tiques, tRNA suppresseurs et mécanismes de régulation. Thèse de Doctorat d'Université Paris 1989.
124 Valle, R. P. C., and Haenni, A. L., Regulation of peptide chain termination in higher eucaryotes; in: Translation in Eucaryotes. Ed. Hans Trachsel. Tedford Press, Caldwell, N.J. 1990 in press.
125 Valle, R. P. C., and Morch, M. D., Stop making sense. FEBS Letters *235* (1988) 1–15.
126 Weiss, W. A., and Friedberg, E. C., Normal yeast $tRNA^{gln}$ can suppress amber codons and is encoded by an essential gene. J. molec. Biol. *192* (1986) 725–735.
127 Weissenbach, J., Dirheimer, G., Falcoff, R., Sanceau, J., and Falcoff, E., Yeast $tRNA^{leu}$ (anticodon UAG) translates all six leucine codons in extracts from interferon treated cells. FEBS Letters *82* (1977) 71–76.
128 Wilson, W., Braddock, M., Adams, S. E., Rathjen, P. D., Kingsman, S. M., and Kingsman, A. J., HIV expression strategies: ribosomal frameshifting is directed by a short sequence in both mammalian and yeast systems. Cell *55* (1988) 1159–1169.
129 Wilson, W., Malim, M. H., Mellor, J., Kingsman, A. J., and Kingsman, S. M., Expression strategies of the yeast retrotransposon Ty: a short sequence directs ribosomal frameshifting. Nucl. Acids Res. *14* (1986) 7001–7016.
130 Woese, C. R., The Genetic Code. Harper & Row, New York 1967.
131 Yamada, K., and Machida, H., Nippon Nogeikagaku Kaishi *36* (1962) 858–860.
132 Yamao, F., Muto, A., Kawauchi, Y., Iwami, M., Iwagami, S., Azumi, Y., and Osawa, S., UGA is read as tryptophan in *Mycoplasma capricolum*. Proc. natl Acad. Sci. USA *82* (1985) 2306–2309.
133 Yoshinaka, Y., Katoh, I., Copeland, T. D., and Oroszlan, S., Murine leukemia virus protease is encoded by the gag-pol gene and is synthesized through suppression of an amber termination codon. Proc. natl Acad. Sci. USA *82* (1985) 1618–1622.
134 Zinoni, F., Birkman, A., Stadtman, T. C., and Bock, A., Nucleotide sequence and expression of the selenocysteine-containing polypeptide of formate dehydrogenase from *Escherichia coli*. Proc. natl Acad. Sci. USA *83* (1986) 4650–4654.

0014-4754/90/11-12/1106-12$1.50 + 0.20/0

The genetic code in mitochondria and chloroplasts

T. H. Jukes and S. Osawa

Space Sciences Laboratory, University of California/Berkeley, 6701 San Pablo Avenue, Oakland (California 94608, USA) and Nagoya University, Laboratory of Molecular Genetics, Department of Biology, Nagoya 464 (Japan)

Summary. The universal genetic code is used without changes in chloroplasts and in mitochondria of green plants. Non-plant mitochondria use codes that include changes from the universal code. Chloroplasts use 31 anticodons in translating the code; a number smaller than that used by bacteria, because chloroplasts have eliminated 10 CNN anticodons that are found in bacteria. Green plant mitochondria (mt) obtain some tRNAs from the cytosol, and genes for some other tRNAs have been acquired from chloroplast DNA. The code in non-plant mt differs from the universal code in the following usages found in various organisms: UGA for Trp, AUA for Met, AGR for Ser and stop, AAA for Asn, CUN for Thr, and possibly UAA for Tyr. CGN codons are not used by *Torulopsis* yeast mt. Non-plant mt, e.g. in vertebrates, may use a minimum of 22 anticodons for complete translation of mRNA sequences. The following possible causes are regarded as contributing to changes in the non-plant mt: directional mutation pressure, genomic economization, changes in charging specificity of tRNAs, loss of release factor RF2, changes in RF1, changes in anticodons, loss of lysidine-forming enzyme system, and disappearance of codons from coding sequences.
Key words. Genetic code; mitochondria; evolution; organelles.

Introduction

Several differences from the universal genetic code have been discovered in mitochondria (mt) of organisms other than green plants. Chloroplasts and green plant mt have retained the universal code, but have undergone genomic economization so that some anticodons in chloroplasts have been discarded, and, in green plant mt, some tRNAs are imported from the cytosol, and some tRNA genes have been acquired from chloroplast DNA.

We have proposed[25] that evolutionary changes in the universal code must have taken place after a codon disappeared from coding sequences, because an abrupt change in assignment of a codon would be disruptive and therefore lethal. Disappearance of a codon, according to our proposal, is accompanied by a change in or loss of the corresponding anticodon. Such changes in the codon or anticodon may sometimes result from directional mutation pressure towards either AT or GC. The deleted codon reappears with a different assignment at locations in the coding sequence corresponding to the amino acid to which it has been reassigned. Examples of these proposed events will be described in this review, occurring in non-plant mt. Similar changes in nuclear codes are described by Jukes elsewhere in this issue of Experientia.

Evolution of mitochondria

The origin of mitochondria (mt) is inferred to be from endosymbiotic bacteria (reviewed by Küntzel and Köchel[15]). Comparison of 16S ribosomal RNA sequences led Yang et al.[37] to conclude that the endosym-

biont that gave rise to mt belonged to the alpha subdivision of purple bacteria, of which a current representative is *Agrobacterium tumefaciens*. Their ribosomal RNA comparisons showed wheat mitochondria branching separately from the line of descent that showed a later divergence between descents of vertebrate and fungal mt. Separate origins for plant, fungal and vertebrate mt are supported by other publications, e.g., Gray et al.[7], and Küntzel and Köchel[15] propose separate bacterial origins for fungal and animal mt.

Plastids, including chloroplasts (cp), are descended from cyanobacteria, independently of the origin of mt[7,8,20].

Plant mitochondria

Plant mt genomes are large, 200–2400 kbp, but contain only a small number of tRNA genes. These are divided into 2 groups. The first ('native') group is considered to be descended from the original symbiotic eubacteria that provided the ancestral DNA of the mt genome. These tRNA genes show only limited sequence similarity (65–80%) to the corresponding eubacterial and chloroplast tRNAs. The second group is clearly related (> 90% sequence identity) to tRNA genes in chloroplasts, and represent cp DNA that has migrated into plant mt. These tRNA genes have been shown to be functional. A third category is tRNA genes in the cell nucleus that are transcribed into tRNA molecules which are imported from the cytoplasm into plant mt.

The genetic code in plant mt is identical with the universal code. Unlike animal and fungal mt, plant mt use codons UGA for stop and, unlike mt of yeast and several animals, plant mt use AUA for isoleucine.

Green and co-workers[9] found a tRNA Leu (NAA) in bean mt that was identical to its cytoplasmic counterpart except for one post-transcriptional modification. Maréchal and co-workers[18] found that four tRNA Leu (NAG) species in bean mt were nuclear-encoded. Several other tRNAs in bean mt (Phe, Trp, Tyr, f-Met, e-Met and Pro) had a high degree of sequence similarity to their chloroplast counterparts leading to the conclusion that they were derived from chloroplast tRNA genes that had been incorporated in mt DNA.

Joyce and Gray[13] examined wheat mt tRNA genes, and sequenced chloroplast-like tRNA genes for Ser (GGA), Phe (GAA) and cysteine (GCA) that were concluded to be 'remnants of 'promiscuous' chloroplast DNA that has been incorporated into wheat mt DNA' during evolution, as deduced from sequence similarities. They added chloroplast-like e-Met, Asn and Trp tRNA genes to this list. Other tRNA species corresponded to native mt and not to cp genes (table 1), and still others were cytoplasmic (encoded by nuclear DNA). Their various findings for chloroplast-like tRNA genes are in table 1.

Maréchal and colleagues[19] extracted tRNAs from potato mt and found 15 'typically mitochondrial', 5 chloroplast-like and 11 nuclear-encoded species (table 1). Of much evolutionary interest was the fact that, in contrast

Table 1. Genetic origin of mitochondrial anticodons in green plants

		Origin			
Amino acid	Anticodon	Native mt	Chloroplast-like mt-encoded	Nuclear	Anticodon present in vertebrate mt
Ala	IGC			P	–
Arg-1	ICG			P	–
Arg-2	NCU			P	–
Asn	GUU		W,P,L		+
Asp	GUC	P			+
Cys	GCA	P,T	W,M		+
Gln	UUG	W,P			+
Glu	UUC	W,P,S			+
Gly	GCC	P		W	–
His	GUG		P,M	W	+
Ile 1	*CAU	P			
Ile 2	GAU?			P	
Leu	CAA			W	–
Leu 1	NAA			P	
Leu 2 to 5	?			P	
Lys	UUU	W,P			+
eMet	CAU	W,M	W,P,A,B,S		+
fMet	CAU	W,P			+
Phe	GAA	P	W		+
Pro	UGG	W,P			+
Ser	GCU	W,P			+
Ser	UGA	W,P			+
Ser	GGA		W,P,O		–
Thr 1,2	?			P	
Trp	CCA		W,P,M,B,O		–
Tyr	GUA	W,P			+
Val 1,2		P			
Val	GAC			W	–

A = *Arabidopsis*; B = bean; O = *Oenothera*; P = potato; S = soybean; T = tomato; *C = lysidine.

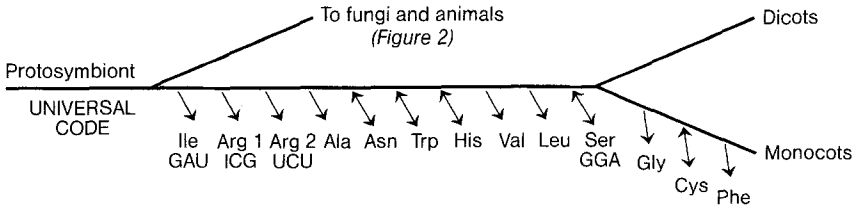

Figure 1. Plant mitochondrial tRNAs changed in evolution. Native tRNA gene in plant mt exchanged for chloroplast gene. Native tRNA gene disappeared. tRNA obtained from nuclear tRNA in cytosol. Anticodons are shown in 4 cases.

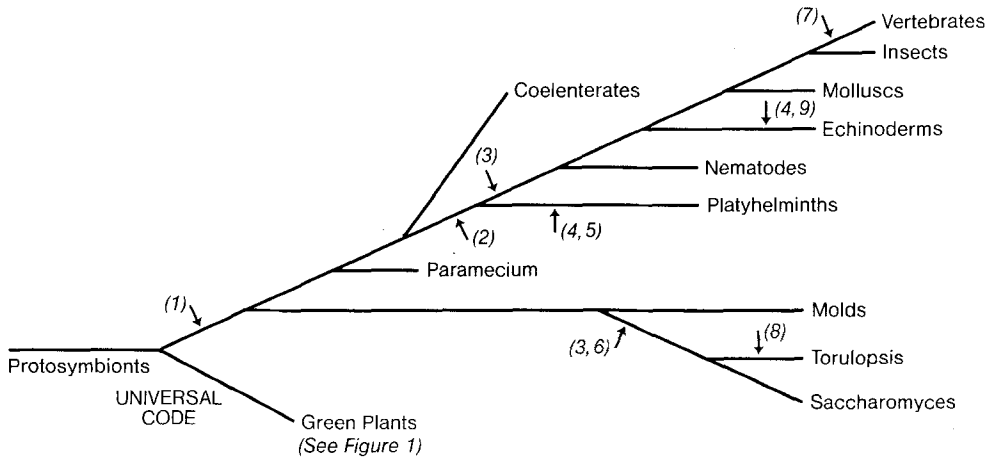

Figure 2. Evolution of code in non-plant mt. (1) UGA, Trp; (2) AGR, Ser; (3) AUA, Met; (4) AAA, Asn; (5) UAA, Tyr (?); (6) CUN, Thr; (7) AGR, Stop; (8) CGN, non-coding; (9) AUA, to Ile from Met.

to wheat, potato mt did not seem to contain chloroplast-like tRNAs for Cys and Phe. This might indicate that the insertion of these two tRNAs took place after the evolutionary separation of monocots and dicots (fig. 1). Maize also had a chloroplast-like tRNA for Lys (table 1). It is also of interest that glycine anticodon GCC listed as native in potato mt is not found in vertebrate mt, which use UCC for glycine, but wheat mt import nuclear-encoded glycine tRNA (GCC) from the cytoplasm.

Plant mitochondrial anticodons identical with those in the vertebrate mt code are indicated in table 1. One exception is GCC, not found in the vertebrate mt code. It is conspicuous that anticodon CCA, apparently obtained from chloroplasts, is used by plant mt for tryptophan, because CCA pairs only with UGG, and plant mt use UGG for Trp and UGA for stop, in contrast to non-plant mt, which use both UGG and UGA for tryptophan, pairing with anticodon UCA. In retrospect, the finding that CCA, pairing only with UGG, is the sole tryptophan anticodon in plant mt may be compared with the discovery that CGG is edited to UGG in plant mt mRNA (reviewed by Jukes in this issue). Anticodons IGC and ICG, present in potato mt, have not been found in non-plant mt. Both are eukaryotic anticodons and ICG is also eubacterial.

Weber and co-workers[34] found that a potato mitochondrial tRNA Ile, translating AUA, contained a lysidine-like nucleotide in position 1 of the anticodon CAU. The C in the anticodon was therefore modified post-transcriptionally, just as in the cases of *E. coli*[21] and *Mycoplasma capricolum*[1], both of which have isoleucine tRNAs with *C (lysidine) in anticodon *CAU, pairing with AUA. The gene in potato mt differed by only one nucleotide from a maize mt tRNA gene with anticodon CAT[30], putatively described as an elongator methionine tRNA, but perhaps this tRNA is also modified post-transcriptionally to function as an isoleucine tRNA *CAU, translating codon AUA.

Use of *CAU as an anticodon for isoleucine codon AUA has been previously found in bacteria, and may also occur in chloroplasts. In contrast, eukaryotes translate AUA, isoleucine, by anticodon IAU, or, in yeast, UAU. Anticodon CAU, unmodified, is used for methionine in all codes.

The observation by Weber and colleagues[34] serves to reinforce the conclusion that the protosymbiotic ancestor of mt was bacterial. The observation also provides an explanation for the assignment of AUA to methionine in yeast mt and in the mt of most metazoa (fig. 2). Conversion of tRNA Met CAU to tRNA Ile *CAU does not occur unless lysine is added to C. Loss of the enzyme system that forms lysidine would convert AUA from an isoleucine to a methionine codon. This would need to be preceded by removal of AUA codons from coding se-

Table 2. Anticodons in the eubacterial code, vertebrate mitochondrial code and chloroplast code

M,C	GAA Phe	C	GGA Ser	M,C	GUA Tyr	M,C	GCA Cys
M,C	UAA Leu	M,C	UGA Ser				
C	CAA Leu		CGA Ser			C	CCA Trp[a]
						C	ICG Arg[a]
	GAG Leu	(C)	GGG Pro	M,C	GUG His		
M,C	UAG Leu	M,C	UGG Pro	M,C	UUG Gln		
	CAG Leu		CGG Pro		CUG Gln	C	CCG Arg
M,C	GAU Ile	C	GGU Thr[b]	M,C	GUU Asn	M,C	GCU Ser
C	*CAU Ile	M,C	UGU Thr	M,C	UUU Lys	C	UCU Arg
M,C	CAU Met		CGU Thr		CUU Lys		CCU Arg
C	GAC Val		GGC Ala	M,C	GUC Asp	C	GCC Gly
M,C	UAC Val	M,C	UGC Ala	M,C	UUC Glu	M,C	UCC Gly
	CAC Val		CGC Ala		CUC Glu		CCC Gly

M = vertebrate mitochondria; C = chloroplast. ICG (from ACG) and *CAU (from CAU) are inferred to exist in chloroplasts. (C) The chloroplast gene containing anticodon GGG is apparently inactive. *C, modified C, probably lysidine. [a] UCA is the sole vertebrate mitochondrial anticodon for tryptophan and UCG for arginine. UCA is the anticodon for tryptophan in *Mycoplasma*. [b] AGU is anticodon for threonine in *Mycoplasma*.

Table 3. Differences from the universal code in non-plant mitochondria

Organisms	UGA Stop	AUA Ile	AAA Lys	AGR Arg	CUN Leu	UAA Stop	Examples
Vertebrates	Trp	Met	Lys	Stop	Leu	Stop	Various
Arthropods	Trp	Met	Lys	Ser(AGA)	Leu	Stop	*Drosophila*
Echinoderms	Trp	Ile	Asn	Ser	Leu	Stop	Sea urchins, starfishes
Molluscs	Trp	Met	Lys	Ser	Leu	Stop	*Doryteuthis*
Nematodes	Trp	Met		Ser			*Ascaris, Caenorhabtidis*
Platyhelminths	Trp	Ile	Asn	Ser	Leu	Tyr?	*Fasciola, Dugesia, Planaria*
Coelenterates				Arg			*Hydra, Metridium*
Yeasts	Trp	Met	Lys	Arg	Thr	Stop	*Saccharomyces, Torulopsis**
Euascomycetes	Trp	Ile	Lys	Arg	Leu	Stop	*Aspergillus, Neurospora*
Protozoa	Trp	Ile	Lys	Arg	Leu	Stop	*Trypanosoma, Paramecium*

*CGN codons are non-coding in *Torulopsis*.

quences by conversion to AUY isoleucine codons. AUA, reappearing by mutation of AUY to AUA, would be translated as methionine.

The entire set of tRNA genes for the chloroplast code was found when *Marchantia* chloroplast DNA was completely sequenced[29]. There are 31 anticodons in the chloroplast code (table 2).

Mitochondria and chloroplasts are regarded as descendants of endosymbiotic bacteria that took up residence in proto-eukaryotic cells. We have proposed that these two organelles underwent genomic economization aided by AT pressure[24]. This eliminated most tRNAs with CNN anticodons and, in the case of non-plant mitochondria, it also eliminated GNN anticodons in family boxes, and, to replace these, UNN anticodons paired with 4 codons in family boxes by four-way wobble. By these deletions, the genetic code in vertebrate mitochondria was reduced to the minimum of 22 anticodons needed for the usual 20 amino acids (table 2).

Anticodon UCU disappeared, and AGR became stop codons. Changes in the code took place in some mt so that AUA and UGA became additional codons for methionine and tryptophan respectively instead of for isoleucine and stop as in the universal code.

In this scenario, it is assumed that the mitochondrial code was originally identical with the universal code in eubacteria, and underwent a series of evolutionary changes that differ in the mitochondria of various species, as explained below.

After elimination of most tRNAs with CNN anticodons the chloroplast code remained identical with the eubacterial code except that UAG and UGC became the only anticodons in the leucine (CUN) and alanine family boxes, pairing by 4-way wobble. The same is probably true for UGG, proline, because anticodon GGG is inactive in chloroplasts[29,36]. The complete DNA sequences of chloroplast genomes from tobacco[36] and rice (Sugiura, personal communication) indicated that there exists only one species of tRNA gene for the CGN arginine family box, tRNA ArgACG. No tRNA ArgCCG was found in spite of a considerable usage of codon CGG. It remains to be seen if tRNA ArgACG (modification unknown) can read all the CGN codons or tRNA ArgCCG is imported into the chloroplasts.

The anticodons used in the mitochondrial and chloroplast codes resemble the eubacterial code rather than the eukaryotic code because they do not use INN anticodons in family boxes. Chloroplasts use GNN and UNN anticodons for valine, serine (UCN codons) and threonine. Eukaryotes use INN anticodons in these cases.

Non-plant mitochondrial codes have evolved to different endpoints, as shown in table 3. We have proposed steps

in these various evolutionary pathways in a series of 4 publications[22,24,26,27]. These are discussed below.

From the information so far available, the coding system used by green plant mitochondria (table 1) is as follows:

a) There are no departures from the universal code;

b) Coding for alanine, arginine (CGN and AGR codons) leucine and threonine is exclusively by tRNAs of nuclear origin imported from the cytosol into mitochondria. Nuclear tRNAs for glycine, isoleucine and valine are in some cases imported from the cytosol and in others are transcribed from genes in mt DNA.

c) Mitochondrial tRNA genes of 'native' mt DNA furnish tRNAs for aspartic acid, glutamic acid, glutamine, lysine, initiator methionine, proline, serine (AGY codons) tyrosine and valine exclusively;

d) Mitochondrial tRNA genes imported from chloroplasts are used exclusively for tRNAs for asparagine, histidine and tryptophan;

e) Coding for cysteine and phenylalanine is by tRNAs with genes in the mt genome of either native or chloroplast origin, depending on the species of organism. Monocots use tRNAs from genes imported from chloroplasts for these amino acids, dicots use native genes;

f) Coding for elongator methionine and serine (UCN codons) is by tRNAs of both mitochondrial and chloroplast origins.

g) Anticodon *CAU (*C = lysidine) is used for AUA, isoleucine.

h) In the unicellular green alga *Chlamydomonas reinhardtii*, only 3 tRNA genes have been detected: for tryptophan, glutamine, and e-methionine. The following codons were absent from the five protein-coding genes: TTA, CTC, ATA, TCR, ACR, GAA, CGG, AGR and GGG[5]. Presumably, *C. reinhardtii* mt import some tRNAs from the cytosol. Gray and Boer suggest 'a separate evolutionary origin of at least the ribosomal RNA genes in *C. reinhardtii* and plant mt'.

The coding system for plant mt is, therefore, a mosaic of anticodons derived from nuclear, chloroplast and mitochondrial genes. The pattern varies with the species of plant and it is probable that it represents an evolutionary flux that is still in progress, and for which the background is unknown.

This separation into various categories under headings (b) through (f) above refers to published information. Further research will probably uncover more examples of all these categories, perhaps in wheat as well as in mitochondria of plants that have not yet been examined, or that only have been partially examined.

Plant mt have followed an entirely different evolutionary pathway from that of animal and fungal mt. Plant mt have imported 'promiscuous' DNA from chloroplast mt carrying tRNA genes, and plant mt also import preformed tRNAs of nuclear origin. The coding system of animal and fungal mt is, in contrast, usually kept separate from that of the nucleus, although importation of nuclear-encoded tRNAs must occur in protozoa: *Tetrahymena*[32] and *Paramecium*[31] mt DNA contain genes for only 8 and 3 species of tRNA, respectively. Evidently, many nuclear-encoded tRNAs are used in these cases. Of course, there is no possibility of acquiring DNA from chloroplasts in the case of non-plant mt.

The origin of mitochondria can be represented as descent from a single proto-symbiont with subsequent branching, or as occurring at least three separate times. The second possibility poses complications, because it implies either that plants, fungi and animals do not have a common ancestor, or that an established population of mitochondria can be displaced by a new species of bacterium that enters cells of eukaryotes. Küntzel and Köchel[15] consider that 'The idea of an independent invasion of different bacteria into already diverged protokaryotic precursors to animal and fungal cells is attractive, because the mitochondrial genomes of the two kingdoms differ in several fundamental properties like gene organization, tRNA structure, genetic code and codon usage'. This may be countered by pointing out that tRNA structure and genetic code vary widely among different metazoan mt, so that the differences between mt genomes of animal and fungal cells may well be the result of divergent evolution rather than of separate origins of animal and fungal mitochondria.

Divergence of plant, fungal and animal mt from a common ancestor, based on comparison of ribosomal RNA sequences, was illustrated in figure 2 of the article by Yang et al.[37]. We have illustrated a similar divergence in figures 1 and 2, which shows that the change in function of UGA from stop to tryptophan occurred after the separation of the plant line.

Evolution of the code in non-plant mitochondria

As evolution progressed in fungi and animals, a series of changes took place in the mitochondrial genetic code. The first change was codon UGA from stop to Trp. This was evidently an early change, because UGA codes for Trp in all non-plant mt. In the animal line, AGR arginine codons became serine starting in *Platyhelminthes* and continuing as serine in nematodes, echinoderms, molluscs and insects until AGR became stop in vertebrates (fig. 2). AAA changed from lysine to asparagine in the branch leading to *Platyhelminthes* and again, as a separate event, in *Echinodermata*. There is some indication that UAA, stop, codes for tyrosine in *Platyhelminthes*. Beginning with nematodes (fig. 2), AUA changed from isoleucine to methionine. This change was reversed in echinoderms (fig. 2).

In the fungal line, changes in yeast mt codes are AUA from isoleucine to methionine and CUN from leucine to threonine. CGN has lost its coding function in *Torulopsis*. No changes from the universal code other than UGA to tryptophan have taken place in molds (fig. 2). The changes in non-plant mt codes are listed in table 3.

Tryptophan, isoleucine and lysine codons all participate in mitochondrial code changes.

We have proposed that these and other changes in the code result from some general pattern of evolutionary events. These include directional mutation pressure, disappearance of codons from coding sequences, changes in anticodons, changes in charging specificity of tRNAs and changes in release factor RF1. The events are followed by reversal of directional mutation pressure and reappearance of a deleted codon. Our proposal, the codon capture hypothesis, provides for codon reassignment without disruptive changes such as alterations in the amino acid sequence of a protein.

AUA for methionine in yeast

Eubacteria use tRNA Ile *CAU for translating AUA as isoleucine, in which *C is lysidine (2-lysyl) cytidine. When *C is experimentally replaced by unmodified C, the tRNA becomes charged by methionine instead of isoleucine. Apparently the tRNA can function as tRNA Met in the absence of the C-modifying (lysylating) enzyme. AUA is a codon for isoleucine and AUG is the codon for methionine in mt of molds (*Aspergillus* and *Neurospora*) but anticodon *CAU has not been found in their mt. However, genes for three species of tRNAs with CAU anticodons have been found in *Aspergillus nidulans* mitochondria. The first one is the initiator methionine tRNA, and the second is the primary elongator methionine tRNA. The function of the third one (tRNA Met 3, of Köchel et al.[14] see ref. 27) has not been identified, but this one may be isoleucine tRNA with modified C (possibly lysidine) at the first anticodon position which would be specific for codon AUA, because no tRNAs other than this have been found for isoleucine codon AUA. Yeast mitochondria, where AUA and AUG are both read as methionine, have only two tRNAs with CAU anticodons. These are elongator and initiator methionine tRNAs. Perhaps tRNA for isoleucine codon AUA disappeared from the yeast mt line before or during the reassignment of AUA to methionine as a result of genomic economization. This reassignment would have been when AUA was removed from coding sequences by conversion to other isoleucine codons AUY[27].

The lysylating enzyme possibly decreased in amount or lost activity so that anticodon CAU was no longer modified to *CAU. As CAU, it then would translate only AUG, and its tRNA would be aminoacylated by methionine. This 'new' tRNA would replace the old tRNA Met (CAU) as the elongator methionine tRNA. The initiator methionine tRNA would be unaffected.

AUA from methionine to isoleucine in echinoderms

In sea urchin mitochondria, there are only two tRNA genes for AUN codons, one with anticodon GAU and the other with CAU. CAU is the same anticodon as that of methionine tRNA in other metazoan mitochondria, where both AUA and AUG are methionine codons. The echinoderm tRNA with anticodon CAU would have changed so as to read only codon AUG as methionine. As a result, in echinoderms, (starfish, and sea urchin), tRNA Ile (GAU) would translate AUA in addition to AUY as isoleucine. Pairing with codons AUA and AUY by tRNA Ile (GAU) could result from modification of G at the first anticodon nucleotide, perhaps to inosine (I) as noted by Cantatore and co-workers[3] and Himeno et al.[10]. Echinoderm mt apparently use modified G in both anticodons GAU and GUU for methionine and asparagine respectively to pair with A. Both changes can be explained by assuming GC pressure causing conversion of AAA and AUA codons to AAG and AUG, followed by lowering of GC pressure, so AAY and AUY codons mutated to AAA (asparagine) and AUA (isoleucine) codons. Alternatively, it is possible that methionine and lysine tRNAs gradually underwent structural changes including mutation of anticodon UUU to CUU (in lysine tRNA), so that codons AUA and AAA disappeared and were later captured by isoleucine and asparagine by modification of G in anticodon GUA (Ile) and GUU (Asn), respectively.

AAA for asparagine in echinoderms and platyhelminths

AAA, a lysine codon in the universal code, is an asparagine codon in mt of a starfish[10] and a sea urchin[12], as predicted from comparisons of amino acid sequences in proteins with the corresponding amino acid sequences in mt of other species. These comparisons show AAA sites in mt of echinoderms corresponding to AAY asparagine sites, and not to AAR lysine sites in other mt. AAG in echinodermal mt codes exclusively for lysine.

We postulated[22] that, in the change of AAA from Lys to Asn, GC pressure converted all AAA lysine codons to AAG, pairing with anticodon CUU, so that anticodon UUU disappeared because it was not needed or used. Anticodon GUU, pairing with asparagine codons AAU and AAC, became modified, perhaps by conversion of G to H (hypoxanthine), the base in inosine (I)[3,10]. IUU would pair with AAA as well as AAU and AAC. AT pressure then replaced GC pressure and some AAC asparagine codons mutated to AAA, pairing with IUU. No changes in amino acid sequences took place. A similar series of events took place in *Fasciola*[22]. In this scheme, disappearance of a codon and of its anticodon took place under GC pressure. Another anticodon became modified. The codon reappeared, under AT pressure, with a new assignment.

In the proposal by Ohama et al.[22], it is pointed out that GUU would pair with AAU, AAC and AAA, while CUU would pair only with AAG in echinoderms and *Fasciola*. In insect mt, CUU pairs with both AAA and AAG, lysine. In bacterial and eukaryotic nuclear systems, AAA and AAG pair with *UUU (*U modified), while CUU, C unmodified, pairs solely with AAG as in

the case of echinoderm mt. These differences may result from changes in tRNA molecules as discussed by Ohama et al.[22].

UAA from Stop to Tyr

UAA is found in the COI gene of an individual *Planaria* mt at a site corresponding to a highly conserved UAY tyrosine site in other individuals (Bessho et al., unpublished observations). Moreover, the only stop codon found so far in *Fasciola* is UAG[35]. Presumably chain termination by release factor RF1, which responds to UAA and UAG, may have become specific for UAG as a result of mutations accumulating in RF1 of platyhelminths. RF2, which responds to both UAA and UGA, has not been detected in mt[16] and, indeed, UGA would not be recognized as a stop codon in non-plant mt, because UGA in these mt is a codon of Trp. If RF1 has become specific for UAG in *Planaria* mt, this would make it possible for UAY Tyr codons to mutate to UAA and be read by anticodon GAU. GAU could pair with all 3 codons UAU, UAC and UAA if G in this anticodon had been modified to pair with U, C and A as in the case of GUU in echinoderms, which functions as an anticodon for AAU, AAC and AAA, lysine in these organisms.

UGA for Try and AUA for Met

The two codon changes that were first discovered in mitochondria, and therefore led to much interest, were UGA from stop to tryptophan and AUA from isoleucine to methionine[2]. The evolutionary histories of these two changes differ markedly.

UGA as a codon for tryptophan is found in all non-plant mt. This change was thoroughly explored in *Mycoplasma*[25]. Evidence was found that it resulted from AT directional mutation pressure that converted all UGA stop codons to UAA, accompanied by deletion of release factor 2, and changed anticodon CCA to UCA, pairing with both UGA and UGG. As a result, some Trp UGG codons that had mutated to UGA under AT pressure were translated as tryptophan. The same sequence of events probably occurred early in the evolution of non-plant mt. In fact, RF2, reacting with UAA and UGA, was not present in vertebrate mt, and RFI, reacting with UAA and UAG, was retained[16].

Codon AUA is translated by *E. coli* to isoleucine by a tRNA with anticodon *CUA, *C = lysidine, 5 lysyl-cytidine[21]. Removal of lysine from this position in this tRNA converts it to a methionyl tRNA. Loss of the enzyme system that converts C to *C would therefore change the assignment of codon AUA from isoleucine to methionine. It seems that this loss has taken place of at least twice in mt evolution (fig. 2): in yeast mt, and in metazoan mt after the separation of the platyhelminth lineage. AUA codes for isoleucine in echinoderms, but perhaps in this case AUA is read by isoleucine anticodon GAU.

This loss of lysidine would be a step of genomic economization. It could not take place, however, until AUA disappeared from coding sequences as a result of A to G mutations.

Changes in the non-plant mt code have taken place during evolution as a sequence of events, to some extent in chronological order, and to some extent occurring on branches that have separated in the evolutionary process, such as the branching between molds and yeasts (*Euascomycetes* and *Saccharomyces*). In addition to the changes discussed above, the following events have taken place.

AGR from Arg to Ser and from Ser to Stop

In metazoan mt, tRNA Arg (UCU) for codons AGR has been deleted by reduction of genome size during evolution. Replacement of AGR by mostly CGN Arg and by a few other codons in metazoan mt suggests that AGR Arg codons were converted mainly to CGN (arginine) upon deletion of tRNA Arg (UCU), so that AGR became unassigned.

Genomes from simpler metazoan mt are usually high in AT as are yeast mt. Probably in the metazoan mitochondrial ancestor, AGR at the important Arg sites mutated, even in the presence of AT pressure, to CGN Arg codons by selective constraints because conversion of AGR to CGN was the only way to conserve these Arg residues. However, the removal of tRNA Arg (UCU) would not have occurred by one-step deletion because predominant AGR codons would have become untranslatable, which would be deleterious. Probably tRNA Arg (UCU) gradually lost pairing ability with AGR codons. Finally, the gene for this tRNA disappeared. By this time, all AGR codons had disappeared from reading frames, by conversion mainly to CGN, and the tRNA for CGN codons, anticodon UCG, would have increased in amount adaptively. In this way, AGR became, without disruption, an unassigned codon pair available for subsequent capture by an amino acid[26].

Following this, the structure of tRNA Ser (GCU), pairing primarily with AGR codons, has an abbreviated and unusual structure, so that the altered Ser tRNA apparently translates AGR codons. AGR (mainly AGA) codons, pairing with this Ser tRNA then appeared by mutations of AGY Ser codons and other codons. AGR codons were thus captured by Ser.

The AGR codons captured by Ser are translated either by tRNA Ser (UCU) (in *Ascaris*) or by tRNA Ser (GCU) (in other invertebrates). It has been proposed that the unusual structure of tRNA Ser (GCU) brings about G:R pairing between anticodon GCU and codons AGR[26]. The anticodon UCU in *Ascaris* mt would have reappeared with a new assignment by mutation of anticodon GCU to UCU under AT pressure and seems to translate all AGN codons as Ser by four-way pairing.

The two 'new' Ser codons AGG and AGA would have been removed from messenger RNA of vertebrate mt before these became stop codons, because appearance of stop codons at Ser sites in reading frames results in cessation of translation, which is deleterious. AGG codons are not found in *Drosophila* mt[35], but we do not know whether AGG can be used as a stop codon in mt of other insects.

Shortly before, or in the early phase of vertebrate mitochondrial evolution, tRNA Ser (GCU) underwent further changes leading to abolition of pairing ability with codons AGR. All AGR (AGA) codons in *Drosophila* were replaced in vertebrates mainly by Ser AGY, Ser UCN, Lys AAA, Thr ACN, and others. Presumably, AGR codons were first removed by mutation mainly to AGY serine (as shown by comparing AGR sites in starfish mt and AGA sites in *Drosophila* with corresponding sites in vertebrates or to ACA by strong selective constraints resulting from the loss of AGR translation by anticodon GCU. Thus, AGR became unassigned codons, and four of the unassigned AGR sites reappeared in vertebrate mt as stop codons. AGY and ACA would further mutate to AGG, ACU, and ACC, and thence to UCN.

During vertebrate mitochondrial evolution, AGR stop codons presumably were formed from UAG stop by deletion of the first nucleotide U and by use of R as the third nucleotide that had existed next to the ancestral UAG stop in the tetranucleotide UAGR[26].

Neither of the changes AGR to serine or stop caused alterations in the amino acid sequences of mt proteins.

CUN from Leu to Thr in yeast mt

Leucine differs from threonine by two codon positions in the universal code, in which leucine has CUN codons and threonine has ACN. We have proposed that during a period of AT pressure, CUN codons disappeared by mutation to UUR[23]. UUR codons are translated by different anticodons, UAA. The ancestral leucine tRNA (UAG) that paired with CUN lost leucine-accepting ability, perhaps by mutation of sites involved in recognition of leucyl-tRNA synthase. The CUN family box of codons was reassigned to threonine when a duplicate threonyl tRNA synthase acquired the ability to charge tRNA (UAG) with threonine. As CUN codons subsequently appeared in reading frames from mutation of various codons, they were translated as threonine by the new tRNA Thr (UAG) that had evolved from tRNA Leu (UAG). Other threonine codons, ACN, are also used for threonine, so that threonine in yeast mt has 8 codons.

The usage of leucine and threonine codons in yeast and molds is such that codons ending in A and T are much preferred over those ending in G and C. This reflects high A + T content of DNA in *Ascomycotina*.

This change in the yeast mitochondrial genetic code is likely to have evolved through a series of nondisruptive nucleotide substitutions that produced no widespread replacement of leucine by threonine in proteins as a consequence.

The reassignment of CUN codons from Leu to Thr in yeast mt (fig. 2) would have been preceded by the silent disappearance of CUN Leu codons by AT pressure. The GC content of silent sites in codons of *Saccharomyces cerevisiae* mt averages less than 10%.

The four yeast (*Saccharomyces* and *Torulopsis*) and fungi (*Aspergillus* and *Neurospora*) mt genomes employ Thr ACA/U codons almost exclusively over the synonymous Thr ACG/C codons.

Similarly, UUR Leu codons (mostly UUA) much predominate over CUN Leu codons (mostly CUA/U) in mt of *Aspergillus* and *Neurospora*. In yeast mt the number of Thr CUN codons is small relative to UUR Leu. We assume that an ancestor of yeast mt used Leu CUN codons only rarely, and these may have disappeared through silent mutations to Leu UUR. Indeed, most Leu CUN sites in fungal mt correspond to Leu UUR in yeast mt.

Almost all yeast mt CUN sites are occupied in fungal mt by Thr ACN, Val GUN, Ser UCN, and other codons. Furthermore, most Leu CUN sites in fungal mt correspond to Leu UUR in yeast mt, in accordance with the suggestion that yeast mt Leu CUN codons disappeared prior to the code change. Some of the existing CUN Thr codons in yeast mt could have been derived from ACN Thr. However, ACN cannot be converted to CUN by a single mutation. Presumably, some of the ACN codons first mutated to AUN (Ile/Met) or UUR (Leu) via UCN and thence to CUR (Thr). Other CUN threonine codons in yeast mt were derived from various codons for other amino acids. Indispensable ACN Thr sites would not have been involved in this reassignment. We infer that most or all existing yeast mt Thr CUN codons arose subsequently to the code change of CUN from Leu to Thr by individual mutations of various other codons rather than through a large-scale replacement of Leu by Thr[23].

AT pressure may affect the composition of tRNA genes in addition to influencing the frequency of their corresponding codons. *Saccharomyces cerevisiae* mt tRNA genes are relatively AT-rich, averaging 30% GC, and tRNA Thr (UAG), like tRNA Arg (ACG) is high in AT. Unusual base composition may indicate that a relaxation of selective constraints has occurred. If a codon is retired from use, the corresponding tRNA gene sequence will no longer be maintained by selective forces and may be lost from the genome, as in the case of *T. glabrata* mt tRNA Arg, or may become free to acquire a new function as in the case of yeast mt tRNA (UAG). In this manner, yeast mt tRNA (UAG) may have lost its ability to interact with leucyl tRNA synthetase, and later acquired the ability to interact with a threonyl synthetase[23].

The CUN codons were reassigned to Thr when tRNA (UAG) acquired the ability to interact with threonyl-

tRNA synthetase to accept Thr. As CUN codons subsequently appeared in reading frames from mutations of various codons, they were 'captured' and translated as Thr by the 'new' tRNA Thr (UAG) that evolved from tRNA Leu (UAG).

Loss of translation by CGN codons in Torulopsis mt

In the yeast, *Torulopsis glabrata*, neither Arg CGN codons nor a tRNA Arg able to decode them have been found[4]. All Arg codons are AGR. In *S. cerevisiae* mt, a close relative of *T. glabrata*, mt genes are AT-rich, averaging 33% GC. However, tRNA Arg (ACG), corresponding to anticodon UCG that translates all CGN codons in other mt, is high in AT (18% GC). This anticodon, ACG, translates rare CGN Arg codons. In *S. cerevisiae* mt, Arg is coded almost entirely by AGR; most Arg CGN codons appear to have mutated silently to Arg AGR. It is possible that when AT pressure causes a codon to fall into disuse, the functional constraints on the corresponding tRNA gene are thereby reduced, allowing it to accumulate mutations that would be otherwise deleterious. Because of AT pressure, most of these mutations are towards AT with the result that the GC content of the tRNA gene decreases. Thus, unusual base composition in the *S. cerevisiae* tRNA Arg (ACG) may indicate that a relaxation of selective constraints has occurred. If a codon is retired from use entirely, the corresponding tRNA gene will no longer be maintained by selective forces and may be lost from the genome as in the case of *T. glabrata* mt tRNA Arg, so that CGN codons become unassigned.

It is thus highly probable that a strong directional mutation pressure would produce unassigned codons. Mitochondria and chloroplasts do not charge tRNA Gln by means of glutamine-tRNA synthase. This enzyme has not been detected in these organelles. Instead, tRNA Gln is charged with Gln [30a].

Why are there changes in the code in non-plant mt?

The nuclear code uses UGA for stop and AUA for isoleucine, and mammalian mt use UGA for tryptophan and AUA for methionine. The significance of this difference has provoked much discussion. Wallace [33] proposed that the differences may have become established because they minimized the deleterious effects of accidental transfer of mRNAs between the mitochondria and the cytosol. His proposal is not supported by the observation that plant mt use the universal code. Also, *Mycoplasma* use UGA for Trp, and they are not subject to accidental transfer of mRNAs. It appears more likely that the changes have evolutionary histories as described in this review.

Acknowledgments. We acknowledge the advice of David Collins, the collaboration of Roxanne Short and Carol Fegté in preparation of the manuscript and support by NIH grant GM 40903 01 to the University of California, Berkeley and by a grant-in-aid from the Ministry of Education, Japan.

1 Andachi, Y., Yamao, F., Muto, A., and Osawa, S., Codon recognition patterns as deduced from sequences of the complete set of transfer RNA species in *Mycoplasma capricolum*: Resemblance to mitochondria. J. molec. Biol. *209* (1989) 37–54.
2 Barrell, G., Bankier, A. T., and Drouin, J., A different genetic code in human mitochondria. Nature *282* (1979) 189–194.
3 Cantatore, P., Roberti, M., Morisco, R., Rainaldi, G., Gadaleta, M. N., and Saccone, C., A novel gene order in the *Paracentrotus lividus* mitochondrial genome. Gene *53* (1987) 41–54.
4 Clark-Walker, G. D., McArthur, C. R., and Sriprakash, K., Location of transcriptional control signals and transfer RNA sequence in *Torulopsis glabrata* mitochondrial DNA. EMBO J. *4* (1985) 465–473.
5 Gray, M. W., and Boer, P. H., Organization and expression of algal (*Chlamydomonas reinhardtii*) mitochondrial DNA. Phil. Trans. R. Soc. Lond. B *319* (1988) 135–147.
6 Gray, M. W., and Doolittle, W. F., Has the endosymbiont hypothesis been proven? Microbiol. Rev. *46* (1982) 1–42.
7 Gray, M. W., Sankoff, D., and Cedergren, R. J., On the evolutionary descent of organisms and organelles: A global phylogeny based on a highly conserved structural core in small subunit ribosomal RNA. Nucl. Acids Res. *12* (1984) 5837–5852.
8 Gray, M. W., Cedergren, R., Abel, Y., and Sankoff, D., On the evolutionary origin of the plant mitochondrion and its genome. Proc. natl Acad. Sci. USA *86* (1989) 2267–2271.
9 Green, A. G., Maréchal, L., Weil, J. H., and Guillemaut, P., A *Phaseolus vulgaris* mitochondrial $tDNA^{Leu}$ is identical to its cytoplasmic counterpart. Plant molec. Biol. *10* (1987) 13–19.
10 Himeno, H., Masaki, H., Ohta, T., Kumagai, I., Miura, K.-I., and Watanabe, K., Unusual genetic codes and a novel genome structure for $tRNA^{Ser}$ AGY in starfish mitochondrial DNA. Gene *56* (1987) 219–230.
11 Izuchi, S.-I., and Sugita, M., Nucleotide sequence of a tomato mitochondrial $tRNA^{Cys}$ (GCA) gene. Nucleic Acids Res. *17* (1990) 1248.
12 Jacobs, H. T., Elliott, D. J., Math, V. B., and Farquharson, A., Nucleotide sequence and gene organization of sea-urchin mitochondrial DNA. J. molec. Biol. *202* (1988) 185–217.
13 Joyce, P. B., and Gray, M. W., Chloroplast-like transfer RNA genes expressed in wheat mitochondria. Nucleic Acids Res. *17* (1989) 5461–5476.
14 Köchel, H. G., Lazarus, C. M., Basak, N., and Küntzel, H., Mitochondrial tRNA gene clusters in *Aspergillus nidulans*: organizations and nucleotide sequence. Cell *23* (1981) 625–633.
15 Küntzel, H., and Köchel, H. G., Evolution of rRNA and origin of mitochondria. Nature *293* (1981) 751–755.
16 Lee, C. C., Timms, K. M., Trotman, C. N. A., and Tate, W. P., Isolation of a rat mitochondrial release factor: Accommodation of the changed genetic code for termination. J. biol. Chem. *262* (1987) 3548–3552.
17 Maréchal-Drouard, L., and Guillemaut, P., Nucleotide sequence of bean mitochondrial $tRNA^{Leu}4$ and of its cytoplasmic counterpart. Re-examination of the modified nucleotide present at position 12 in bean mitochondrial and cytoplasmic $tRNA^{Leu}1$ sequences. Nucleic Acids Res. *16* (1988) 11 812.
18 Maréchal-Drouard, L., Weil, J. H., and Guillemaut, P., Import of several tRNAs from the cytoplasm into the mitochondria in bean *Phaseolus vulgaris*. Nucleic Acids Res. *16* (1988) 4777–4788.
19 Maréchal-Drouard, L., Guillemaut, P., Cosset, A., Arbogast, M., Weber, F., Weil, J., and Dietrich, A., Transfer RNAs of potato (*Solanum tuberosum*) mitochondria have different genetic origins. Nucleic Acids Res. *18* (1990) 3689–3696.
20 McCarroll, R., Olsen, G. J., Stahl, Y. D., Woese, C. R., and Sogin, M. L., Nucleotide sequence of the *Dictyostelium discoideum* small-subunit ribosomal ribonucleic acid inferred from the gene sequence: Evolutionary implications. Biochemistry *22* (1983) 5858–5868.
21 Muramatsu, T., Nishikawa, K., Nemoto, F., Kuchino, Y., Nishimura, A., Miyazawa, T., and Yokoyama, S., Codon and amino acid specificities of a transfer RNA are both converted by a single post-transcriptional modification. Nature *336* (1988) 179–181.
22 Ohama, T., Osawa, S., Watanabe, K., and Jukes, T. H., Evolution of the mitochondrial genetic code IV. AAA as an asparagine codon in some animal mitochondria. J. molec. Evol. *30* (1990) 329–332.
23 Osawa, S., Collins, D., Ohama, T., Jukes, T. H., and Watanabe, K., Evolution of the mitochondrial genetic code III. Reassignment of CUN codons from leucine to threonine during evolution of yeast mitochondria. J. molec. Evol. *30* (1990) 322–328.
24 Osawa, S., and Jukes, T. H., Evolution of the genetic code by anticodon content. Trends Genet. *4* (1988) 191–198.

25 Osawa, S., and Jukes, T. H., Codon reassignment (codon capture) in evolution. J. molec. Evol. *28* (1989) 271–278.
26 Osawa, S., Ohama, T., Jukes, T. H., and Watanabe, K., Evolution of the mitochondrial genetic code I. Origin of AGR serine and stop codons in metazoan mitochondria. J. molec. Evol. *29* (1989) 202–207.
27 Osawa, S., Ohama, T., Jukes, T. H., Watanabe, K., and Yokoyama, S., Evolution of the mitochondrial genetic code II. Reassignment of codon AUA from isoleucine to methionine. J. molec. Evol. *29* (1989) 373–380.
28 Osawa, S., Muto, A., Jukes, T. H., and Ohama, T., Evolutionary changes in the genetic code. Proc. Roy. Soc. Lond. B *241* (1990) 19–28.
29 Ozeki, H., Ohyama, K., Inokuchi, H., Fukuzawa, H., Kiochi, T., Sano, T., Nakahigashi, K., and Umesono, K., Genetic system of chloroplasts. Cold Spring Harbor Symp. quant. Biol. *52* (1987) 791–804.
30 Parks, T. D., Dougherty, W. G., Levings, C. S. III, and Timothy, D. H., Identification of two methoinine transfer RNA genes in the maize mitochondrial genome. Plant Physiol. *76* (1984) 1079–1082.
30a Schön, A., Kannangara, C. G., Gough, S., and Söll, D., Protein biosynthesis in organelles requires misaminoacylation of tRNA. Nature *331* (1988) 187ff.
31 Seilhamer, J. J., and Cummings, D. J., Altered genetic code in *Paramecium* mitochondria: Possible evolutionary trends. Med. gen. Genet. *187* (1982) 236–239.
32 Suyama, Y., Two-dimensional polyacrylamide gel electrophoresis analysis of *Tetrahymena* mitochondrial tRNA. Curr. Genet. *10* (1986) 411–420.
33 Wallace, D. C. W., Structure and evolution of organelle genomes. Microbiol. Rev. *46* (1982) 208–240.
34 Weber, F., Dietrich, A., Weil, J.-H., and Maréchal-Drouard, L., A potato mitochondrial isoleucine tRNA is coded for by a mitochondrial gene possessing a methionine anticodon. Nucl. Acids Res. *18* (1990) 5027–5030.
35 Wolstenholme, D. R., Okimoto, R., Macfarlane, J. L., Pont, G. A., Chamberlin, H. M., Garey, J. R., and Okada, N. A., Unusual features of lower invertebrate mitochondrial genomes, in: Structure, Function and Biogenesis of Energy Transfer Systems. Eds E. Quagriello, S. Papa, F. Palmieri and C. Saccone. Elsevier, Amsterdam 1990.
36 Wakasugi, T., Ohme, M., Shinozaki, K., and Sugiura, M., 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 molec. Biol. *7* (1986) 385–392.
37 Yang, D., Oyaizu, Y., Oyaizu, H., Olsen, G. J., and Woese, C. R., Mitochondrial origins. Proc. natl Acad. Sci. USA *82* (1985) 4443–4447.

0014-4754/90/11-12/1117-10$1.50 + 0.20/0

Codon context

R. H. Buckingham

URA 1139 du CNRS, Institut de Biologie Physico-Chimique, 13 rue Pierre et Marie Curie, Paris (France)

Summary. The analysis of coding sequences reveals nonrandomness in the context of both sense and stop codons. Part of this is related to nucleotide doublet preference, seen also in non-coding sequences and thought to arise from the dependence of mutational events on surrounding sequence. Another nonrandom context element, relating the wobble nucleotides of successive codons, is observed even when doublet preference, codon usage and bias in amino acid doublets are all allowed for. Several phenomena related to protein synthesis have been shown in vivo to be affected by the nucleotide sequence around codons. Thus, nonsense and missense suppression, elongation rate, precision of tRNA selection and polypeptide chain termination are all affected by codon context. At present, it remains unclear how these phenomena may influence the evolution of nonrandomness in the context of codons in natural sequences.
Key words. Context effects; nucleotide sequence nonrandomness; translational suppression; polypeptide chain elongation; polypeptide chain termination.

Introduction

Effects of neighbouring sequences on codon translation, usually referred to as context effects, have frequently been evoked to explain puzzling phenomena even when direct evidence of their responsibility has been weak. As will be seen below, two quite different approaches have yielded information about the significance of codon context in translation. Experimental evidence from studies both in vivo and in vitro shows that context can affect nonsense suppression, missense suppression, translational errors and frameshifting. Secondly, statistical analysis of coding sequences reveals that the context around codons is not random, and that the tendencies towards nonrandom contexts are different according to the level of expression of the genes concerned. Part of this nonrandomness has to do with mutational pressure, and shows similarities to nonrandomness in non-coding sequences. Other elements may be due to selective pressure related to translation.

Evidence from statistical analysis for constraints on codon context

Several major studies of natural coding sequences have been performed and point to significant tendencies towards nonrandom codon context. A detailed discussion of codon preferences and codon context preferences needs to be prefaced by a consideration of factors that influence the evolution of non-coding sequences, since