J Mol Evol (1992) 34:331–335

Journal of Molecular Evolution

Planarian Mitochondria II. The Unique Genetic Code as Deduced from Cytochrome c Oxidase Subunit I Gene Sequences

Yoshitaka Bessho, Takeshi Ohama, and Syozo Osawa

Laboratory of Molecular Genetics, Department of Biology, School of Science, Nagoya University, Chikusa-ku, Nagoya 464-01, Japan

Summary. The cytochrome c oxidase subunit I (COI) gene sequences from planarian (*Dugesia japonica*) DNA, most probably of mitochondrial origin, are heterogeneous. Taking advantage of the heterogeneity that occurs primarily in silent sites of the COI DNA sequences, amino acid assignments of several codons have been deduced as nonuniversal: UGA = Trp, AAA = Asp, and AGR (R: A or G) = Ser. In addition, UAA, a stop codon in the universal genetic code, is tentatively assumed to be a tyrosine codon, because three of the sequences examined have UAA at the well-conserved tyrosine site of UAY (Y: U or C) in other planarian sequences as well as in the mitochondria of human, *Xenopus*, sea urchin, *Drosophila*, *Trypanosoma*, and *Saccharomyces cerevisiae*. AUA would most probably be an isoleucine codon in these mitochondria, whereas it is a methionine codon in the majority of nonplant mitochondria.

Key words: Genetic code — Mitochondria — Planarian — *Dugesia japonica* — Cytochrome c oxidase subunit I gene — UAA tyrosine codon

Introduction

Heterogeneous sequences of the cytochrome c oxidase subunit I (COI) gene in planarian mitochondrial DNA (mtDNA) should be almost equivalent functionally, as heterogeneity occurs mostly at silent codon sites in this gene (Bessho et al. 1992). Furthermore, the COI gene is one of the most well-conserved proteins, so that reliable alignment within various planarian DNA sequences and even with those of various other organisms is easily obtained. These situations make it possible to deduce amino acid assignments for codons in planarian mitochondria.

Offprint requests to: Y. Bessho

Materials and Methods

The Materials and Methods are discussed in detail in Bessho et al. (1992). In addition, two primers, pr-c (5′-ACATATGATGGGCCCAAAC-3′) and pr-d (5′-TTTTCTACAAATCATAAAGATATTGG-3′), were constructed to amplify the region upstream of Seq-A as shown in Fig. 2 of Bessho et al. (1992). The polymerase chain reaction (PCR) product obtained with the template DNA from a group of mixed individuals was analyzed by the direct sequencing method. The sequence obtained was the same as for type A in the overlapping region, so that we assumed this sequence to be the upstream origin of type A.

Results and Discussion

UGA

UGA was detected four times in the planarian COI gene, i.e., at positions 186, 288, 323, and 340. These positions are highly conserved tryptophan sites (UGA or UGG) throughout various nonplant mitochondria. Thus, in addition to codon UGG, UGA most probably codes for tryptophan in planarian mitochondria.

UAA

UAA is a stop codon in all of the genetic codes that have been reported. In the reading frames of the planarian COI gene sequences, UAA was detected at position 304 in type E (Seq-E1, Seq-E2, and Seq-

Table 1. Comparison of six codon sites in various cytochrome c oxidase subunit I (COI) gene sequences of planarian mitochondria with corresponding amino acid sites for other mitochondria

	UGR						UAN	
	186	193	236	288	323	340		244
Seq-A	TGA	TGG	TGG	TGG	TGA	TGG	Seq-A	TAT
Seq-B				TGG	TGA	TGG	Seq-B	
Seq-C				TGA	TGG	TGA	Seq-C	
Seq-D1				TGG	TGG		Seq-D1	
Seq-E1				TGG	TGA	TGG	Seq-E1	
Seq-E2				TGG	TGA	TGG	Seq-E2	
Seq-E3				TGG	TGA	TGG	Seq-E3	
Human	W	V	W	W	W	W	Human	Y
Xenopus	W	V	W	W	W	W	*Xenopus*	Y
Sea	W	F	W	W	W	W	Sea	Y
Dry	W	L	W	W	W	W	Dry	Y
Try	W	I	W	W	W	F	Try	Y
Yeast	W	F	W	W	W	Y	Yeast	Y

	AUN										
	190	208	210	234	245	247	248	253	259	271	273
Seq-A	ATA	ATG	ATT	ATG	ATT	ATA	ATT	ATT	ATG	ATA	ATG
Seq-B					ATT	ATA	ATT	ATA	---	ATT	ATG
Seq-C					ATT	ATT	ATT	ATA	ATG	TTG	ATG
Seq-D1								ATT	ATG	ATT	ATG
Seq-E1					ATT	ATT	ATT	ATA	ATG	CTT	ATG
Seq-E2					ATT	ATT	ATT	ATA	ATG	CTT	ATG
Seq-E3					ATT	ATT	ATT	ATA	ATG	CTT	ATG
Human	I	M	L	L	I	I	L	M	T	M	M
Xenopus	I	M	L	L	I	I	L	M	T	M	M
Sea	V	M	L	L	I	I	L	M	A	L	L
Dry	I	M	L	L	I	I	L	M	S	L	M
Try	L	L	L	L	I	I	L	L	E	V	M
Yeast	I	M	L	L	I	I	I	I	S	I	M

	AGN										
	218	223	262	279	322	331	332a	333	335	362	366
Seq-A	AGT	AGG	AGT	AGC	AGT	AGT	TTA	AGT	AGT	AGG	AGT
Seq-B			AGT	AGC	AGT	AGT	TTA	AGT	AGT	AGG	AGT
Seq-C			AGT	AGT	AGT	AGC	AGA	AAT	AGA	AGG	AGT
Seq-D1			AGT	AGT	AGT						
Seq-E1			AGT	AGT	AGT	AGT	GTT	AAT	AGT		
Seq-E2			AGT	AGT	AGT	AGT	GTT	AAT	AGT		
Seq-E3			AGT	AGT	AGT	AGT	GTT	AAT	AGT		
Human	T	A	S	S	S	N	–	K	S	S	V
Xenopus	T	A	S	S	S	T	–	K	D	S	M
Sea	T	A	S	A	S	N	–	Q	S	S	V
Dry	S	A	S	A	S	Q	–	S	S	S	I
Try	S	V	S	L	N	D	–	C	C	G	L
Yeast	S	A	S	S	S	S	–	R	A	S	A

Seq-A to Seq-E3 are representative sequences of a part of the heterogeneous COI gene from planarian mtDNA. The amino acid residue numbers follow those given for human COI (Anderson et al. 1981). 332a is an extra amino acid site in planarian COI; a blank indicates that no data were obtained at that site; – indicates a gap. TGA, TAA, ATA, AGR, and AAA are boxed. Source of amino

E3), whereas this position is UAU tyrosine in all other planarian sequences. Moreover, this site corresponds to a highly conserved tyrosine site in the mitochondria of humans (Anderson et al. 1981), *Xenopus* (Roe et al. 1985), sea urchin (Jacobs et al. 1988), *Drosophila* (Clary and Wolstenholme 1985), *Trypanosoma* (Hensgens et al. 1984), and yeast (Bonitz et al. 1980) (Table 1). Therefore, UAA probably codes for tyrosine rather than for a stop codon, at least in type-E sequences, although more evidence is needed. It may be argued that type E sequences could be a part of pseudogenes, or that UAA acts as a stop codon producing a shorter peptide. Either possibility is not very likely, as the deduced amino

Table 1. Extended

UAN									
260	261	265	293	299	304	328	330	371	372
TAT	TAT	GAT	TAC	TAC	TAT	TGT	TCT	TAC	TAT
TAC	TAT	GAT	TAC	TAC	TAT	TGT	TCT	TAC	TAT
TAC	TAT	GAC	TAT	TAT	TAT	TGT	TCT	TAT	TAT
TAT	TAT	GAT	TAT	TAT	TAT	TTT			
TAC	TAT	TAT	TAT	TAT	TAA	TAT	TAT		
TAC	TAT	TAT	TAT	TAT	TAA	TAT	TAT		
TAC	TAT	TAT	TAT	TAT	TAA	TAT	TAT		
Y	Y	K	F	V	Y	H	S	Y	Y
Y	Y	K	F	V	Y	H	G	Y	Y
H	Y	R	F	V	Y	Q	S	Y	Y
Q	E	K	F	V	Y	H	T	Y	Y
V	T	–	F	V	Y	L	T	Y	F
T	Y	–	F	A	Y	Y	G	Y	Y

AUN											
277	280	286	292	310	311	312	318	324	345	350	365
ATG	ATT	ATT	ATG	ATG	ATT	ATT	ATT	ATT	ATT	ATA	ATT
ATG	ATT	ATT	ATG	ATG	ATT	ATT	ATT	ATT	ATT	ATA	ATT
ATG	ATT	ATT	ATG	ATG	ATT	ATT	ATT	ATA	ATT	GTT	ATA
ATG	ATT	ATT	ATG	ATG	ATT	ATT	ATT	ATT			
ATG	ATT	ATT	ATG	ATG	ATT	ATT	ATT	ATT	ATC		
ATG	ATT	ATT	ATG	ATG	ATT	ATT	ATT	ATT	ATC		
ATG	ATT	ATT	ATG	ATG	ATT	ATT	ATT	ATT	ATC		
M	I	I	M	M	I	I	V	L	I	V	I
M	I	I	M	M	I	I	V	L	I	V	I
M	I	L	M	M	I	I	L	M	V	L	F
M	I	I	M	M	I	I	I	L	V	V	I
M	I	F	M	V	L	I	I	I	I	A	I
M	I	L	M	M	I	I	I	L	L	M	V

CGN				AAN							
	213	302	334		214	216	264	319	333	339	360
Seq-A	CGT	CGT	CGT	Seq-A	AAT	AAA	AAG	AAG	AGT	AAT	AAT
Seq-B		CGT	CGT	Seq-B			AAG	AAG	AGT	AAT	AAT
Seq-C		CGT	CGT	Seq-C			AAG	AAG	AAT	AAT	AAT
Seq-D1		CGT		Seq-D1			AAG	AAG			
Seq-E1		CGT	CGT	Seq-E1			AAG	AAG	AAT	AAT	
Seq-E2		CGT	CGT	Seq-E2			AAG	AAG	AAT	AAT	
Seq-E3		CGT	CGT	Seq-E3			AAG	AAG	AAT	AAT	
Human	R	R	W	Human	N	N	K	K	K	L	N
Xenopus	R	R	W	*Xenopus*	N	N	K	K	K	L	N
Sea	R	R	W	Sea	N	N	K	K	Q	L	N
Dry	R	R	Y	Dry	N	N	K	K	S	L	N
Try	R	R	I	Try	N	N	R	K	C	Y	N
Yeast	R	R	L	Yeast	N	N	K	K	R	L	N

acid sequences: Human (Anderson et al. 1981), *Xenopus laevis* (Roe et al. 1985), sea urchin (*Strongylocentrotus purpuratus*: Sea) (Jacobs et al. 1988), *Drosophila yakuba* (Dry) (Clary and Wolstenholme 1985), *Trypanosoma brucei* (Try) (Hensgens et al. 1984), and *Saccharomyces cerevisiae* (Yeast) (Bonitz et al. 1980)

acid sequences are well conserved before and after UAA, if UAA is assumed to be a tyrosine codon. It then follows that planarian mitochondria might have only one stop codon, UAG, because UGA is a tryptophan codon.

Dugesia japonica and *Fasciola hepatica* belong to the same phylum Platyhelminthes. In the three reported mitochondrial genes of *Fasciola* (ND1, ND3, and COI) (Garey and Wolstenholme 1989), all of the stop codons are UAG, which is the most rarely used stop codon throughout mitochondrial genes in other species. It is possible that UAA, although not yet found in *Fasciola,* is a tyrosine codon as is suspected in planarian mitochondria.

AUA

AUA is a methionine codon in all nonplant mitochondria except for echinoderms. Presumably, the conversion of AUA isoleucine to methionine took place during the early stages of mitochondrial evolution. In the echinoderm line, a reversal of AUA from methionine to isoleucine could have taken place after its separation from the main metazoan line of descent (Jukes and Osawa 1990). In planarian COI sequences, AUA is used at positions 190 (type A), 247 (types A and B), and 365 (type C). Positions 247 and 365 in all planarian mitochondrial sequences are occupied either by AUA or AUU (Ile). These three positions are well conserved as isoleucine in other organisms. AUA was also detected at positions 253, 271, 324, and 350, and these positions are not well conserved in other organisms. Among planarian sequences, positions 253 and 324 are either AUU or AUA, and position 271 is AUU, AUA, or other codons. There are no AUA sites replacing AUG. From these facts, AUA would code for isoleucine and not for methionine. The same reversal of AUA from methionine to isoleucine as in echinoderms could have occurred independently in the planarian line. The assignment of AUA in *Fasciola* has not been settled (Garey and Wolstenholme 1989).

AGR

AGG would code for serine and not for arginine, as position 362 is AGG in all of the planarian sequences, and the corresponding amino acid of this site is serine in human, *Xenopus, Drosophila,* sea urchins, and yeast. Furthermore, well-conserved arginine sites at positions 213 and 302 in other mitochondria are all CGU in planarian mitochondria, and no AGG codons are used. AGA would also code for serine, as position 335 of type C is AGA, whereas it is AGU (serine) in other planarian sequences. Codons AGA and AGG are also assigned to serine in other animal species except vertebrates in which these are stop codons (see Jukes and Osawa 1990).

AAA

AAA occurred only once in type A, but it may be for asparagine and not for lysine. AAA at position 216 in planarian mitochondria corresponds to an asparagine site in mitochondria of all of the organisms throughout. In addition, AAA is not found at positions 264 and 319 in other planarian sequences, where these sites are occupied only by AAG lysine. In other organisms, these sites are occupied also by a lysine codon of either AAA (except echinoderms; see below) or AAG. AAA was proposed to be an asparagine codon in the mitochondria of echinoderms (Himeno et al. 1987; Jacobs et al. 1988; Cantatore et al. 1989) and *Fasciola* (Ohama et al. 1990).

The deduced amino acid assignments, AUA for isoleucine, UAA for tyrosine, and AAA for asparagine, could have resulted from similar mechanisms. Codon NNA (N: U, C, A, or G) might have disappeared from the coding frames concomitant with a loss of the coding function of the corresponding tRNA (for AAA or AUA) or a loss of the UAA-recognition function of release factor I (RF-I) as a result of mutations accumulating in the tRNA or RF-I. Then, the NNA codon might have been captured later by an amino acid that had previously been assigned only by NNY a two-codon set (Y: U or C). This might have occurred following a change in the coding capacity of the tRNA for NNY codons, so that the NNA codon would be translated in addition to NNY. Usually, the anticodon GNN recognizes NNU and NNC codons by wobble. If the anticodon first nucleoside G is converted, for example, to I (inosine), such an anticodon could wobble-pair with U, C, and A, as proposed for the translation of AAA as asparagine in echinoderm mitochondria (Himeno et al. 1987; Jacobs et al. 1988; Cantatore et al. 1989; see also Jukes and Osawa 1990). As a result, such a codon capture would have produced a three-codon set, consisting of codons NNU, NNC, and NNA, all of which are now translated to the same amino acid, i.e., AUY/A to isoleucine, AAY/A to asparagine, and UAY/A to tyrosine.

In the present paper, the assignment of the "changed" codon is based on comparisons of the amino acid sequences as predicted from nucleotide sequences of the DNA of other species. This procedure has a number of potential pitfalls. One of these would be that observed changes are due to RNA editing (for review, see Cattaneo 1990) and not to changes in the genetic code. However, RNA editing has not been reported to occur in the mitochondria of multicellular animals.

Acknowledgments. This work was supported by Research Aid from the Inoue Foundation for Science, Inamori Science Foundation, and grants from the Ministry of Education, Science and Culture, Japan.

References

Anderson S, Bankier AT, Barrell BG, de Bruijn MHL, Coulson AR, Drouin J, Eperon IC, Nierlich DP, Roe BA, Sanger F, Schreier PH, Smith ALH, Staden R, Young IG (1981) Sequence and organization of the human mitochondrial genome. Nature 290:457–464

Bessho Y, Ohama T, Osawa S (1992) Planarian mitochondria I. Heterogeneity of cytochrome c oxidase subunit I gene sequences in the freshwater planarian, *Dugesia japonica.* J Mol Evol 34:000–000

Bonitz SG, Coruzzi G, Thalenfeld BE, Tzagoloff A, Macino G (1980) Assembly of the mitochondrial membrane system:

structure and nucleotide sequence of the gene coding for subunit I of yeast cytochrome oxidase. J Biol Chem 255:11927–11941

Cantatore P, Roberti M, Rainaldi G, Gadaleta MN, Saccone C (1989) The complete nucleotide sequence, gene organization, and genetic code of the mitochondrial genome of *Paracentrotus lividus*. J Biol Chem 264:10965–10975

Cattaneo R (1990) Messenger RNA editing and the genetic code. Experientia 46:1142–1148

Clary DO, Wolstenholme DR (1985) The mitochondrial DNA molecule of *Drosophila yakuba*: nucleotide sequence, gene organization, and genetic code. J Mol Evol 22:252–271

Garey JR, Wolstenholme DR (1989) Platyhelminth mitochondrial DNA: evidence for early evolutionary origin of $tRNA^{ser}$ AGN that contains a dihydrouridine arm replacement loop, and serine-specifying AGA and AGG codons. J Mol Evol 28: 374–387

Hensgens LAM, Brakenhoff J, de Vries BF, Sloof P, Tromp MC, van Boom JH, Benne R (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

Himeno H, Masaki H, Kawai T, Ohta T, Kumagai I, Miura I, Watanabe K (1987) Unusual genetic codes and a novel gene structure for $tRNA_{AGY}^{Ser}$ in starfish mitochondrial DNA. Gene 56:219–230

Jacobs HT, Elliot DJ, Math VB, Farquharson A (1988) Nucleotide sequence and gene organization of sea urchin mitochondrial DNA. J Mol Biol 202:185–217

Jukes TH, Osawa S (1990) The genetic code in mitochondria and chloroplasts. Experientia 46:1117–1126

Ohama T, Osawa S, Watanabe K, Jukes TH (1990) Evolution of the mitochondrial genetic code IV: AAA as an asparagine codon in some animal mitochondria. J Mol Evol 30:329–332

Roe BA, Ma DP, Wilson RK, Wong JFH (1985) The complete nucleotide sequence of the *Xenopus laevis* mitochondrial genome. J Biol Chem 260:9759–9774

Received March 27, 1991/Revised November 25, 1991