anticodon base and pairs with U, C, A, and G in the third position of codons. He hypothesized that this code expanded by dividing some of the family boxes into pairs of two-codon sets. This process continued, probably over a long period, until the number of amino acids in the code increased to 20, as shown in Table 15 (109). For this expansion, the single tRNA genes had to duplicate, and under GC pressure one of the duplicates underwent a mutational change from UNN to GNN in the anticodon. For example, if glutamine was in a family box with anticodon UUG, the anticodon would duplicate to UUG and GUG, and the tRNA with anticodon GUG would be captured by histidine (Table 1). The other anticodon, UUG, became modified to *UUG so that it would not continue to pair with all four CAN codons, but only with CAA and CAG.

In proposing this expansion of the code, it is tempting to

guess which of the two amino acids in a pair of two-codon sets was the earlier, but there is no experimental basis for making such decisions. For example, either histidine or glutamine may have been the predecessor in the histidine-glutamine box. The family box containing four stop codons UAN became divided between UAY (tyrosine) and UAR (stop).

When this expansion was complete, there should have been two codons for Met, AUA and AUG, and two codons for Trp, UGA and UGG.

The next step was dividing the two codons for Met between Ile and Met and the two codons for Trp between Trp and stop. This took place under GC pressure that led to development of anticodons CAU and CCA from UAU and UCA, respectively. CAU paired only with codon AUG, and CCA paired only with UGG, so that codons AUA and UGA were untranslatable and became unassigned. Next, tRNA Met CAU duplicated and one of the duplicates underwent structural changes so that its anticodon was modified to LAU (where L is lysidine) upon emergence of the lysidine-forming enzyme. Its tRNA accepts Ile and recognizes only codon AUA. AUA was thus captured by Ile, and AUA codons appeared by mutations of AUU. In the eukaryotic lineage, LAU was discarded and was replaced by IAU, pairing with isoleucine codons AUU and AUA. At this stage, unassigned codon UGA was captured by stop via mutations of UAA to UGA under GC pressure and evolution of the release factor to recognize UGA as well as UAA. It is of interest that UGA is the predominant stop codon in G+C-rich organisms.

We postulate that GC pressure was important in the evolution of the genetic code (early code) after it had first acquired 20 amino acids (181). A clue to this is the vertebrate mitochondrial code, which contains only 22 anticodons and is able to translate all the amino acid codons except those for AGA and AGG. We propose that the early code was similar to the vertebrate mitochondrial code and that it expanded under GC pressure so that CNN anticodons, pairing only with NNG codons, were produced following duplication of UNN anticodons. In support of this proposal, it has been found that the number of CNN anticodons used by microorganisms increases with increasing G+C content of their DNA (CNN rule). By this expansion, the code became less subject to error, because G pairs strongly with C.

In mitochondria, the amount of DNA is small as a result of genomic economization and the number of tRNA genes has decreased, presumably as a reversal of the steps in evolution of the code as discussed above. To reduce the anticodon list, CNN anticodons were discarded.

EPILOGUE

In some recent textbooks, nonuniversal codes, such as UGA Trp in *Mycoplasma* spp. and UAR Gln in ciliated protozoans, are dealt with simply as exceptions. Such a treatment gives the impression that nearly 100% of organisms use the universal code. Is it true? It is believed that there are more than 10 million species of organisms now living on Earth, all of which are derived from a single pool of the ancestor. Despite such an enormous diversity, until recently only a handful of standard organisms, such as *E. coli*, *B. subtilis*, *S. cerevisiae*, *Drosophila* spp., and vertebrates, had been examined genetically. Along with the development of molecular phylogenetic studies and the rapid progress of gene technology, our interest has begun to focus on various odd organisms other than the standard. As a

TABLE 15. Anticodons of proposed archetypal code^a

Anticodon	Amino acid						
UAA	Phe or Leu	UGA	Ser	GUA	Tyr	UCA	Cys or Trp
UAG	Leu	UGG	Pro	UUG	His or Gln	UCG	Arg
UAU	Ile or Met	UGU	Thr	UUU	Asn or Lys	UCU	Arg or Ser
UAC	Val	UGC	Ala	UUC	Asp or Glu	UCC	Gly

^a Each anticodon pairs with four codons for a single amino acid, except for GUA, which pairs only with AUU and AUC.

result, nonuniversal codes have been discovered at a relatively high incidence. Codon UGA Trp has been found in seven *Mycoplasma* species and related bacteria; at least two kinds of nonuniversal code are independently used in ciliated protozoans; the same code change was found in two different organismic lines, ciliated protozoans, and unicellular green algae; a yeast line uses a still different code. All nonplant mitochondria that have been examined use nonuniversal codes, which are more or less characteristic for each line. It is remarkable that mitochondria from one species use more than two nonuniversal codons: six in yeasts, four or five in many invertebrates, and four in vertebrates. Thus, nonuniversal codes are widely distributed in various groups of organisms and organelles.

The nonuniversal codes are not randomly produced, but are derived from the universal code as a result of a series of nondisruptive changes. As discussed in detail above, many codons potentially have their assignments changed by disappearing from mRNA sequences and then reappearing by codon capture. The codon capture is driven by directional mutation pressure, modification of RF, or genomic economization, producing an unassigned codon which can be reassigned upon appearance of a new tRNA or RF recognizing the codon.

Some of the principles governing the evolution of the code, upon which our codon capture theory is based, have come to light with organisms having very high genomic A+T or G+C contents. With such extreme organisms, the existence of an unassigned codon has been demonstrated and mechanisms of the code change from UGA stop to Trp have been uncovered. It should be stressed that such laboratory data would never have been obtained with the standard organisms alone.

New changes will be discovered when more organisms or organelles are examined. A hint to finding new changes in nuclear codes would be to find organisms having high genomic G+C or A+T content. Indeed, all the organisms in which changes are found carry such genomes.

The knowledge now available and described in this review has made it possible to deduce the early evolution of the code in a much more reasonable way than before.

The use of UGA as a Secys codon is one of the active fields in the study of evolution of the code, and it is still under much discussion. Generally, a codon cannot have two meanings simultaneously. UGA is a remarkable exception; in-frame UGA is read as Secys, and UGA at the termination site functions as a stop codon. How is in-frame UGA recognized by Secys tRNA? Is Secys a relic in the "anaerobic world," or is it a recent evolutionary improvement?

RNA editing, which we have not discussed in this review, is another important field, and its mechanism as well as its evolutionary meaning will hopefully be explored in the near future. Of much interest is the finding that liverwort mitochondria do not edit their mRNA in the manner used by phanerogams (178).

ACKNOWLEDGMENTS

We are grateful to Shigeyuki Yokoyama and Takashi Ohama for discussions and to Yayoi Takahashi for devoted editorial work. We thank Carol Fegté for editing and typing the manuscript, Roxanne Short and Nancy Laleau for typing, and David Collins for criticism and discussion.

Research in the laboratories of S. Osawa, K. Watanabe, and A. Muto was supported by a grant-in-aid for scientific research on the priority area Diversity of the Genetic Code, from the Ministry of Education, Science and Culture of Japan (Grants 01656004, 02238104, and 03222104). Research in the laboratory of T. H. Jukes was supported by NIH grant RO1 HG00312 and NASA grant NAGW-1521.

REFERENCES

- Ainley, W. M., I. G. Macreadie, and R. A. Butowa. 1985. *var1* gene on the mitochondrial genome of *Torulopsis glabrata*. *J. Mol. Biol.* **184**:565-576.
- Akiyama, M., H. Maki, M. Sekiguchi, and T. Horiuchi. 1989. A specific role of MutT protein: to prevent dG-dA mispairing in DNA replication. *Proc. Natl. Acad. Sci. USA* **86**:3949-3952.
- Andachi, Y., F. Yamao, M. Iwami, A. Muto, and S. Osawa. 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.
- Andachi, Y., F. Yamao, A. Muto, and S. Osawa. 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., A. T. Bankier, B. G. Barrell, M. H. L. de Bruijn, A. R. Coulson, J. Drouin, I. C. Eperon, D. P. Nierlich, B. A. Roe, F. Sanger, P. H. Schreier, A. J. H. Smith, R. Staden, and I. G. Young. 1981. Sequence and organization of the human mitochondrial genome. *Nature (London)* **290**:457-465.
- Anderson, S., M. H. L. de Bruijn, A. R. Coulson, I. C. Eperon, F. Sanger, and I. G. Young. 1982. Complete sequence of bovine mitochondrial DNA: conserved features of the mammalian mitochondrial genome. *J. Mol. Biol.* **156**:683-717.
- Anfinsen, C. B., and L. G. Corley. 1969. An active variant of staphylococcal nuclease containing norleucine in place of methionine. *J. Biol. Chem.* **244**:5149-5152.
- Aota, S.-I., and T. Ikemura. 1986. Diversity in G + C content at the third position of codons in vertebrate genes and its cause. *Nucleic Acids Res.* **14**:6345-6355.
- Araki, T., S. Asakawa, Y. Kumazawa, K.-I. Miura, and K. Watanabe. 1988. Gene organization of tRNAs and their secondary structures found in starfish mitochondrial genome. *Nucleic Acids Res. Symp. Ser.* **20**:93-94.
- Asakawa, S., Y. Kumazawa, H. Araki, T. Himeno, K. Miura, and K. Watanabe. 1991. Strand-specific nucleotide composition bias in echinoderm and vertebrate mitochondrial genomes. *J. Mol. Evol.* **32**:511-520.
- Barahona, I., H. Soares, L. Cyrne, D. Penque, P. Denoulet, and C. Rodrigues-Pousada. 1988. Sequence of one alpha- and two beta-tubulin genes of *Tetrahymena pyriformis*. *J. Mol. Biol.* **202**:365-382.
- Baron, C., J. Heider, and A. Böck. 1990. Mutagenesis of selC, the gene for the selenocysteine-inserting tRNA-species in *E. coli*: effects in *in vivo* function. *Nucleic Acids Res.* **18**:6761-6766.

13. Barrell, B. G., S. Anderson, A. T. Bankier, M. H. L. de Bruijn, E. Chen, A. R. Coulson, J. Drouin, I. C. Eperon, D. P. Nierlich, B. A. Roe, F. Sanger, P. H. Schreier, A. J. H. Smith, R. Staden, and I. G. Yang. 1980. Different pattern of codon recognition by mammalian mitochondrial tRNAs. *Proc. Natl. Acad. Sci. USA* 77:3164–3166.
14. Barrell, B. G., A. T. Bankier, and J. Drouin. 1979. A different genetic code in human mitochondria. *Nature (London)* 282:189–194.
15. Bernardi, G. 1989. The isochore organization of the human genome. *Annu. Rev. Genet.* 23:637–661.
16. Bernardi, G., and G. Bernardi. 1986. Compositional constraints and genome evolution. *J. Mol. Evol.* 24:1–11.
17. Bernardi, G., B. Olofsson, J. Filipski, M. Zerial, J. Salinas, G. Cuny, M. Meunier-Rotival, and F. Rodier. 1985. The mosaic genome of warm-blooded vertebrates. *Science* 228:953–958.
18. Berry, M. J., L. Banu, Y. Chen, S. J. Mandel, J. D. Kieffer, J. W. Harney, and P. R. Larsen. 1991. Recognition of UGA as a selenocysteine codon in Type 1 deiodinase requires sequences in the 3' untranslated region. *Nature (London)* 353:273–276.
19. Berry, M. J., L. Banu, and P. R. Larsen. 1991. Type 1 iodothyronine deiodinase is a selenocysteine-containing enzyme. *Nature (London)* 349:438–440.
20. Bessho, Y., T. Ohama, and S. Osawa. Planarian mitochondria. II. The unique genetic code as deduced from COI gene sequences. *J. Mol. Evol.*, in press.
21. Bibb, M. J., R. A. VanEtten, C. T. Wright, M. W. Walberg, and D. A. Clayton. 1981. Sequence and gene organization of mouse mitochondrial DNA. *Cell* 26:167–180.
22. Björk, G. J. 1987. Modification of stable RNA, p. 719–731. In F. C. Neidhardt, J. L. Ingraham, K. B. Low, B. Magasanik, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli and Salmonella typhimurium: cellular and molecular biology*, vol. 1. American Society for Microbiology, Washington, D.C.
23. Böck, A., and T. C. Stadtman. 1988. Selenocysteine, a highly specific component of certain enzymes, is incorporated by a UGA-directed co-translational mechanism. *BioFactors* 1:245–250.
- 23a. Böck, A., K. Forchhammer, J. Heider, W. Leinfelder, G. Sawers, B. Veprek, and F. Zinoni. 1991. Selenocysteine: the 21st amino acid. *Mol. Microbiol.* 5:515–520.
24. Borsook, H., C. L. Deasy, A. J. Haagen-Smit, G. Keighley, and P. H. Lowy. 1949. The uptake *in vitro* of C¹⁴-labeled glycine, L-leucine, and L-lysine by different components of guinea pig liver homogenate. *J. Biol. Chem.* 184:529–543.
25. Brown, C. M., P. A. Stockwell, C. N. A. Trotman, and W. P. Tate. 1990. The signal for the termination of protein synthesis in prokaryotes. *Nucleic Acids Res.* 18:2079–2086.
26. Brown, C. M., P. A. Stockwell, C. N. A. Trotman, and W. P. Tate. 1990. Sequence analysis suggests that tetra-nucleotides signal termination of protein synthesis in eukaryotes. *Nucleic Acids Res.* 18:6339–6345.
27. Browning, K. S., and U. L. RajBhandary. 1982. Cytochrome oxidase subunit III gene in *Neurospora crassa* mitochondria: location and sequence. *J. Biol. Chem.* 257:5253–5256.
28. Burger, G., C. Scriven, W. Machleidt, and S. Werner. 1982. Subunit 1 of cytochrome oxidase from *Neurospora crassa*: nucleotide sequence of the coding gene and partial amino acid sequence of the protein. *EMBO J.* 1:1385–1391.
29. Canaday, J., G. Dirheimer, and R. P. Martin. 1980. Yeast mitochondrial methionine initiator tRNA: characterization and nucleotide sequence. *Nucleic Acids Res.* 8:1445–1457.
30. Cantatore, P., M. Roberti, G. Rainaldi, M. N. Gadaleta, and C. Saccone. 1989. The complete nucleotide sequence, gene organization, and genetic code of the mitochondrial genome of *Paracentrotus lividus*. *J. Biol. Chem.* 264:10965–10975.
31. Caron, F., and E. Meyer. 1985. Does *Paramecium primaurelia* use a different genetic code in its macronucleus? *Nature (London)* 314:185–188.
32. Caskey, C. T. 1980. Peptide chain termination. *Trends Biochem. Sci.* 5:234–237.
33. Cerretti, D. P., D. Dean, G. R. Davis, D. M. Bedwell, and M. Nomura. 1983. The spc ribosomal protein operon of *Escherichia coli*: sequence and cotranscription of the ribosomal protein genes and a protein export gene. *Nucleic Acids Res.* 11:2599–2616.
34. Chambers, I., J. Frampton, P. Goldfarb, N. Affara, W. McBain, and P. R. Harrison. 1986. The structure of the mouse glutathione peroxidase gene: the selenocysteine in the active site is encoded by the “termination” codon, TGA. *EMBO J.* 5:1221–1227.
35. Chevalier, C., C. Saillard, and J. M. Bové. 1990. Organization and nucleotide sequences of the *Spiroplasma citri* genes for ribosomal protein S2, elongation factor Ts, spiralin phosphofructokinase, pyruvate kinase and an unidentified protein. *J. Bacteriol.* 172:2693–2703.
36. Clark-Walker, G. D., C. R. McArthur, and K. Spriprakash. 1985. Location of transcriptional control signals and transfer RNA sequence in *Torulopsis glabrata* mitochondrial DNA. *EMBO J.* 4:465–473.
37. Clary, D. O., and D. R. Wolstenholme. 1985. The mitochondrial DNA molecule of *Drosophila yakuba*: nucleotide sequence, gene organization, and genetic code. *J. Mol. Evol.* 22:252–271.
38. Cone, J. E., R. M. del Rio, J. N. Davis, and T. C. Stadtman. 1976. Chemical characterization of the selenoprotein component of clostridial glycine reductase: identification of selenocysteine as the organoselenium moiety. *Proc. Natl. Acad. Sci. USA* 73:2659–2663.
39. Cowie, D. B., G. N. Cohen, E. T. Bolton, H. De Robichon-Szulmajster. 1959. Amino acid analog incorporation into bacterial proteins. *Biochim. Biophys. Acta* 34:39–46.
40. Cox, E. C., and C. Yanofsky. 1967. Altered base ratios in the DNA of an *Escherichia coli* mutator strain. *Proc. Natl. Acad. Sci. USA* 58:1895–1902.
41. Crick, F. H. C. 1966. Codon-anticodon pairing: the wobble hypothesis. *J. Mol. Biol.* 19:548–555.
42. Crick, F. H. C. 1968. The origin of the genetic code. *J. Mol. Biol.* 38:367–379.
43. Crick, F. H. C. 1988. What mad pursuit: a personal view of scientific discovery. Basic Books, Inc., New York.
44. Crick, F. H. C., S. Brenner, A. Klug, and C. Piezenik. 1976. A speculation on the origin of protein synthesis. *Origins Life* 7:389–397.
45. Crick, F. H. C., J. S. Griffith, and L. E. Orgel. 1957. Codes without commas. *Proc. Natl. Acad. Sci. USA* 43:416–421.
46. Cronin, J. R., and S. Pizzarello. 1983. Amino acids in meteorites. *Adv. Space Res.* 3:5–18.
47. de Bruijn, M. H. L. 1983. *Drosophila melanogaster* mitochondrial DNA, a novel organization and genetic code. *Nature (London)* 304:234–241.
48. Desjardins, P., and R. Morais. 1990. Sequence and gene organization of the chicken mitochondrial genome. A novel gene order in higher vertebrates. *J. Mol. Biol.* 212:599–634.
49. Diamond, A., B. Dudock, and D. Hatfield. 1981. Structure and properties of a bovine liver UGA suppressor serine tRNA with a tryptophan anticodon. *Cell* 25:497–506.
50. Dirheimer, G., and R. P. Martin. 1990. Mitochondrial tRNAs: structure, modified nucleosides and codon reading patterns. *J. Chromatogr.* 45B:197–264.
51. Dounce, A. L. 1952. Duplicating mechanism for peptide chain and nucleic acid synthesis. *Enzymologia* 15:251–258.
52. Dubin, D. T., and C.-C. HsuChen. 1984. Sequence and structure of a methionine transfer RNA from mosquito mitochondria. *Nucleic Acids Res.* 12:4185–4189.
53. Dudler, R., C. Schmidhauser, R. W. Parish, R. E. Wettenhall, and J. Schmidt. 1988. A mycoplasma high-affinity transport system and the *in vitro* invasiveness of mouse sarcoma cells. *EMBO J.* 7:3963–3970.
54. Eigen, M., B. F. Lindemann, M. Tietze, R. Winkler-Oswatitsch, A. Dress, and A. von Haeseler. 1989. How old is the genetic code? Statistical geometry of tRNA provides an answer. *Science* 244:673–679.
55. Elliott, M. S., and R. W. Trewyn. 1984. Inosine biosynthesis in transfer RNA by an enzymatic insertion of hypoxanthine. *J. Biol. Chem.* 259:2407–2418.

56. Fiers, W., R. Contreras, F. Duerinck, G. Haegeman, D. Iserentant, J. Merregaert, W. Min Jou, F. Molemans, A. Raeymaekers, A. Van den Berghe, G. Volckaert, and M. Ysebaert. 1976. Complete nucleotide sequence of bacteriophage MS2 RNA: primary and secondary structure of the replicase gene. *Nature (London)* **260**:500-507.
57. Filipinski, J. 1991. Evolution of DNA sequences. Contribution of mutation bias and selection to the origin of chromosomal compartments. *Adv. Mutagenesis Res.* **2**:1-54.
58. Flohe, L., W. A. Günzler, and H. H. Shock. 1973. Glutathione peroxidase: a selenoenzyme. *FEBS Lett.* **32**:132-134.
- 58a. Forchhammer, K., and A. Böck. 1991. Selenocysteine synthase from *Escherichia coli*. *J. Biol. Chem.* **266**:6324-6328.
59. Forstrom, J. W., J. J. Zakowski, and A. L. Tappel. 1978. Identification of the catalytic site of rat liver glutathione peroxidase as selenocysteine. *Biochemistry* **17**:2639-2644.
60. Fox, T. D. 1987. Natural variation in the genetic code. *Annu. Rev. Genet.* **21**:67-91.
61. Gadaleta, G., G. Pepe, G. De Candia, C. Quagliariello, E. Sbisà, and C. Saccone. 1989. The complete nucleotide sequence of the *Rattus norvegicus* mitochondrial genome: cryptic signals revealed by comparative analysis between vertebrates. *J. Mol. Evol.* **28**:497-516.
62. Gamow, G. 1954. Possible relation between deoxyribonucleic acid and protein structures. *Nature (London)* **173**:318.
63. Garesse, R. 1988. *Drosophila melanogaster* mitochondrial DNA: gene organization and evolutionary considerations. *Genetics* **118**:649-683.
64. Garey, J. R., and D. R. Wolstenholme. 1989. Platyhelminth mitochondrial DNA: evidence for early origin of a tRNA^{Ser} AGN that contains a dihydrouridine arm replacement loop, and serine-specifying AGA and AGG codons. *J. Mol. Evol.* **28**:374-387.
65. Grisi, E., T. A. Brown, R. B. Waring, C. Scazzocchio, and R. W. Davies. 1982. Nucleotide sequence of a region of the mitochondrial genome of *Aspergillus nidulans*, including the gene for ATPase subunit 6. *Nucleic Acids Res.* **10**:3531-3539.
66. Grivell, L. A. 1986. Deciphering divergent codes. *Nature (London)* **324**:109-110.
67. Günzler, W. A., G. L. Steffens, A. Grossman, S.-M. A. Kim, F. Otting, A. Wendel, and L. Flöhe. 1984. The amino acid sequence of bovine glutathione peroxidase. *Hoppe-Seyler's Z. Physiol. Chem.* **365**:195-212.
68. Gupta, R. 1984. *Halobacterium volcanii* tRNAs: identification of 41 tRNAs covering all amino acids, and the sequences of 33 Class 1 tRNAs. *J. Biol. Chem.* **259**:9461-9471.
69. Guthrie, C., and J. Abelson. 1982. Organization and expression of tRNA genes in *Saccharomyces cerevisiae*, p. 487-528. *In* J. N. Strathern, E. Jones, and J. Broach (ed.), *The molecular biology of the yeast Saccharomyces: metabolism and gene expression*. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
70. Hanyu, N., Y. Kuchino, S. Nishimura, and H. Beier. 1986. Dramatic events in ciliate evolution: alteration of UAA and UAG termination codons to glutamine codons due to anticodon mutations in two *Tetrahymena* tRNAs Gln. *EMBO J.* **5**:1307-1311.
71. Hara-Yokoyama, M., S. Yokoyama, T. Watanabe, K. Watanabe, M. Kitazumi, Y. Mitamura, T. Morii, S. Takahashi, Y. Kuchino, S. Nishimura, and T. Miyazawa. 1986. Characteristic anticodon sequences of major tRNA species from an extreme thermophile, *Thermus thermophilus* HB8. *FEBS Lett.* **202**:149-152.
72. Harper, D. S., and C. L. Jahn. 1989. Differential use of termination codons in ciliated protozoa. *Proc. Natl. Acad. Sci. USA* **86**:3252-3256.
73. Hartmanis, M. G. N., and T. C. Stadtman. 1982. Isolation of a selenium-containing thiolase from *Clostridium kluyveri*: identification of the selenium moiety as selenomethionine. *Proc. Natl. Acad. Sci. USA* **79**:4912-4916.
74. Hatfield, D., A. Diamond, and B. Dudock. 1982. Opal suppressor serine tRNAs from bovine liver form phosphoseryl-tRNA. *Proc. Natl. Acad. Sci. USA* **79**:6215-6219.
75. Heckman, J. E., J. Sarnoff, B. Alzner-DeWeerd, S. Yin, and U. L. Rajbhandary. 1980. Novel features in the genetic code and codon reading patterns in *Neurospora crassa* mitochondria based on sequences of six mitochondrial tRNAs. *Proc. Natl. Acad. Sci. USA* **77**:3159-3163.
76. Heider, J., W. Leinfelder, and A. Böck. 1989. Occurrence and functional compatibility within Enterobacteriaceae of a tRNA species which inserts selenocysteine into protein. *Nucleic Acids Res.* **17**:2529-2540.
77. Helftenbein, E. 1985. Nucleotide sequence of a macronuclear DNA molecule coding for alpha-tubulin from the ciliate *Stylonychia lemnae*. Special codon usage: TAA is not a translation termination codon. *Nucleic Acids Res.* **13**:415-433.
78. Helftenbein, E., and E. Müller. 1988. Both alpha-tubulin genes are transcriptionally active in *Stylonychia lemnae*. *Curr. Genet.* **13**:425-432.
79. Hendriks, L., C. Van Broeckhoven, A. Vandenberghe, Y. Van De Peer, and R. DeWachter. 1988. Primary and secondary structure of the 18S ribosomal RNA of the bird spider *Eurytelma californica* and evolutionary relationship among eukaryotic phyla. *Eur. J. Biochem.* **177**:15-20.
80. Hensgens, L. A. M., J. Brakenhoff, B. F. De Vries, P. Sloof, M. C. Tromp, J. H. Van Boon, and R. Benne. 1984. The sequence of the gene for cytochrome c oxidase subunit I, a frameshift containing gene for cytochrome c oxidase subunit II and seven unassigned reading frames in *Trypanosoma brucei* mitochondrial maxi-circle DNA. *Nucleic Acids Res.* **12**:7327-7344.
81. Hensgens, L. A. M., L. A. Grivell, P. Borst, and J. L. Bos. 1979. Nucleotide sequence of the mitochondrial structural gene for subunit 9 of yeast ATPase complex. *Proc. Natl. Acad. Sci. USA* **76**:1663-1667.
82. Herrick, G., D. Hunter, K. Williams, and K. Kotter. 1987. Alternate processing during development of a macronuclear chromosome family in *Oxytricha fallax*. *Genes Dev.* **1**:1047-1058.
83. Hill, K. E., R. S. Lloyd, J.-G. Yang, R. Read, and R. F. Burk. 1991. The cDNA for rat selenoprotein P contains 10 TGA codons in the open reading frame. *J. Biol. Chem.* **266**:10050-10053.
84. Himeno, H., H. Masaki, T. Ohta, I. Kumagai, K.-I. Miura, and K. Watanabe. 1987. Unusual genetic codes and a novel genome structure for tRNA^{Ser}_{AGY} in starfish mitochondrial DNA. *Gene* **56**:219-230.
- 84a. Hiratsuka, J., H. Shimada, R. Whittier, T. Ishibashi, M. Sakamoto, M. Mori, C. Kondo, Y. Honji, C. Sun, B. Meng, Y. Li, A. Kanno, Y. Nishizawa, A. Hirai, K. Shinozaki, and M. Sugiura. 1989. The complete sequence of the rice (*Oryza sativa*) chloroplast genome: intermolecular recombination between distinct tRNA genes accounts for a major plastic DNA inversion during the evolution of the cereals. *Mol. Gen. Genet.* **217**:185-194.
85. Hoekstra, W. G., D. Hafeman, S. H. Oh, R. A. Sunde, and H. E. Ganther. 1973. Effect of dietary selenium on liver and erythrocyte glutathione peroxidase in the rat. *Fed. Proc.* **132**:885.
86. Holmquist, G. P. 1989. Evolution of chromosome bands: molecular ecology of noncoding DNA. *J. Mol. Evol.* **28**:469-486.
87. Holmquist, R., T. H. Jukes, and S. Pangburn. 1973. Evolution of transfer RNA. *J. Mol. Biol.* **78**:91-116.
88. Hori, H., and S. Osawa. 1979. Evolutionary change in 5S RNA secondary structure and a phylogenetic tree of 54 5S RNA species. *Proc. Natl. Acad. Sci. USA* **76**:381-385.
89. Hori, H., and S. Osawa. 1987. Origin and evolution of organisms as deduced from 5S ribosomal RNA sequences. *Mol. Biol. Evol.* **4**:445-472.
90. Horowitz, S., and M. A. Gorovsky. 1985. An unusual genetic code in nuclear genes of *Tetrahymena*. *Proc. Natl. Acad. Sci. USA* **82**:2452-2455.
91. HsuChen, C.-C., G. R. Cleaves, and D. T. Dubin. 1983. A major lysine tRNA with a CUU anticodon in insect mitochondria. *Nucleic Acids Res.* **11**:8659-8662.

92. HsuChen, C.-C., and D. T. Dubin. 1984. A cluster of four transfer RNA genes in mosquito mitochondrial DNA. *Biochem. Int.* **8**:385-391.
93. HsuChen, C.-C., R. M. Kotin, and D. T. Dubin. 1984. Sequences of the coding and flanking regions of the large ribosomal subunit RNA gene of mosquito mitochondria. *Nucleic Acids Res.* **12**:7771-7785.
94. Hudspeth, M. E. S., W. M. Ainley, D. S. Shumard, R. A. Butow, and L. I. Grossman. 1982. Location and structure of the *var1* gene on yeast mitochondrial DNA: nucleotide sequence of the 40.0 allele. *Cell* **30**:617-626.
95. Ikeda, R., T. Ohama, A. Muto, and S. Osawa. 1990. Nucleotide sequences of nine tRNA genes from *Micrococcus luteus*. *Nucleic Acids Res.* **18**:7154.
96. Ikeda, R., T. Ohama, A. Muto, and S. Osawa. 1990. Nucleotide sequences of two tRNA gene clusters from *Micrococcus luteus*. *Nucleic Acids Res.* **18**:7155.
97. Ikemura, T. 1981. Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes. *J. Mol. Biol.* **146**:1-21.
98. Ikemura, T. 1981. Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes: a proposal for a synonymous codon choice that is optimal for the *E. coli* translational system. *J. Mol. Biol.* **151**:389-409.
99. Ikemura, T. 1982. Correlation between the abundance of yeast transfer RNAs and the occurrence of the respective codons in protein genes. *J. Mol. Biol.* **158**:573-597.
100. Ikemura, T., and S.-I. Aota. 1988. Global variation in G + C content along vertebrate genome DNA. Possible correlation with chromosome band structure. *J. Mol. Biol.* **203**:1-13.
101. Inamine, J. M., T. P. Denny, S. Loechel, U. Schaper, C.-H. Huang, K. F. Bott, and P.-C. Hu. 1988. Nucleotide sequence of the P1 attachment-protein gene of *Mycoplasma pneumoniae*. *Gene* **64**:217-229.
102. Inamine, J. M., K. Ho, S. Loechel, and P. Hu. 1990. Evidence that UGA is read as tryptophan rather than stop by *Mycoplasma pneumoniae*, *Mycoplasma genitalium*, and *Mycoplasma gallisepticum*. *J. Bacteriol.* **172**:504-506.
103. Iwabe, N., K. Kuma, M. Hasegawa, S. Osawa, and T. Miyata. 1989. Evolutionary relationship of archaeobacteria, eubacteria, and eukaryotes inferred from phylogenetic trees of duplicated genes. *Proc. Natl. Acad. Sci. USA* **86**:9355-9359.
104. Jacobs, H. T., S. Asakawa, T. Araki, K. Miura, M. Smith, and K. Watanabe. 1989. Conserved tRNA gene cluster in starfish mitochondrial DNA. *Curr. Genet.* **15**:193-206.
105. Jacobs, H. T., D. J. Elliott, V. B. Math, and A. Farquharson. 1988. Nucleotide sequence and gene organization of sea-urchin mitochondrial DNA. *J. Mol. Biol.* **202**:185-217.
106. Jukes, T. H. 1965. The genetic code, II. *Am. Sci.* **53**:477-487.
107. Jukes, T. H. 1966. *Molecules and evolution*. Columbia University Press, New York.
108. Jukes, T. H. 1973. Possibilities for the evolution of the genetic code from a preceding form. *Nature (London)* **246**:22-26.
109. Jukes, T. H. 1981. Amino acid codes in mitochondria as possible clues to primitive codes. *J. Mol. Evol.* **18**:15-17.
110. Jukes, T. H. 1983. Evolution of the amino acid code: inferences from mitochondrial codes. *J. Mol. Evol.* **19**:219-225.
111. Jukes, T. H. 1985. A change in the genetic code in *Mycoplasma capricolum*. *J. Mol. Evol.* **22**:361-362.
112. Jukes, T. H. 1990. Genetic code 1990. Outlook. *Experientia* **46**:1149-1157.
113. Jukes, T. H., Y. Bessho, T. Ohama, and S. Osawa. 1991. Release factors and genetic code. *Nature (London)* **352**:575.
114. Jukes, T. H., and V. Bhushan. 1986. Silent nucleotide substitutions and G + C content of some mitochondrial and bacterial genes. *J. Mol. Evol.* **24**:39-44.
115. Jukes, T. H., R. Holmquist, and H. Moise. 1975. Amino acid composition of proteins: selection against the genetic code. *Science* **189**:50-51.
116. Jukes, T. H., and S. Osawa. 1990. The genetic code in mitochondria and chloroplasts. *Experientia* **46**:1117-1126.
117. Jukes, T. H., and S. Osawa. 1991. Recent evidence for evolution of the genetic code, p. 79-95. *In* S. Osawa and T. Honjo (ed.), *Evolution of life: fossils, molecules and culture*. Springer-Verlag, Tokyo.
118. Jukes, T. H., S. Osawa, and A. Muto. 1987. Divergence and directional mutation pressures. *Nature (London)* **325**:668.
119. Jukes, T. H., S. Osawa, A. Muto, and N. Lehman. 1987. Evolution of anticodons: variations in the genetic code. *Cold Spring Harbor Symp. Quant. Biol.* **52**:769-776.
120. Jungck, J. R. 1978. The genetic code as a periodic table. *J. Mol. Evol.* **11**:211-224.
121. Kagawa, Y., H. Nojima, N. Nukiwa, M. Ishizuka, T. Nakajima, T. Yasuhara, T. Tanaka, and T. Oshima. 1984. High guanine plus cytosine content in the third letter of codons of extreme thermophiles. DNA sequence of the isopropylmalate dehydrogenase of *Thermus thermophilus*. *J. Biol. Chem.* **259**:2956-2960.
122. Kano, A., Y. Andachi, T. Ohama, and S. Osawa. 1991. Novel anticodon composition of transfer RNAs in *Micrococcus luteus*, a bacterium with a high genomic GC-content: correlation with codon usage. *J. Mol. Biol.* **221**:387-401.
123. Kawaguchi, Y., H. Honda, J. Taniguchi-Morimura, and S. Iwasaki. 1989. The codon CUG is read as serine in an asporogenic yeast *Candida cylindracea*. *Nature (London)* **341**:164-166.
124. Kimura, M. 1983. *The neutral theory of molecular evolution*. Cambridge University Press, Cambridge.
125. King, J. L., and T. H. Jukes. 1969. Non-Darwinian evolution. *Science* **164**:788-798.
126. Köchel, H. G., C. M. Lazarus, N. Basak, and H. Küntzel. 1981. Mitochondrial tRNA gene clusters in *Aspergillus nidulans*: organization and nucleotide sequence. *Cell* **23**:625-633.
- 126a. Kojima, A., Y. Andachi, and S. Osawa. Unpublished data.
127. Komine, Y., T. Adachi, H. Inokuchi, and H. Ozeki. 1990. Genomic organization and physical mapping of the transfer RNA genes in *Escherichia coli* K12. *J. Mol. Biol.* **212**:579-598.
128. Kondo, K., H. Sone, H. Yoshida, T. Toida, K. Kanatani, Y.-M. Hong, N. Nishino, and J. Tanaka. 1990. Cloning and sequence analysis of the arginine deaminase gene from *Mycoplasma arginini*. *Mol. Gen. Genet.* **221**:81-86.
129. Kuchino, Y., N. Hanyu, F. Tashiro, and S. Nishimura. 1985. *Tetrahymena thermophila* glutamine tRNA and its gene that corresponds to UAA termination codon. *Proc. Natl. Acad. Sci. USA* **82**:4758-4762.
130. Lagerkvist, U. 1978. "Two out of three": an alternative method for codon reading. *Proc. Natl. Acad. Sci. USA* **75**:1759-1762.
131. Lagerkvist, U. 1981. Unorthodox codon reading and the evolution of the genetic code. *Cell* **23**:305-306.
132. Lee, B. J., M. Rajagopalan, Y. S. Kim, H. Kwang, K. You, K. B. Jacobson, and D. Hatfield. 1990. Selenocysteine tRNA^{[Ser]Sec} gene is ubiquitous within the animal kingdom. *Mol. Cell. Biol.* **10**:1940-1949.
133. Lee, B. J., P. J. Worland, J. N. Davis, T. C. Stadtman, and D. Hatfield. 1989. Identification of a selenocysteine tRNA^[Ser] in mammalian cells that recognizes the nonsense codon, UGA. *J. Biol. Chem.* **264**:9724-9727.
134. Lee, C. C., K. M. Timms, C. N. A. Trotman, and W. P. Tate. 1987. Isolation of a rat mitochondrial release factor: accommodation of the changed genetic code for termination. *J. Biol. Chem.* **262**:3548-3552.
135. Lehman, N., and T. H. Jukes. 1988. Genetic code development by stop codon takeover. *J. Theor. Biol.* **135**:204-214.
136. Leinfelder, W., K. Forchhammer, B. Veprek, E. Zehelein, and A. Böck. 1990. *In vitro* synthesis of selenocysteinyl-tRNA_{UCA} from seryl-tRNA_{UCA}: involvement and characterization of selD gene product. *Proc. Natl. Acad. Sci. USA* **87**:543-547.
137. Leinfelder, W., E. Zehelein, M. A. Mandrand-Berthelot, and A. Böck. 1988. Gene for a novel tRNA species that accepts L-serine and co-translationally inserts selenocysteine. *Nature (London)* **331**:723-725.
138. Lewin, B. 1990. *Genes IV*, p. 148. Oxford University Press, Oxford.

139. Li, M., and A. Tzagoloff. 1979. Assembly of the mitochondrial membrane system: sequences of yeast mitochondrial valine and an unusual threonine tRNA gene. *Cell* **18**:47–53.
140. Loftfield, R. B., and D. A. Vanderjagt. 1972. The frequency of errors in protein biosynthesis. *Biochem. J.* **128**:1353–1356.
141. Macino, G., G. Coruzzi, F. G. Nobrega, M. Li, and A. Tzagoloff. 1979. Use of the UGA terminator as a tryptophan codon in yeast mitochondria. *Proc. Natl. Acad. Sci. USA* **76**:3784–3785.
142. Macino, G., and A. Tzagoloff. 1979. Assembly of the mitochondrial membrane system—partial sequence of a mitochondrial ATPase gene in *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. USA* **76**:131–135.
143. Maréchal, L., P. Guillemaut, J.-M. Grienenberger, G. Jeannin, and J.-H. Weil. 1985. Sequence and codon recognition of bean mitochondria and chloroplast tRNAs Trp: evidence for a high degree of homology. *Nucleic Acids Res.* **13**:4411–4416.
144. Maréchal-Drouard, L., P. Guillemaut, A. Cosset, M. Arbogast, F. Weber, J. Weil, and A. Dietrich. 1990. Transfer RNAs of potato (*Solanum tuberosum*) mitochondria have different genetic origins. *Nucleic Acids Res.* **18**:3689–3696.
145. Martin, R. O., A.-P. Sibling, C. W. Gehrke, K. Kuo, C. G. Edmonds, J. A. McCloskey, and G. Dirheimer. 1990. 5-[[[(carboxymethyl)amino]methyl]uridine is found in the anticodon of yeast mitochondrial tRNAs recognizing two-codon families ending in a purine. *Biochemistry* **29**:956–959.
146. Martin, R. P., J.-M. Schneller, A. J. C. Stahl, and G. Dirheimer. 1979. Import of nuclear DNA coded lysine-accepting tRNA (anticodon C-U-U) into yeast mitochondria. *Biochemistry* **18**:4600–4605.
147. Menon, N. K., H. D. Peck, Jr., J. Le Gall, and A. E. Przybyla. 1987. Cloning and sequencing of the genes encoding the large and small subunits of the periplasmic (NiFeSe) hydrogenase of *Desulfovibrio baculatus*. *J. Bacteriol.* **169**:5401–5407.
148. Meyer, F. M., H. J. Schmidt, E. Plümper, A. Hasilik, G. Mersmann, H. E. Meyer, A. Engström, and K. Heckmann. 1991. UGA is translated as cysteine in pheromone 3 of *Euplotes octocarinatus*. *Proc. Natl. Acad. Sci. USA* **88**:3758–3761.
149. Miller, S. L. 1953. A production of amino acids under possible primitive earth conditions. *Science* **117**:528–529.
150. Miyata, T., N. Iwabe, K. Kuma, Y. Kawanishi, M. Hasegawa, H. Kishino, Y. Mukohata, K. Ihara, and S. Osawa. 1991. Evolution of archaeobacteria: phylogenetic relationships among archaeobacteria, eubacteria, and eukaryotes, p. 337–351. *In* S. Osawa and T. Honjo (ed.), *Evolution of life: fossils, molecules and culture*. Springer-Verlag, Tokyo.
151. Mizutani, T., H. Kurata, and K. Yamada. 1991. Study of mammalian selenocysteyl-tRNA synthesis with (75 Se) HSe⁻. *FEBS Lett.* **289**:59–63.
152. Motchnik, P. A., and A. L. Tappel. 1989. Rat plasma selenoprotein P properties and purification. *Biochim. Biophys. Acta* **993**:27–35.
153. Motchnik, P. A., and A. L. Tappel. 1990. Multiple selenocysteine content of selenoprotein P in rats. *J. Inorg. Biochem.* **40**:265–269.
154. Muramatsu, T., K. Nishikawa, F. Nemoto, Y. Kuchino, A. Nishimura, T. Miyazawa, and S. Yokoyama. 1988. Codon and amino acid specificities of a transfer RNA are both converted by a single post-transcriptional modification. *Nature (London)* **336**:179–181.
155. Muramatsu, T., S. Yokoyama, N. Horie, A. Matsuda, T. Ueda, Z. Yamaizumi, Y. Kuchino, S. Nishimura, and T. Miyazawa. 1988. A novel lysine-substituted nucleoside in the first position of the anticodon of minor isoleucine tRNA from *Escherichia coli*. *J. Biol. Chem.* **263**:9261–9267.
156. Murao, K., and H. Ishikura. 1978. A new uridine derivative located in the anticodon of tRNA^{Gly} from *Bacillus subtilis*. *Nucleic Acids Res.* **5**:S333–S336.
157. Muto, A., Y. Andachi, H. Yuzawa, F. Yamao, and S. Osawa. 1990. The organization and evolution of transfer RNA genes of *Mycoplasma capricolum*. *Nucleic Acids Res.* **18**:5037–5043.
158. Muto, A., Y. Kawauchi, F. Yamao, and S. Osawa. 1984. Preferential use of A- and U-rich codons for *Mycoplasma capricolum* ribosomal proteins S8 and L6. *Nucleic Acids Res.* **12**:8209–8217.
159. Muto, A., T. Ohama, Y. Andachi, F. Yamao, R. Tanaka, and S. Osawa. 1991. Evolution of codons and anticodons in eubacteria, p. 179–193. *In* M. Kimura and N. Takahata (ed.), *New aspects in population genetics and molecular evolution*. Japan Scientific Societies Press/Springer-Verlag, Tokyo.
160. Muto, A., and S. Osawa. 1987. The guanine and cytosine content of genomic DNA and bacterial evolution. *Proc. Natl. Acad. Sci. USA* **84**:166–169.
161. Nellen, W., and D. Gallwitz. 1982. Actin genes and actin messenger RNA in *Acanthamoeba castellanii*: nucleotide sequence of the split actin gene I. *J. Mol. Biol.* **159**:1–18.
162. Netzker, R., H. G. Köchel, N. Basak, and H. Küntzel. 1982. Nucleotide sequence of *Aspergillus nidulans* mitochondrial genes coding for ATPase subunit 6, cytochrome oxidase subunit 3, seven unidentified proteins, four tRNAs and L-rRNA. *Nucleic Acids Res.* **10**:4783–4794.
163. Nghiem, Y., M. Cabrera, C. G. Cupples, and J. H. Miller. 1988. The mutY gene: a mutator locus in *Escherichia coli* that generates G · C → T · A transversions. *Proc. Natl. Acad. Sci. USA* **85**:2709–2713.
164. Nirenberg, M. W., and J. H. Matthaei. 1961. The dependence of cell-free protein synthesis in *E. coli* upon naturally occurring or synthetic polyribonucleotides. *Proc. Natl. Acad. Sci. USA* **47**:1588–1602.
165. Nishimura, S. 1979. Modified nucleosides in tRNA, p. 59–79. *In* P. R. Schimmel, D. Söll, and J. M. Abelson (ed.), *Transfer RNA: structure, properties, and recognition*. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
166. Nishimura, S. 1983. Structure, biosynthesis and function of queuosine in tRNA. *Prog. Nucleic Acid Res. Mol. Biol.* **28**:49–73.
167. Nomoto, M., N. Imai, H. Saiga, T. Matsui, and T. Mita. 1987. Characterization of two types of histone H2B genes from macronuclei of *Tetrahymena thermophila*. *Nucleic Acids Res.* **15**:5681–5698.
168. Oba, T., Y. Andachi, A. Muto, and S. Osawa. 1991. CGG, an unassigned or nonsense codon in *Mycoplasma capricolum*. *Proc. Natl. Acad. Sci. USA* **88**:921–925.
169. Oba, T., Y. Andachi, A. Muto, and S. Osawa. 1991. Translation in vitro of codon UGA as tryptophan in *Mycoplasma capricolum*. *Biochimie* **73**:1109–1112.
170. Oh, S. H., H. E. Ganther, and W. G. Hoekstra. 1974. Selenium as a component of glutathione peroxidase isolated from bovine erythrocytes. *Biochemistry* **13**:1825–1828.
171. Ohama, T., A. Muto, and S. Osawa. 1989. Spectinomycin operon of *Micrococcus luteus*: evolutionary implications of organization and codon usage. *J. Mol. Evol.* **29**:381–395.
172. Ohama, T., A. Muto, and S. Osawa. 1990. Role of GC-biased mutation pressure on synonymous codon choice in *Micrococcus luteus*, a bacterium with a high genomic GC-content. *Nucleic Acids Res.* **18**:1565–1569.
173. Ohama, T., S. Osawa, K. Watanabe, and T. H. Jukes. 1990. Evolution of the mitochondrial genetic code. IV. AAA as an asparagine codon in some animal mitochondria. *J. Mol. Evol.* **30**:329–332.
174. Ohama, T., F. Yamao, A. Muto, and S. Osawa. 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.
175. Ohkubo, S., A. Muto, Y. Kawauchi, F. Yamao, and S. Osawa. 1987. The ribosomal protein gene cluster of *Mycoplasma capricolum*. *Mol. Gen. Genet.* **210**:314–322.
176. Ohyama, K., H. Fukuzawa, T. Kohchi, T. Sano, S. Sano, H. Shirai, K. Umesono, F. Shiki, M. Takeuchi, Z. Chang, S. Aota, H. Inokuchi, and H. Ozeki. 1988. Structure and organization of *Marchantia polymorpha* chloroplast genome. I. Cloning and gene identification. *J. Mol. Biol.* **203**:281–298.
177. Ohyama, K., H. Fukuzawa, T. Kohchi, H. Shirai, T. Sano, S. Sano, K. Umesono, Y. Shiki, M. Takeuchi, Z. Chang, S.-I. Aota, H. Inokuchi, and H. Ozeki. 1986. Chloroplast gene organization

- deduced from complete sequence of liverwort *Marchantia polymorpha* chloroplast DNA. *Nature* (London) **322**:572-574.
178. Ohyama, K., Y. Ogura, K. Oda, K. Yamato, E. Ohta, Y. Nakamura, M. Takemura, N. Nozato, K. Akashi, T. Kanegae, and Y. Yamada. 1991. Evolution of organellar genomes, p. 187-198. *In* S. Osawa and T. Honjo (ed.), *Evolution of life: fossils, molecules, and culture*. Springer-Verlag, Tokyo.
 179. Old, J. M., and D. S. Jones. 1975. The recognition of methionine analogues by *Escherichia coli* methionyl-transfer ribonucleic acid synthetase. *Biochem. Soc. Trans.* **3**:659-660.
 180. Osawa, S., D. Collins, T. Ohama, T. H. Jukes, and K. Watanabe. 1990. Evolution of the mitochondrial genetic code. III. Reassignment of CUN codons from leucine to threonine during evolution of yeast mitochondria. *J. Mol. Evol.* **30**:322-328.
 181. Osawa, S., and T. H. Jukes. 1988. Evolution of the genetic code as affected by anticodon content. *Trends Genet.* **4**:191-198.
 182. Osawa, S., and T. H. Jukes. 1989. Codon reassignment (codon capture) in evolution. *J. Mol. Evol.* **28**:271-278.
 183. Osawa, S., A. Muto, T. H. Jukes, and T. Ohama. 1990. Evolutionary changes in the genetic code. *Proc. R. Soc. London Ser. B* **241**:19-28.
 184. Osawa, S., A. Muto, T. Ohama, Y. Andachi, R. Tanaka, and F. Yamao. 1990. Prokaryotic genetic code. *Experientia* **46**:1097-1106.
 185. Osawa, S., T. Ohama, T. H. Jukes, and K. Watanabe. 1989. Evolution of the mitochondrial genetic code. I. Origin of AGR serine and stop codons in metazoan mitochondria. *J. Mol. Evol.* **29**:202-207.
 186. Osawa, S., T. Ohama, T. H. Jukes, K. Watanabe, and S. Yokoyama. 1989. Evolution of the mitochondrial genetic code. II. Reassignment of codon AUA from isoleucine to methionine. *J. Mol. Evol.* **29**:373-380.
 187. Osawa, S., T. Ohama, F. Yamao, A. Muto, T. H. Jukes, H. Ozeki, and K. Umesonono. 1988. Directional mutation pressure and transfer RNA in choice of the third nucleotide of synonymous two-codon sets. *Proc. Natl. Acad. Sci. USA* **85**:1124-1128.
 188. Ozeki, H., K. Ohyama, H. Inokuchi, H. Fukuzawa, T. Kohchi, T. Sato, K. Nakahigashi, and K. Umesonono. 1987. Genetic system of chloroplasts. *Cold Spring Harbor Symp. Quant. Biol.* **52**:791-804.
 189. Pape, L. K., and A. Tzagoloff. 1985. Cloning and characterization of the gene for the yeast cytoplasmic threonyl-tRNA synthetase. *Nucleic Acids Res.* **13**:6171-6183.
 190. Patterson, E. L., R. Milstrey, and E. L. R. Stokstad. 1957. Effect of selenium in preventing exudative diathesis in chicks. *Proc. Soc. Exp. Biol. Med.* **95**:617-620.
 191. Pedersen, N. D., P. D. Whanger, P. H. Weswig, and O. H. Muth. 1972. Selenium binding proteins in tissues of normal and selenium responsive myopathic lambs. *Bioinorg. Chem.* **2**:33-45.
 192. Pinsent, J. 1954. The need for selenite and molybdate in the formation of formic dehydrogenase by members of the *coli-aerogenes* group of bacteria. *Biochem. J.* **57**:10-16.
 193. Prat, A., M. Katinka, F. Caron, and E. Meyer. 1986. Nucleotide sequence of the *Paramecium primaurelia* G surface protein. A huge protein with a highly periodic structure. *J. Mol. Biol.* **189**:47-60.
 194. Preer, J. R., Jr., L. B. Preer, B. M. Rudman, and A. J. Barnett. 1985. Deviation from the universal code shown by the gene for surface protein 51A in *Paramecium*. *Nature* (London) **314**:188-190.
 195. Pritchard, A. E., C. L. Sable, S. E. Venuti, and D. J. Cummings. 1990. Analysis of NADH dehydrogenase proteins, ATPase subunit 9, cytochrome b, and ribosomal protein L14 encoded in the mitochondrial DNA of *Paramecium*. *Nucleic Acids Res.* **18**:163-171.
 196. Pritchard, A. E., J. J. Seilhamer, R. Mahalingam, C. L. Sable, S. E. Venuti, and D. J. Cummings. 1990. Nucleotide sequence of the mitochondrial genome of *Paramecium*. *Nucleic Acids Res.* **18**:173-180.
 197. Read, R., T. Bellew, J. G. Yang, K. E. Hill, I. S. Palmer, and R. F. Burk. 1990. Selenium and amino acid composition of selenoprotein P, the major selenoprotein in rat serum. *J. Biol. Chem.* **265**:17899-17905.
 198. Renbaum, P., D. Abrahamobe, A. Fainsod, G. G. Wilson, S. Rottem, and A. Razin. 1990. Cloning, characterization, and expression in *Escherichia coli* of the gene coding for the CpG DNA methylase from *Spiroplasma sp.* strain MQ1 (M. Sssl). *Nucleic Acids Res.* **18**:1145-1152.
 199. Roe, B. A., D.-P. Ma, R. K. Wilson, and F.-H. Wong. 1985. The complete nucleotide sequence of the *Xenopus laevis* mitochondrial genome. *J. Biol. Chem.* **260**:9759-9774.
 200. Roe, B. A., J. F. H. Wong, E. Y. Chen, and P. W. Armstrong. 1981. Sequence analysis of mammalian mitochondrial tRNAs, p. 167-176. *In* A. G. Walton (ed.), *Recombinant DNA. Proceedings of the Third Cleveland Symposium on Macromolecules*. Elsevier Science Publishing Co., Amsterdam.
 201. Rotruck, J. T., A. L. Pope, H. E. Ganther, A. B. Swanson, D. G. Hafeman, and W. G. Hoekstra. 1973. Selenium: biochemical role as a component of glutathione peroxidase. *Science* **179**:588-590.
 202. Safran, M., A. P. Farwell, and J. L. Leonard. 1991. Evidence that type II 5'-deiodinase is not a selenoprotein. *J. Biol. Chem.* **266**:13477-13480.
 203. Samuelsson, T., Y. S. Guindy, F. Lustig, T. Boren, and U. Lagerkvist. 1987. Apparent lack of discrimination in the reading of certain codons in *Mycoplasma mycoides*. *Proc. Natl. Acad. Sci. USA* **84**:3166-3170.
 204. Sanger, F., G. G. Brownlee, and B. G. Barrell. 1965. A two-dimensional fractionation procedure for radioactive nucleotides. *J. Mol. Biol.* **13**:373-388.
 205. Schneider, S. U., M. B. Leible, and X.-P. Yang. 1989. Strong homology between the small subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase of two species of *Acetabularia* and the occurrence of unusual codon usage. *Mol. Gen. Genet.* **218**:445-452.
 206. Schön, A., C. G. Kannangara, S. Gough, and D. Söll. 1988. Protein biosynthesis in organelles requires misaminoacylation of tRNA. *Nature* (London) **331**:187-190.
 207. Schwarz, K., and C. M. Foltz. 1957. Selenium as an integral part of Factor 3 against dietary necrotic liver degeneration. *J. Am. Chem. Soc.* **79**:3292-3293.
 208. Sebald, W., E. Wachter, and A. Tzagoloff. 1979. Identification of amino acid substitutions in the dicyclohexylcarbodiimide binding subunit of the mitochondrial ATPase complex from oligomycin resistant mutants of *Saccharomyces cerevisiae*. *Eur. J. Biochem.* **100**:599-607.
 209. Seilhamer, J. J., and D. J. Cummings. 1982. Altered genetic code in *Paramecium* mitochondria: possible evolutionary trends. *Mol. Gen. Genet.* **187**:236-239.
 210. Sharp, P. M., and W.-S. Li. 1986. An evolutionary perspective of synonymous codon usage in unicellular organisms. *J. Mol. Evol.* **24**:28-38.
 211. Shepherd, J. C. W. 1984. Fossil remnants of a primeval genetic code in all forms of life? *Trends Biochem. Sci.* **9**:8-10.
 212. Shimayama, T., H. Himeno, J. Sasuga, S. Yokobori, T. Ueda, and K. Watanabe. 1990. The genetic code of a squid mitochondrial gene. *Nucleic Acids Res. Symp. Ser.* **22**:73-74.
 213. Shinozaki, K., M. Ohme, M. Tanaka, T. Wakasugi, N. Hayashida, T. Matsubayashi, N. Zaita, J. Chunwongse, J. Obokata, K. Yamaguchi-Shinozaki, C. Ohto, K. Torazawa, B. Y. Meng, M. Sugita, H. Deno, T. Kamogashira, K. Yamada, J. Kusuda, J. Kusuda, F. Takaiwa, A. Kato, N. Tohdoh, H. Shimada, and M. Sugiura. 1986. The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and expression. *EMBO J.* **5**:2043-2049.
 214. Shuber, A. P., E. C. Orr, M. A. Recny, P. F. Schendel, H. D. May, N. L. Schauer, and J. G. Ferry. 1986. Cloning, expression, and nucleotide sequence of the formate dehydrogenase genes from *Methanobacterium formicicum*. *J. Biol. Chem.* **261**:12942-12947.
 215. Sibler, A. P., G. Dirheimer, and R. P. Martin. 1981. Nucleotide sequence of a yeast mitochondrial threonine-tRNA able to decode the CUN leucine codons. *FEBS Lett.* **132**:344-348.
 216. Sibler, A. P., G. Dirheimer, and R. P. Martin. 1985. Yeast mitochondrial tRNA^{Ile} and tRNA^{Met}_m: nucleotide sequence

- and codon recognition patterns. *Nucleic Acids Res.* **13**:1341–1345.
217. Sibler, A. P., G. Dirheimer, and R. P. Martin. 1986. Codon reading patterns in *Saccharomyces cerevisiae* mitochondria based on sequences of mitochondrial tRNAs. *FEBS Lett.* **194**:131–138.
 218. Smith, M. J., D. K. Banfield, K. Doteval, S. Gorshi, and D. J. Kowbel. 1990. Nucleotide sequence of nine protein-coding genes and 22 tRNAs in the mitochondrial DNA of the sea star *Pisaster ochraceus*. *J. Mol. Evol.* **31**:195–204.
 219. Sprinzl, M., T. Hartmann, J. Weber, J. Blank, and R. Zeidler. 1989. Compilation of tRNA sequences and sequences of tRNA genes. *Nucleic Acids Res.* **17**(Suppl.):1–173.
 220. Stadtman, T. C. 1979. Some selenium-dependent biochemical processes. *Adv. Enzymol.* **48**:1–28.
 221. Stadtman, T. C. 1980. Selenium-dependent enzymes. *Annu. Rev. Biochem.* **49**:93–110.
 222. Stadtman, T. C. 1987. Specific occurrence of selenium in enzymes and amino acid tRNAs. *FASEB J.* **1**:375–379.
 223. Stadtman, T. C. 1990. Selenium biochemistry. *Annu. Rev. Biochem.* **59**:111–127.
 224. Sueoka, N. 1961. Correlation between base composition of deoxyribonucleic acid and amino acid composition of protein. *Proc. Natl. Acad. Sci. USA* **47**:1141–1149.
 225. Sueoka, N. 1962. On the genetic basis of variation and heterogeneity of DNA base composition. *Proc. Natl. Acad. Sci. USA* **48**:582–592.
 226. Sueoka, N. 1988. Directional mutation pressure and neutral molecular evolution. *Proc. Natl. Acad. Sci. USA* **85**:2653–2657.
 227. Sukenaga, Y., K. Ishida, T. Takeda, and K. Takagi. 1987. cDNA sequence coding for human glutathione peroxidase. *Nucleic Acids Res.* **15**:7178.
 228. Tanaka, R., A. Muto, and S. Osawa. 1989. Nucleotide sequence of tryptophan tRNA gene in *Acholeplasma laidlawii*. *Nucleic Acids Res.* **17**:5842.
 229. Uhlenbeck, O. C., L. Bare, and A. G. Bruce. 1984. Anticodon loop substitution in tRNA, p. 163–174. *In* B. F. C. Clark and H. U. Petersen (ed.), *Gene expression*. Alfred Benzon Symposium 1983. Munksgaard, Copenhagen.
 230. Umesono, K., and H. Ozeki. 1987. Chloroplast gene organization in plants. *Trends Genet.* **3**:281–287.
 231. Vold, B. 1985. Structure and organization of genes for transfer ribonucleic acid in *Bacillus subtilis*. *Microbiol. Rev.* **49**:71–80.
 232. Wada, K., S. Aota, R. Tsuchiya, F. Ishibashi, T. Gojobori, and T. Ikemura. 1990. Codon usage tabulated from GenBank genetic sequence data. *Nucleic Acids Res.* **18**(Suppl.):2367–2411.
 233. Wakasugi, T., M. Ohme, K. Shinozaki, and M. Sugiura. 1986. Structure of tobacco chloroplast genes of tRNA Ile(CAU), tRNA Leu(CAA), tRNA Cys(GCA), tRNA Ser(UGA) and tRNA Thr(GGU): a compilation of tRNA genes from tobacco chloroplasts. *Plant Mol. Biol.* **7**:385–392.
 234. Waring, R. B., R. W. Davies, S. Lee, E. Grisi, M. M. Berks, and C. Scazzocchio. 1981. The mosaic organization of the apocytochrome b gene of *Aspergillus nidulans* revealed by DNA sequencing. *Cell* **27**:4–11.
 235. Warrick, H., and J. A. Spudich. 1988. Codon preference in *Dictyostelium discoideum*. *Nucleic Acids Res.* **14**:6617–6635.
 236. Watson, J. D., and F. H. C. Crick. 1953. Molecular structure of nucleic acids. *Nature (London)* **171**:738–740.
 237. Weast, R. C. 1964. *Handbook of chemistry and physics*, 45th ed., p. F82. Chemical Rubber Co., Cleveland.
 238. Weber, A. L., and S. L. Miller. 1981. Reasons for the occurrence of the twenty coded protein amino acids. *J. Mol. Evol.* **17**:273–284.
 239. Weber, F., A. Dietrich, J.-H. Weil, and L. Marechal-Drouard. 1990. A potato mitochondrial isoleucine tRNA is coded for by a mitochondrial gene possessing a methionine anticodon. *Nucleic Acids Res.* **18**:5027–5030.
 240. Weiss, W. A., and E. C. Friedberg. 1986. Normal yeast tRNA^{Gln} (CAG) can suppress amber codons and is encoded by an essential gene. *J. Mol. Biol.* **192**:725–735.
 241. Williams, K. R., and G. Herrick. 1991. Expression of the gene encoded by a family of macromolecular chromosomes generated by alternative DNA processing in *Oxytricha fallax*. *Nucleic Acids Res.* **19**:4717–4724.
 242. Woese, C. R., D. H. Dugre, S. A. Dugre, M. Kondo, and W. C. Saxinger. 1966. On the fundamental nature and evolution of the genetic code. *Cold Spring Harbor Symp. Quant. Biol.* **31**:723–736.
 - 242a. Wolstenholme, D. Unpublished data.
 243. Wolstenholme, D. R., J. L. Macfarlane, R. Okimoto, D. O. Clary, and J. A. Wahleithner. 1987. Bizarre tRNAs inferred from DNA sequences of mitochondrial genomes of nematode worms. *Proc. Natl. Acad. Sci. USA* **84**:1324–1328.
 244. Wolstenholme, D. R., R. Okimoto, J. L. Macfarlane, G. A. Pont, H. M. Chamberlin, J. R. Garey, and N. A. Okada. 1990. Unusual features of lower invertebrate mitochondrial genomes, p. 103–106. *In* E. Quagliariello, S. Papa, F. Palmieri, and C. Saccone (ed.), *Structure, function and biogenesis of energy transfer systems*. Elsevier Biomedical Press, Amsterdam.
 245. Wong, J. T.-F. 1976. The evolution of a universal genetic code. *Proc. Natl. Acad. Sci. USA* **73**:2336–2340.
 246. Wong, J. T.-F., and R. Cedergren. 1986. Natural selection versus primitive gene structure as determinant of codon usage. *Eur. J. Biochem.* **159**:175–180.
 - 246a. Yamao, F., Y. Andachi, A. Muto, T. Ikemura, and S. Osawa. 1991. Levels of tRNAs in bacterial cells as affected by amino acid usage in proteins. *Nucleic Acids Res.* **19**:6119–6122.
 247. Yamao, F., S. Iwagami, Y. Azumi, A. Muto, S. Osawa, N. Fujita, and A. Ishihama. 1988. Evolutionary dynamics of tryptophan tRNAs in *Mycoplasma capricolum*. *Mol. Gen. Genet.* **212**:364–369.
 248. Yamao, F., A. Muto, Y. Kawachi, M. Iwami, S. Iwagami, Y. Azumi, and S. Osawa. 1985. UGA is read as tryptophan in *Mycoplasma capricolum*. *Proc. Natl. Acad. Sci. USA* **82**:2306–2309.
 249. Yokogawa, T., T. Suzuki, T. Ueda, M. Mori, T. Ohama, Y. Kuchino, S. Yoshinari, I. Mitoki, K. Nishikawa, S. Osawa, and K. Watanabe. Serine tRNA for translation of non-universal codon CUG in *Candida cylindracea*: evolutionary implications. *Proc. Natl. Acad. Sci. USA*, in press.
 250. Yokoyama, S., T. Watanabe, K. Murao, H. Ishikura, Z. Yamaizumi, S. Nishimura, and T. Miyazawa. 1985. Molecular mechanism of codon recognition by tRNA species with modified uridine in the first position of the anticodon. *Proc. Natl. Acad. Sci. USA* **82**:4905–4909.
 251. Zinoni, F., A. Birkmann, T. C. Stadtman, and A. Böck. 1986. Nucleotide sequence and expression of selenocysteine-containing polypeptide of formate dehydrogenase (formate-hydrogen-lyase-linked) from *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **83**:4650–4654.
 252. Zinoni, F., J. Heider, and A. Böck. 1990. Features of the formate-dehydrogenase mRNA necessary for decoding of the UGA codon as selenocysteine. *Proc. Natl. Acad. Sci. USA* **87**:4660–4664.