An Extreme Codon Preference Strategy: Codon Reassignment[1]

Siv G. E. Andersson and C. G. Kurland

Department of Molecular Biology, Uppsala University

We argue that in animal mitochondria codon reassignments, such as those for AGA and AGG from arginine to serine or of AUA from isoleucine to methionine, are the result of an interplay between biased mutational forces and selective ones. In particular, there is a marked tendency for animal mitochondria to have very small genomes and to minimize their investment in components required for gene expression. These tendencies are expressed as a reduction in the diversity of tRNA isoacceptor species. In our view, the pressure to simplify tRNA populations, together with mutational bias against certain codons, will account for the codon reassignments observed in animal mitochondria. A parallel to the major codon bias in microorganisms, which likewise tends to reduce the diversity of the tRNA isoacceptor populations under fast growth conditions, may be drawn. Therefore, we suggest that codon reassignments are usefully viewed as an extreme form of codon bias.

Introduction

The genetic code is used in a number of different ways that are characteristic of particular genomic families, and these genomic biases often are colored by the G+C composition of the DNA (Grantham et al. 1980*a*, 1980*b*). For example, among prokaryotes there is a rough correlation between overall codon composition and genomic G+C content (Muto and Osawa 1987). Similarly, the chromosomes of vertebrates are compositional mosaics with regions of DNA that have characteristically different base compositions and correspondingly biased codon preferences (Aota and Ikemura 1986). In addition, within these compositional constraints are groups of genes that tend to have their own particular, extreme codon biases (Sharp et al. 1988). Finally, at the far end of this spectrum there are in the genomes of mitochondria, as well as in those of a few organisms, stop and sense codons that are translated in nonstandard ways (Fox 1987). We are concerned here with the origins of such codon reassignments.

It is clear that mutational bias may drive the composition of genomes to extremes of G+C content—and, accordingly, that such bias must influence codon frequencies (Aota and Ikemura 1986; Kurland 1987; Muto and Osawa 1987). In addition, it has been suggested that such compositional drift generates codon replacements (Jukes et

1. Key words: codon reassignment, mitochondrial tRNA, selection.

Address for correspondence and reprints: C. G. Kurland, Department of Molecular Biology, Uppsala University, Biomedical Center, Box 590, Uppsala, Sweden S751 24.

Mol. Biol. Evol. 8(4):530–544. 1991.

0737-4038/91/0804-0010$02.00

al. 1987; Osawa et al. 1987; Osawa and Jukes 1988, 1989). According to this interpretation, extremes of G+C content, particularly in small genomes, can lead to the disappearance of the rarest codons from the coding sequences. Then, the chance return of the transient codons may be accompanied by reassignment to other amino acids or punctuation signals. This interpretation is useful to explain the reassignment of some codons, but it will not account for other characteristic reassignments, such as those of AGR from arginine to serine, in animal mitochondria.

Comparisons of substitution frequencies between homologous genes suggest that

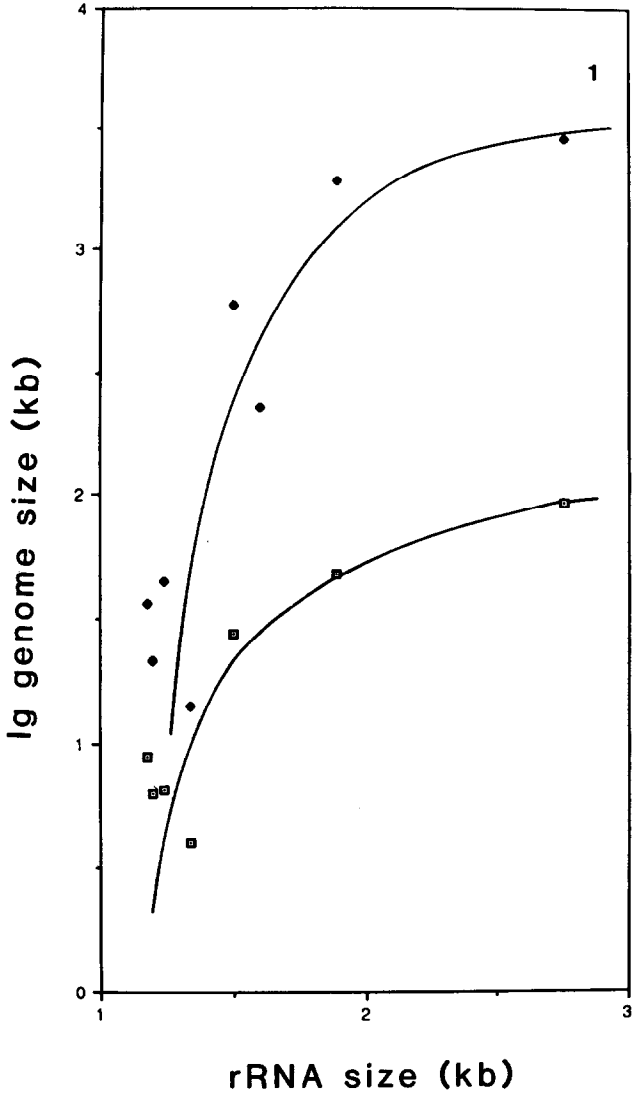

FIG. 1.—Size of 16 S and 23 S rRNAs in relation to genomic size of mitochondria. The curves have been fitted by eye. Data on the sizes of the 16 S rRNAs (♦) are those of Dams et al. (1988), and data on the sizes of the 23S rRNAs (⊡) are those of Gutell and Fox (1988). Included are mitochondrial rRNAs from human, *Xenopus laevis, Drosophila yakuba, Paramecium aurelia, Aspergillis nidulans, Saccharomyces cerevisiae,* and maize.

Table 1
Sense Codon Reassignments in Mitochondria

Biological System	Genome Size (kb)	Sense Codon Reassignment[a] UGA Stop	AUA Ile	AGA Arg	AGG Arg	AAA Lys	CUN Leu
Animaliae:							
Mammals (human)	16	Trp	Met	Stop	Stop	++	++
Amphibians (*Xenopus laevis*)	16	Trp	Met	Stop	−−	++	++
Insects (*Drosophila yakuba*)	16	Trp	Met	Ser	−−	++	++
Echinoderms (*Asterina pectinifera*)	16	Trp	++	Ser	Ser	Asn	++
Nematodes (*Ascaris suum*)	14	Trp	??	Ser	Ser	??	++
Platyhelminths (*Fasciola hepatica*)	14	Trp	??	Ser	Ser	Asn	++
Protista:							
Zoo mastigina (*Trypanosoma brucei*)	22	Trp	++	++	++	++	++
Ciliphora (*Paramecium aurelia*)	41	Trp	++	++	++	++	++
Fungi:							
Molds (*Neurospora crassa*)	60	Trp	++	++	++	++	++
Yeasts (*Saccharomyces cerevisiae*)	78	Trp	Met	++	−−	++	Thr
Higher plants:							
Maize (*Zea mays*)	570	++	++	++	++	++	++

[a] ++ = Standard assignment; −− = codons so far not observed; ?? = not known.

additional selective forces have shaped the marked codon biases of the highly expressed genes in such organisms as *Escherichia coli* and *Saccharomyces cerevisiae* (Sharp and Li 1987). One selective advantage of a major codon bias has been identified by a growth-rate optimization of the translation system that involves a reduction of the complexity of the tRNA populations under favorable growth conditions (Kurland 1987; Andersson and Kurland 1990; Emilsson and Kurland 1990). Here, we wish to suggest that similar selective pressures together with mutational bias drive the evolution of codon reassignments in mitochondria. Some of the relevant facts are as follows:

Codon Reassignments and Genomic Economy

Animal mitochondrial genomes are organized with extreme economy, particularly the mitochondrial DNAs (mtDNAs) of the vertebrates, insects, and echinoderms, in which codon reassignments are most prominent. All of these mitochondria usually have a genome size <20 kb (Anderson et al. 1981; Clary and Wolstenholme 1985; Jacobs et al. 1988), although large size variants have been observed occasionally (Moritz and Brown 1987; Boyce et al. 1989). Furthermore, noncoding information between genes is almost absent, and in several cases the stop codons are added posttranscriptionally (Ojala et al. 1981). Even the translational system reflects this economy. For example, many tRNA genes are exceedingly short in size and lack most of the modifications found in the cytoplasmic systems (Sprinzl et al. 1987). Most striking is the direct correlation between the size of the ribosomal RNA (rRNAs) and the size of the corresponding genomes (fig. 1). For the genomes that are <20 kb the 16 S and 23 S rRNAs are only ~850 and ~1,500 bases long, respectively, as compared with sizes of ~1,500 and ~3,000 bases in genomes >50 kb (Dams et al. 1988; Gutell and Fox 1988).

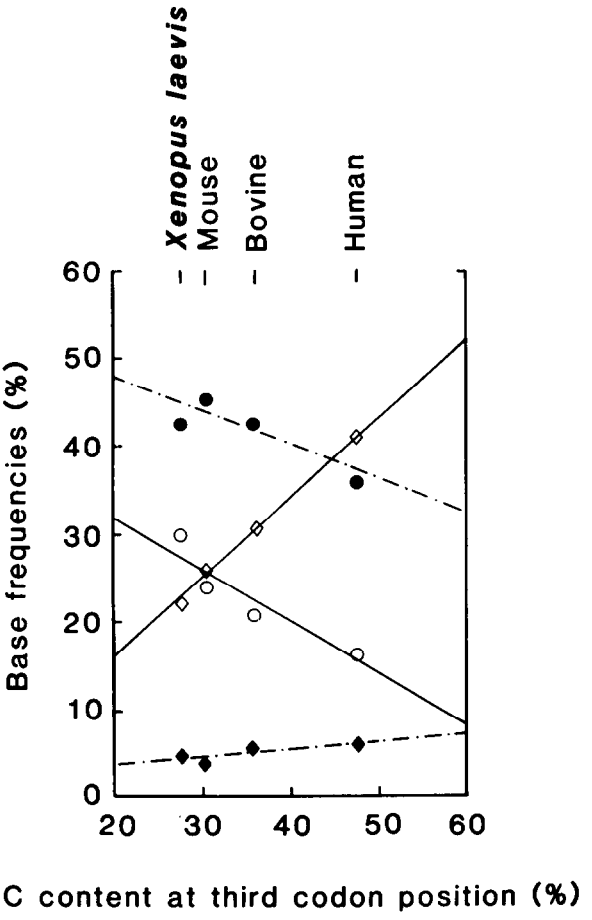

FIG. 2.—Overall codon frequencies in relation to average G+C content at third codon position in vertebrate mitochondria. The lines have been fitted by linear regression. Data have been taken from the references cited in table 3 and are % NNU codons (○), % NNC codons (◇), % NNA codons (●), and % NNG codons (♦).

Thus, the mitochondria of vertebrates, insects, and echinoderms seem to have developed under strong pressure for reduced genome size (Gray and Doolittle 1982). We correlate this tendency with one that minimizes the number of coded tRNA genes (Bulmer 1988; Andersson and Kurland 1990). This tendency can be related to codon reassignments as follows:

A minimal tRNA set would consist of one tRNA per amino acid, i.e., 20 different tRNAs. Indeed, by selecting suitable anticodon loop structures it is apparently possible for a single aminoacyl-tRNA species to translate a four-codon box of the form XYN, and this reduces the number of tRNA species needed to translate all codons (Lagerkvist 1986; Samuelsson et al. 1987). This sort of reduction in the diversity of the tRNA population is limited by the constraints of the conventional code because here a minimum of two different tRNAs are required for the reading of the arginine, isoleucine, serine, and leucine codons. If the codon reassignments have evolved to overcome this

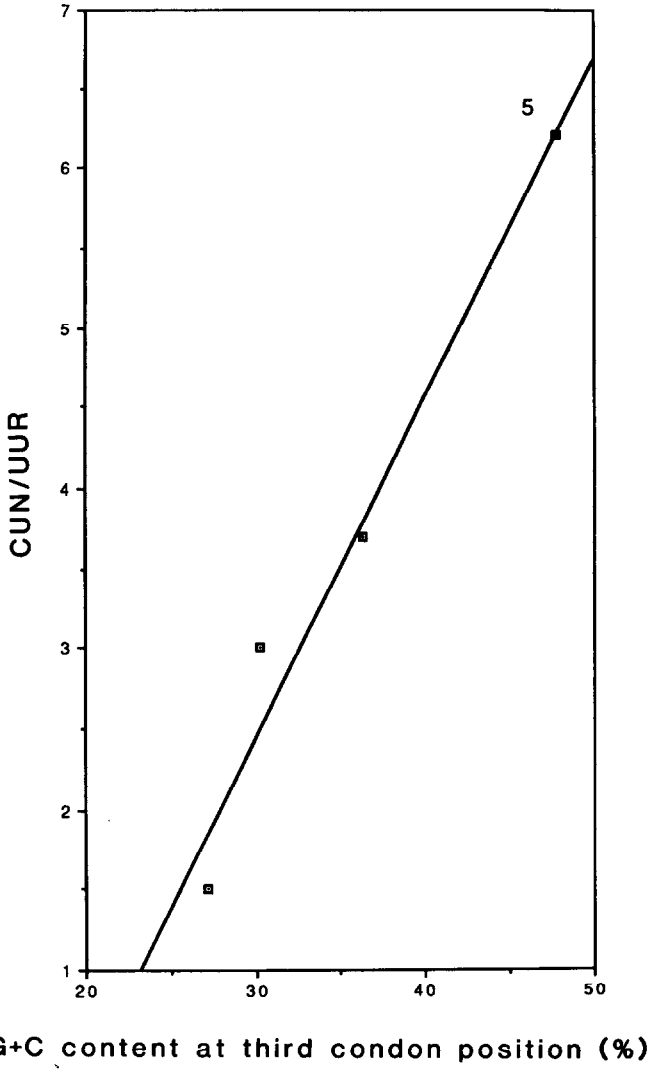

FIG. 3.—Relative frequencies of leucine CUN and UUR codons in relation to average G+C content at third codon position in vertebrate mitochondria. The lines have been fitted by linear regression. Data have been taken from the references cited in table 3.

limitation, they should be confined to these four amino acid families and to the most compact genomes. In fact, a majority of the sense codon reassignments are alterations of arginine and isoleucine codons in the small-sized mitochondrial genomes of vertebrates, insects, echinoderms, nematodes, and platyhelminths. (table 1).

One way of eliminating tRNA genes would be through a conversion of sense codons into stop codons. In vertebrate mitochondria the arginine codons AGA and AGG have been converted into termination codons, thereby eliminating the need for one $tRNA^{Arg}$. The other way of eliminating tRNAs would be to increase the codon degeneracy of the tRNAs already employed. There is evidence that this strategy has been used by the mitochondrial genomes of insects and echinoderms (table 1). These

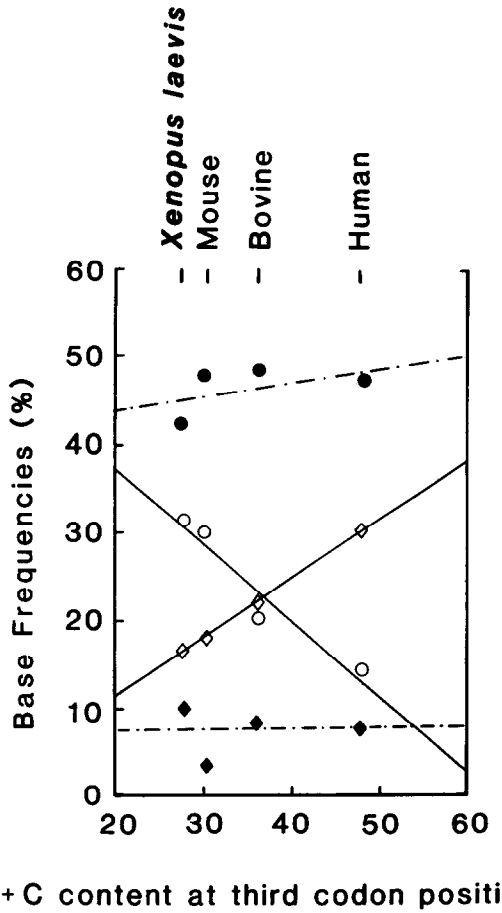

FIG. 4.—Base frequencies in two single-stranded regions of 23S rRNA in relation to G+C content at third codon position in vertebrate mitochondria. The lines have been fitted by linear regression. Data are those of Gutell and Fox (1988) and include the following sequences: human—469–497 and 543–596; Bovine—478–507 and 553–607; mouse—471–507 and 553–606; and *Xenopus laevis*—487–528 and 574–628. Data are % U (○), % C (◇), % A (●), and % G (◆).

genomes have engaged $tRNA^{Ser}$ for the reading of AGA and AGG, which are normally translated as arginine. Finally, the recognition of AUA by $tRNA^{Met}$ eliminates the demand for one additional isoacceptor tRNA, leading to a total of only 22 different tRNAs.

In summary, we favor the interpretation that the reduction of the number of tRNA genes attending the reassignments of the AGA, AGG, and AUA codons is selected during the evolution of the mitochondria (Bulmer 1988; Andersson and Kurland 1990). These coupled reassignments of codons and tRNA species provide a specific example of the more general drive to reduce the size of the mitochondrial genome (Gray and Doolittle 1982).

Codon Reassignments and Codon Frequencies

Since a majority of the sense codon reassignments are associated with the animal mitochondrial genes, we have inspected the codon usage frequencies of these genes

Table 2
Nucleotide Composition of tRNAs and mRNAs, in Relation to Overall Nucleotide Frequencies of Corresponding DNA Strands in Mitochondria of *Xenopus laevis*

	Nucleotide Composition (%)					
Sequence[a]	G	C	A	U	G/C	A/U
L-strand	14	24	33	30	0.6	1.1
$mRNA_H$	13	26	31	26	0.5	1.2
$tRNA_H$	19	22	33	26	0.9	1.3
H-strand	24	14	30	33	1.7	0.9
$mRNA_L$	27	9	19	43	3.0	0.4
$tRNA_L$	27	17	24	32	1.6	0.8

Source.—Roe et al. (1985).

[a] $mRNA_H$ = mRNAs encoded by the H-strand; $tRNA_H$ = tRNAs encoded by the H-strand; $mRNA_L$ = mRNA encoded by the L-strand; $tRNA_L$ = tRNAs encoded by the L-strand.

more closely. The mitochondrial sequences that were examined were human (Anderson et al. 1981), cow (Anderson et al. 1982), mouse (Bibb et al. 1981), toad (Roe et al. 1985), starfish (Himeno et al. 1987), fruit fly (Clary and Wolstenhome 1985), and flatworm (Garey and Wolstenholme 1989).

The mitochondrial codon usage frequencies show that the pattern of codon bias is markedly different in the different taxa. No organelle-specific codon-choice patterns can be identified. Rather, the G+C content at the third codon position varies from <10% in the mitochondria of *D. yakuba* to ~50% in the human mitochondria (Aota et al. 1988). This indicates that codon usage in organellar genes is strongly influenced by biased mutation pressure. Several other observations support this interpretation.

First, the overall frequency of NNC codons versus NNU codons in the human, bovine, mouse, and toad mitochondria varies in direct proportion to the G+C content at the third codon position (fig. 2). However, the frequency of guanosines does not parallel that of cytidines, and the frequency of adenosines is strikingly different from that of uridines. Especially notable is the extremely low (4%–7%) frequency of codons ending in guanosine, compared with the very high (35%–45%) frequency of codons ending in adenosine. Even in the first codon position there is an overrepresentation of adenosines, coupled with a corresponding underrepresentation of guanosines. The frequencies of cytidines and uridines in the first codon position vary as well among the genomes. This can partly be ascribed to differences in the relative usage of UUR codons versus CUN codons (fig. 3).

Second, the bias of base composition is strand specific for the mitochondria, and this relates to the fact that the H-strand is the principle coding strand of these genomes. The base composition bias of single-stranded regions in the 23 S rRNAs, which are coded by the H-strand, resembles that of the messenger RNAs (mRNAs) coded by the same strand (fig. 4). The frequencies of cytidines and uridines vary considerably between the different genomes, while the incidence of adenosines is always high (40%–45%) and that of guanosines is always low (5%–10%) for genes coded by the H-strand. Similarly, tRNAs transcribed from the H-strand have a higher content of adenosines

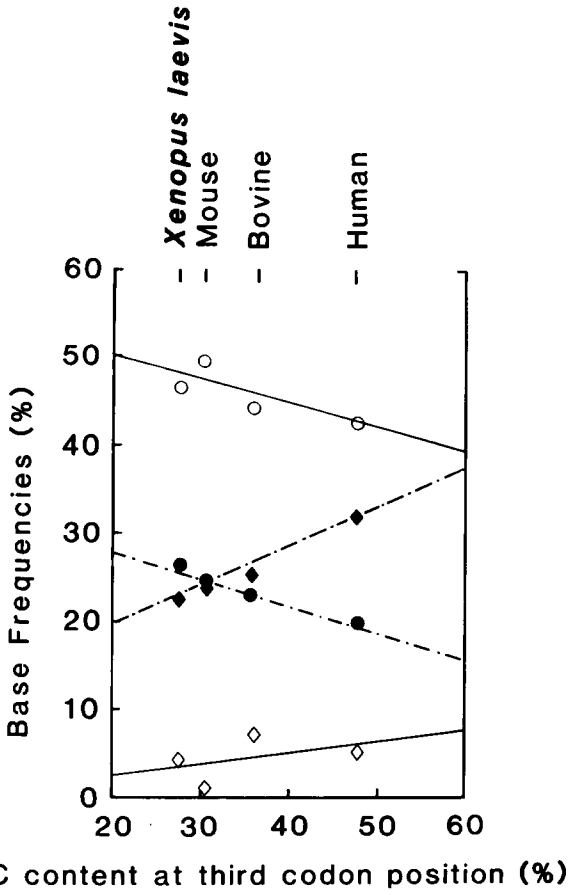

FIG. 5.—Codon frequencies in urf 6 gene in relation to average G+C content at third codon position in vertebrate mitochondria. The lines have been fitted by linear regression. Data have been taken from the references cited in table 3 and are % NNU codons (○), % NNC codons (◇), % NNA codons (●), and % NNG codons (◆).

(33%) than of guanosines (19%). In contrast, tRNAs transcribed from the L-strand tend to be somewhat richer in guanosines than in adenosines. For both groups of tRNAs there is a direct correlation with the overall strand-specific base frequencies (table 2). Likewise, a gene (referred to as urf 6) unusual because it is found on the L-strand has an overrepresentation of uridines, along with an underrepresentation of cytidines (fig. 5 and table 2). Thus, the base composition of the mitochondrial tRNAs, rRNAs, and mRNAs seems to be greatly influenced by mutation pressures which are different for the two DNA strands of these genomes.

The magnitudes of the resulting nucleotide biases are different in different coding positions. For example, the bias is particularly evident in third codon positions and in certain single-stranded regions of the rRNAs. Furthermore, the G+C content is lower in mRNA sequences corresponding to protein regions that are highly variable than it is in regions that are strongly conserved (table 3). This can be attributed to a difference in the relative proportion of adenosines and guanosines. The A/G bias is, on the average, 2.5-fold higher in the variable regions than in the strongly conserved

Table 3
Relative Frequencies of UGA, UGG, and Tryptophan in Organellar Genetic Systems

Biological System	Assignment of UGA	UGA/NNA[a]	UGG/NNG[b]	Trp (%)
Human	Trp	1.0	0.7	2.7
Xenopus laevis	Trp	0.9	1.1	3.1
Drosophila yakuba	Trp	0.8	0.9	2.7
Asterina pectinifera	Trp	1.0	0.6	2.5
Fasciola hepatica	Trp	1.5	1.9	4.0
Trypanosoma brucei	Trp	1.3	1.0	3.4
Paramecium aurelia	Trp	0.7	0.6	1.7
Neurospora crassa	Trp	0.6	0.3	1.5
Saccharomyces cerevisiae	Trp	0.5	1.3	1.7
Average		0.9	0.9	2.6
Maize (mt)	Stop	<0.05	3.1	2.6
Maize (chl)	Stop	0.06	2.0	2.0
Average		<0.05	2.5	2.3

SOURCE.—Clary and Wolstenholme (1985), Roe et al. (1985), Himeno et al. (1987), Aota et al. (1988), and Garey and Wolstenholme (1989).

[a] NNA = average amount of NNA codons/box (UAA, UGA, AUA, and AGA have been excluded from the analysis).

[b] NNG = average amount of NNG codons/box (UAG, UGG, AUG, and AGG have been excluded from the analysis).

regions (table 4). This is consistent with the notion that functionally less important nucleotides are in general more strongly affected by mutation pressures than are the functionally important ones.

We have inspected the codon pairs UGA and UGG, AUA and AUG, and AGA and AGG to determine whether the reassignments have allowed the overall frequencies of adenosines and guanosines characteristic of the different genomes to be expressed in the usage of these particular codons. The usage frequencies of the tryptophan codon UGA, the methionine codon AUA, and the serine codon AGA follow the overall usage frequencies of the NNA type of codons (tables 5–7). Similarly, the codons

Table 4
Relative Frequencies of A and G in Three Codon Positions in Highly Conserved and Highly Variable Regions of Vertebrate Mitochondrial mRNAs

	A/G					
	Conserved Region			Variable Region		
BIOLOGICAL SYSTEM	First Codon Position	Second Codon Position	Third Codon Position	First Codon Position	Second Codon Position	Third Codon Position
Xenopus laevis	0.9	1.2	10.6	2.7	2.4	22.6
Mouse	0.8	1.0	10.0	3.7	3.6	36.5
Bovine	0.7	1.1	9.5	2.6	2.5	12.5
Human	0.7	1.2	6.5	4.1	3.3	20.7

SOURCE.—References cited in table 3.

Table 5
Overall G + C Content in Conserved and Variable Regions of Vertebrate Mitochondrial mRNAs

	G+C Content (%)	
Biological System	Conserved Region[a]	Variable Region[b]
Xenopus laevis	39.8	33.4
Mouse	42.4	34.4
Bovine	46.2	34.8
Human	49.2	37.6

Source.—Anderson et al. (1981, 1982), Bibb et al. (1981), and Roe et al. (1985).

[a] Represents compilation of codons at which same amino acid is present in all four species. The conserved region corresponds to the following amino acid positions in the mitochondrial proteins of *X. laevis:* urf 1—33–37, 136–151, 192–218, 224–229, and 282–293; col—5–20, 22–26, 28–33, 35–40, 64–71, 73–114, 117–134, 138–153, 155–166, 168–175, 77–187, 88–226, 227–250, 256–279, 281–293, 297–301, 303–324, 338–356, 366–389, 423–433, and 435–447; coll—76–85 and 159–170; colll—75–88, 97–103, 128–134, 145–150, 202–214, and 228–238; ATPase 6—144–150 and 210–215; urf 3—61–73; urf 4—132–140, 142–147, and 219–230; urf 5—145–157, 240–265, and 296–321; and cyt b—32–39, 100–107, 130–152, 166–180, 187–195, and 235–243.

[b] Represents compilation of codons at which same amino acid is present in all four species. The variable region corresponds to the following amino acid positions in mitochondrial proteins of *X. laevis:* urf 2—72–103 and 305–345; ATPase 6—19–29, 35–40, 49–83, and 176–196; urf 4—40–62 and 166–190; urf 5—34–45 and 413–425.

UGG, AUG, and AGG are used along with the rest of the NNG codon class. Here it should be recalled that the NNA/NNG ratio is in the range of 10–20 for animal mitochondrial coding sequences. This strengthens the interpretation that the usage frequencies of the reassigned codons are indeed determined by the overall mutation pressures. The reassigned codons AUA, UGA, and AGA are frequently used where adenosines are frequent and are infrequently used where adenine is infrequent. Thus, reassigned codons are not rare codons. Rather, they appear at frequencies higher than or equal to those expected on the basis of the base compositions of the mitochondrial genes in which they appear (table 8).

Codon Reassignments and Amino Acid Abundances

Since the reassignments of AGA, AGG, AUA, and UGA have led to an increased number of codons in the serine, methionine, and tryptophan families, we may ask whether the protein content of these amino acids has been altered correspondingly. The frequency of tryptophan is approximately 2% of all amino acid residues in mitochondrial proteins; this is independent of whether one or two codons are used to specify this amino acid in the mitochondrial genome (table 5). In contrast, the acquisition of AUA by methionine is paralleled by a twofold-increased protein content of that amino acid (table 6). The engagement of two additional serine codons is associated with a slightly higher total abundance of serine (table 7). Tryptophan is

Table 6
Relative Frequencies of AUA, AUG, and Methionine in Organellar Genetic Systems

Biological System	Assignment of AUA	AUA/NNA[a]	AUG/NNG[b]	Met (%)
Human	Met	1.8	2.5	5.5
Xenopus laevis	Met	1.4	2.5	5.2
Drosophila yakuba	Met	1.7	2.6	5.7
Saccharomyces cerevisiae	Met	0.5	24.2	4.7
Average[c]		1.4	6.4	5.3
Asterina pectinifera	Ile	2.6	2.7	1.5
Trypanosoma brucei	Ile	3.2	6.7	3.8
Paramecium aurelia	Ile	1.4	1.2	1.8
Neurospora crassa	Ile	2.2	3.7	1.6
Maize (mt)	Ile	1.1	3.7	3.1
Maize (chl)	Ile	0.7	2.2	2.4
Average[d]		1.9	2.7	2.1

SOURCE.—References cited in table 5.
[a] NNA = average amount of NNA codons/box (UAA, UGA, AUA, and AGA have been excluded from the analysis).
[b] NNG = average amount of NNG codons/box (UAG, UGG, AUG, and AGG have been excluded from the analysis).
[c] Average/codon = 2.6.
[d] Average/codon = 2.3.

the most highly conserved amino acid in the vertebrate mitochondrial genes, and therefore, it is presumably of great functional importance (Anderson et al. 1982). In contrast, methionine, isoleucine, and serine have been found to be interchangeable with each other as well as with alanine, valine, threonine, and leucine (Anderson et al. 1982). Accordingly, hydrophobicity rather than individual amino acid frequencies may here have been the selected parameter.

Table 7
Relative Frequencies of AGA, AGG, and Serine in Organellar Genetic Systems

Biological System	Assignment of AGR	AGA/NNA[a]	AGG/NNG[b]	Ser (%)
Drosophila yakuba	Ser	0.7	<0?	9.0
Asterina pectinifera	Ser	0.8	0.6	10.4
Fasciola hepatica	Ser	0.2	0.8	8.5
Average[c]		0.6	<0.5	9.3
Trypanosoma brucei	Arg	0.6	0.6	6.7
Paramecium aurelia	Arg	0.7	1.1	8.3
Neurospora crassa	Arg	0.9	1.2	7.2
Saccharomyces cerevisiae	Arg	0.6	<0	6.2
Maize (mt)	Arg	0.3	0.3	7.8
Maize (chl)	Arg	0.6	0.2	5.8
Average[d]		0.6	<0.6	7.0

SOURCE.—References cited in table 5.
[a] NNA = average amount of NNA codons/box (UAA, UGA, AUA, and AGA have been excluded from the analysis.
[b] NNG = average amount of NNG codons/box (UAG, UGG, AUG, and AGG have been excluded from the analysis).
[c] Average/codon = 1.2.
[d] Average/codon = 1.2.

Table 8
Relative Usage of Reassigned and Conventionally Assigned Tryptophan, Methionine, and Serine Codons

Biological System	UGA/UGG	AUA/AUG	(AGR × 3)/(AGY+CUN)
Human	8.4	4.2	
Xenopus laevis	6.3	4.6	
Drosophila yakuba	16.0	10.8	0.8
Asterina pectinifera	6.2		0.6
Trypanosoma brucei	0.3		0.6
Paramecium aurelia	4.4		
Aspergillis nidulans	0.7		
Neurospora crassa	9.5		
Saccharomyces cerevisiae	8.8	0.4	

SOURCE.—References cited in table 5.

Discussion

In general, codon usage is determined by a balance between selective and mutational pressures on genomes (Andersson and Kurland 1990). In microorganisms, the codon choices are constrained in highly expressed genes by selective pressures whereas in lowly expressed genes mutational pressures are more evident. In contrast, much has been written to support the interpretation that codon reassignments in mitochondria and elsewhere are determined solely by mutational pressures (Jukes 1985; Jukes et al. 1987; Osawa et al. 1987, 1989*a*, 1989*b*, 1990; Osawa and Jukes 1988, 1989). Nevertheless, we are persuaded that the reassignment of codons is the result of both selective and mutational pressures.

First, while mutational bias will go far in accounting for the tryptophan and methionine codon reassignments, it will not explain by itself the reassignments of the arginine codons AGR to termination or to serine. Second, mutational pressure alone will not explain the simplification, of the tRNA population, that is coupled with codon reassignments in mitochondria. In this context the observed codon reassignments and the reduction of tRNA number represent the smaller part of a larger pattern, which is the overall tendency for genomic reduction.

Discussions of the evolutionary divergence of both the translation system and code in *Mycoplasma* have tended also to be fixed on the compositional bias of the mycoplasmic genome (Jukes 1985; Osawa and Jukes 1988, 1989). It is likely that the pressure to accumulate AT pairs has influenced mycoplasmic evolution. Nevertheless, this is a group of organisms with many other features similar to those of mitochondria: they are degenerate forms of gram-positive bacteria that can multiply as intracellular parasites, and they have reduced genome sizes as well as minimal systems for gene expression (Muto 1987). In particular, a number of the tRNA species of *M. mycoides* have been found to be able to read four-codon boxes of the form XYN and to use anticodon loops of the same sort as those found in homologous tRNA species from mitochondria (Samuelsson et al. 1987). These similarities suggest that the evolution of both *Mycoplasma* mitochondria and animal mitochondria has converged under

pressure for a genomic minimization that seems to be correlated with an intracellular life-style.

In multiple-codon families a reduction of the frequency of guanosines in the third codon position can simply be compensated by an increase in the usage of synonymous codons. In contrast, there are two radically different ways to adjust to a pronounced avoidance of unique codons such as AUG and UGG: one way is to reduce the protein content of methionine and tryptophan, and the other way is to reassign these amino acids to new codons. We also have noted a strong correlation between the reassignment of AUA to methionine and a greater abundance of this amino acid in mitochondrial proteins. This suggests that codon reassignment permits the amino acid composition of proteins to change in a direction that is opposite to that in which standard codon assignments would be driven by mutational bias.

In summary, we find evidence for an interplay of mutational bias and functionally selective pressures in the evolution of the codon reassignments of mitochondria. This interplay is reminiscent of that seen in the evolution of the codon biases of microorganisms. While the major codon bias of microorganisms is constrained by selective pressures on the highly expressed genes, the identity of the preferred codons varies from organism to organism and reflects the compositional bias of each system (Kurland 1987; Muto and Osawa 1987; Bulmer 1988; Andersson and Kurland 1990).

There is now direct evidence that the major codon bias of microorganisms is associated with a tendency to lower the heterogeneity of the tRNA species that are expressed at high concentration at the fastest growth rates (Emilsson and Kurland 1990). Thus, for the eight tRNA isoacceptors that have been fully characterized, it is found that, as the growth rates are increased by systematically changing the quality of the growth media, the concentrations of those that translate major (preferred) codons increase whereas the concentrations of those that translate minor codons decrease. We wish to draw attention to the parallel between the rearrangements of the tRNA population associated with the major codon bias and the simplifications of the tRNA populations in animal mitochondria. In light of this, it seems useful to view the codon reassignments of mitochondria as extreme forms of codon bias.

Finally, a comment should be made about the functional consequences of the simplified translation systems characteristic of many mitochondria. If there were no negative consequences of the tRNA simplifications characteristic of mitochondrial and mycoplasmic systems, it is likely that they would be advantageous for other translation systems, since they would lead to a greater economy in the biomass invested in gene expression (Ehrenberg and Kurland 1984). Therefore, we are persuaded that there must be significant negative consequences for mitochondrial tRNA populations, either with respect to translational efficiency or with respect to precision. Indeed, recent estimates with yeast mitochondria have confirmed this expectation. Here, it has been found that mitochondrial ribosomes translate at rates that are between one and two orders of magnitude slower than those of *Escherichia coli* (M. Zagorski, personal communication).

Acknowledgments

We thank Paul Sharp, Michael Bulmer, Ford Doolittle, Michael Gray, and Manolo Gouy for criticism and helpful discussions. This work was supported by grants from The Swedish Cancer Society and The Natural Science Research Council.

LITERATURE CITED

ANDERSON, S., A. T. BANKIER, B. G. BARELL, 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 **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.

ANDERSSON, S. G. E., and C. G. KURLAND. 1990. Codon preferences in freeliving microorganisms. Microbiol. Rev. **54**:198–210.

AOTA, S., T. GOJOBORI, F. ISHIBASHI, T. MARUYAMA, and T. IKEMURA. 1988. Codon usage tabulated from the GenBank genetic sequence data. Nucleic Acids Res. **16**:315–402.

AOTA, S., 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.

BIBB, M. J., R. A. VAN ETTEN, C. T. WRIGHT, M. W. WALBERG, and D. A. CLAYTON. 1981. Sequence and gene organization of mouse mitochondrial DNA. Cell **26**:167–180.

BOYCE, T. M., M. E. ZWICK, and C. F. AQUADRO. 1989. Mitochondrial DNA in the bark weevils: size, structure and heteroplasmy. Genetics **123**:825–836.

BULMER, M. 1988. Evolutionary aspects of protein synthesis. Oxf. Surv. Evol. Biol. **5**:1–40.

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.

DAMS, E. I., L. HENDRIKS, Y. VAN DE PEER, J.-M. NEEFS, G. SMITS, I. VANDENBEMPT, and R. D. WACHTER. 1988. Compilation of small ribosomal subunit RNA sequences. Nucleic Acids Res. **16**:r87–173.

EHRENBERG, M., and C. G. KURLAND. 1984. Cost of accuracy determined by a maximal growth rate constraint. Q. Rev. Biophys. **17**:45–82.

EMILSSON, V., and C. G. KURLAND. 1990. Growth rate dependence of tRNA abundance in *Escherichia coli*. EMBO J. **9**:4359–4366.

FOX, T. D. 1987. Natural variation in the genetic code. Annu. Rev. Genet. **21**:67–91.

GAREY, J. R., and D. R. WOLSTENHOLME. 1989. Platyhelminth mitochondrial DNA: evidence for early evolutionary origin of a tRNA^{ser}AGN that contains a dihydrouridine arm replacement loop, and of serine-specifying AGA and AGG codons. J. Mol. Evol. **28**:374–387.

GRANTHAM, R., C. GAUTIER, and M. GOUY. 1980*a*. Codon frequencies in 169 individual genes confirm consistent choices of degenerate bases according to genome type. Nucleic Acids Res. **8**:1893–1911.

GRANTHAM, R., C. GAUTIER, M. GOUY, R. MERCIER, and A. PAVE. 1980*b*. Codon catalog usage and the genome hypothesis. Nucleic Acids Res. **8**:49–53.

GRAY, M. W., and W. F. DOOLITTLE. 1982. Has the endosymbiont hypothesis been proven? Microbiol. Rev. **46**:1–30.

GUTELL, R., and G. E. FOX. 1988. A compilation of large subunit RNA sequences presented in a structural format. Nucleic Acids Res. **16**:r175–r269.

HIMENO, H., H. MASAKI, T. KAWAI, T. OHTA, I. KUMAGAI, K. MIURA, and K. WATANABE. 1987. Unusual genetic codes and a novel gene structure for tRNA$^{Ser}_{AGY}$ in starfish mitochondrial DNA. Gene **56**:219–230.

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.

JUKES, T. H. 1985. A change in the genetic code in mycoplasma capricolum. J. Mol. Evol. **22**: 361–362.

JUKES, T. H., S. OSAWA, and A. MUTO. 1987. Divergence and directional mutation pressures. Nature **325**:668.

KURLAND, C. G. 1987. Strategies for efficiency and accuracy in gene expression. I. The major codon preference: a growth optimization strategy. Trends Biochem. Sci. **12**:126–128.

LAGERKVIST, U. 1986. Unconventional methods in codon reading. Bioessays **4**:223–226.

MORITZ, C., and W. M. BROWN. 1987. Tandem duplications in animal mitochondrial DNAs: variation in incidence and gene content among lizards. Proc. Natl. Acad. Sci. USA **84**:7183–7187.

MUTO, A. 1987. The genome structure of *Mycoplasma capricolum.* Isr. J. Med. Sci. **23**:334–341.

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.

OJALA, D., J. MONTOYA, and G. ATTARDI. 1981. tRNA punctuation model of RNA processing in human mitochondria. Nature **290**:470–474.

OSAWA, S., D. COLLINS, T. OHAMA, T. 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.

OSAWA, S., and T. JUKES. 1988. Evolution of the genetic code as affected by anticodon content. Trends Genet. **4**:191–198.

———. 1989. Codon reassignment (codon capture) in evolution. J. Mol. Evol. **28**:271–278.

OSAWA, S., T. H. JUKES, A. MUTO, F. YAMAO, T. OHAMA, and Y. ANDACHI. 1987. Role of directional mutation pressure in the evolution of the eubacterial genetic code. Cold Spring Harbor Symp. Quant. Biol. **52**:777–789.

OSAWA, S., T. OHAMA, T. H. JUKES, and K. WATANABE. 1989*a*. Evolution of mitochondrial genetic code. I. Origin of AGR serine and stop codons in metazoan mitochondria. J. Mol. Evol. **29**:202–207.

OSAWA, S., T. OHAMA, T. H. JUKES, K. WATANABE, and S. YOKOYAMA. 1989*b*. Evolution of the mitochondrial genetic code. II. Reassignment of codon AUA from isoleucine to methionine. J. Mol. Evol. **29**:373–380.

ROE, B. A., D.-P. MA, R. K. WILSON, and J. F.-H. WONG. 1985. The complete nucleotide sequence of the *Xenopus laevis* mitochondrial genome. J. Biol. Chem. **260**:9759–9774.

SAMUELSSON, T., Y. S. GUINDY, F. LUSTIG, T. BORÉN, 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.

SHARP, P. M., E. COWE, D. G. HIGGINS, D. C. SHIELDS, K. H. WOLFE, and F. WRIGHT. 1988. Codon usage patterns in *Escherichia coli, Bacillus subtilis, Saccharomyces cerevisiae, Schizosaccharomyces pombe, Drosophila melanogaster* and *Homo sapiens:* a review of the considerable within-species diversity. Nucleic Acids Res. **16**:8207–8211.

SHARP, P. M., and W.-H. LI, 1987. The rate of synonymous substitution in enterobacterial genes is inversely related to codon usage bias. Mol. Biol. Evol. **4**:222–230.

SPRINZL, M., T. HARTMAN, F. MEISSNER, J. MOLL, and T. VORDEWULBECKE. 1987. Compilation of tRNA sequences and sequences of tRNA genes. Nucleic Acids Res. **15**:r53–r188.

BARRY G. HALL, reviewing editor

Received August 7, 1990; revision received January 18, 1991

Accepted February 4, 1991