Complete DNA Sequence of the Mitochondrial Genome of the Ascidian *Halocynthia roretzi* (Chordata, Urochordata)

Shin-ichi Yokobori,*,† Takuya Ueda,‡ Gertraud Feldmaier-Fuchs,† Svante Pääbo,†,1 Rei Ueshima,§ Akiko Kondow, Kazuya Nishikawa†† and Kimitsuna Watanabe****

**Department of Molecular Biology, School of Life Science, Tokyo University of Pharmacy and Life Science, Horinouchi, Hachioji, Tokyo 192-0392, Japan, †Institute of Zoology, University of Munich, Munich, D-80333 Germany, ‡Department of Integrated Biosciences, Graduate School of Frontier Sciences, University of Tokyo, Hongo, Bunkyo-ku, Tokyo 113-8656, Japan, §Division of Evolutionary Biology, Department of Biological Sciences, Graduate School of Science, University of Tokyo, Hongo, Bunkyo-ku, Tokyo 113-8656, Japan, **Department of Chemistry and Biotechnology, Graduate School of Engineering, University of Tokyo, Hongo, Bunkyo-ku, Tokyo 113-8656, Japan and ††Department of Biomolecular Science, Faculty of Engineering, Gifu University, Yanagido, Gifu, 501-1193 Japan*

Manuscript received April 5, 1999
Accepted for publication August 10, 1999

ABSTRACT

The complete nucleotide sequence of the 14,771-bp-long mitochondrial (mt) DNA of a urochordate (Chordata)—the ascidian *Halocynthia roretzi*—was determined. All the Halocynthia mt-genes were found to be located on a single strand, which is rich in T and G rather than in A and C. Like nematode and *Mytilus edulis* mtDNAs, that of Halocynthia encodes no ATP synthetase subunit 8 gene. However, it does encode an additional tRNA gene for glycine (anticodon TCT) that enables Halocynthia mitochondria to use AGA and AGG codons for glycine. The mtDNA carries an unusual $tRNA^{Met}$ gene with a TAT anticodon instead of the usual $tRNA^{Met}_{CAT}$ gene. As in other metazoan mtDNAs, there is not any long noncoding region. The gene order of Halocynthia mtDNA is completely different from that of vertebrate mtDNAs except for $tRNA^{His}$–$tRNA^{Ser}_{GCU}$, suggesting that evolutionary change in the mt-gene structure is much accelerated in the urochordate line compared with that in vertebrates. The amino acid sequences of Halocynthia mt-proteins deduced from their gene sequences are quite different from those in other metazoans, indicating that the substitution rate in Halocynthia mt-protein genes is also accelerated.

ALTHOUGH the complete sequences of some 60 metazoan mitochondrial (mt) DNAs have been reported, the species completely sequenced belong to only 7 out of over 30 existing phyla. Metazoan mtDNAs are typically circular, range in size from 14 to 18 kbp, and encode 13 protein genes [ATP synthetase subunits 6 and 8 (*ATPase6* and *8*), cytochrome oxidase subunits I–III (*COI–COIII*), apocytochrome b (*Cytb*), and NADH dehydrogenase subunits 1–6 and 4L (*ND1–6* and *4L*)], 2 ribosomal RNA genes [small and large subunit rRNAs (*srRNA* and *lrRNA*)], 22 tRNA genes, and no introns (Wolstenholme 1992). Exceptions have been reported in a few metazoan lineages: certain cnidarian mitochondria have linear genomes (Warrior and Gall 1985); the *ATPase8* is lost from mtDNAs of nematodes (Okimoto *et al.* 1991, 1992) and the bivalve *Mytilus edulis* (Hoffmann *et al.* 1992); Mytilus mtDNA carries an additional $tRNA^{Met}$ (Hoffmann *et al.* 1992); and the mtDNA of the cnidarian *Metridium senile* encodes only two tRNA genes but has an additional open reading frame (Wolstenholme 1992; Beagley *et al.* 1998).

Corresponding author: Shin-ichi Yokobori, Department of Molecular Biology, School of Life Science, Tokyo University of Pharmacy and Life Science, 1432-1 Horinouchi, Hachioji, Tokyo 192-0392, Japan. E-mail: yokobori@ls.toyaku.ac.jp

[1]*Present address:* Max-Planck-Institute for Evolutionary Anthropology, Leipzig, D-04103 Germany.

The arrangement of mt-genes is essentially identical among vertebrates, although some minor differences have been found in several taxa (see Desjardins and Morais 1990; Janke *et al.* 1994). The echinoderms whose mtDNAs have been completely sequenced also exhibit very similar mt-gene arrangements, the only major variation being one inversion between the sea urchin and starfish mt-genomes (Asakawa *et al.* 1991, 1995). In arthropods, only the locations of the tRNA genes differ slightly in the examples reported so far (see Flook *et al.* 1995). On the other hand, the arrangement of molluscan mt-genes shows much variation among a polypracophoran (*Katharina tunicata*), a bivalve (Mytilus), and three pulmonate land snails (*Albinaria coerulea*, *Cepaea nemoralis*, and *Euhadra herklotsi*; Yamazaki *et al.* 1997). In nematodes, although the mt-genes of *Caenorhabditis elegans* and *Ascaris suum* are arranged in a very similar manner (Okimoto *et al.* 1992), the mt-gene arrangement of a third nematode, *Meloidogyne javanica*, is quite different (Okimoto *et al.* 1991).

Because of their unique characteristics, the primary sequences of metazoan mtDNAs have been widely used for phylogenetic analyses (Avise 1994). Other characteristics, such as gene arrangement, are also utilized in molecular phylogenetic studies (see Boore *et al.* 1995).

We previously found that the mt-genetic system of the ascidian *Halocynthia roretzi* uses a unique genetic code in that AG*R* (*R* = A or G) codons specify Gly, which is

different from the cases in vertebrates and other invertebrates (Yokobori *et al.* 1993). On the basis of our finding, we predicted that Halocynthia mtDNA encodes an additional $tRNA^{Gly}$ to enable the AGR codons to be read for Gly (Yokobori *et al.* 1993). To verify this prediction, we determined the complete nucleotide sequence of Halocynthia mtDNA. As expected, we found an additional $tRNA^{Gly}$. To our surprise, the mtDNA showed considerable differences in a variety of features (including gene content, gene arrangement, nucleotide composition, and putative tRNA structure) compared to those of other deuterostome mtDNAs. Here the unusual characteristics, including the additional $tRNA^{Gly}$, of Halocynthia mtDNA are reported and discussed.

MATERIALS AND METHODS

Preparation, cloning, and sequencing of Halocynthia mtDNA: mtDNA was prepared according to Himeno *et al.* (1987) from the hepatopancreases of Halocynthia specimens obtained in Mutsu Bay, Japan. The mtDNA was digested with *Sal*I or with a combination of *Bam*I and *Xho*I. The *Sal*I and *Bam*HI–*Xho*I fragments were ligated with *Sal*I- and *Sal*I *Bam*HI-treated pUC 18, respectively. Recombinant plasmids were transformed with JM109 and screened by the alkali-lysis method (Sambrook *et al.* 1989). A 2-kbp *Sal*I fragment and two types of 4-kbp *Bam*HI–*Xho*I fragments (~10 kbp in total) were obtained. The other regions were amplified by PCR (Saiki *et al.* 1988) using specific primers that included the sequence of each end of the inserts. The PCR fragments were cloned after digestion with suitable restriction enzymes. Oligonucleotides were synthesized with a DNA synthesizer (Applied Biosystems, Foster City, CA). Sequencing was carried out by constructing deletion mutants and primer walking, using Sequenase ver. 2 (United States Biochemicals, Cleveland) and [^{35}S]dCTPαS or [α-^{33}P]dATP. Nucleotide sequences were assembled using either the GENETYX (Software Development Co. Ltd., Tokyo) or DANASIS (Hitachi Software, Tokyo) program packages.

Data analysis: Protein and rRNA genes were identified by comparison with the mtDNA sequences of *Xenopus laevis* (Roe *et al.* 1985), *Strongylocentrotus purpuratus* (Jacobs *et al.* 1988a), and *Drosophila yakuba* (Clary and Wolstenholme 1985). For *ND4L* and *ND6*, the hydrophobic profiles (Kyte and Doolittle 1982) were also compared. tRNA genes were identified by constructing their cloverleaf configurations. Amino acid sequences were aligned using Clustal W (Thompson *et al.* 1994).

Phylogenetic analysis: To construct a maximum likelihood (ML) tree of the *COIII* nucleotide sequences (first and second codon positions) among Halocynthia, *Pyura stolonifera* (Durrheim *et al.* 1993), *Homo sapiens* (Anderson *et al.* 1981), Xenopus (Roe *et al.* 1985), Strongylocentrotus (Jacobs *et al.* 1988a), Asterina (Asakawa *et al.* 1995), and *D. yakuba* (Clary and Wolstenholme 1985), DNAML in PHYLIP ver. 3.572c (Felsenstein 1995) was used under the default condition. *D. yakuba* was employed as an outgroup. In this analysis, the nucleotide sequences of COIII were aligned according to the amino acid sequence alignment obtained with Clustal W (Thompson *et al.* 1994). The first position of Leu codons (TTR and CTN) was then replaced with Y, and the first position of Gly codons of urochordate sequences (AGR and GGN) was replaced with R.

Combined data of the inferred amino acid sequences for all the mt-protein genes except *ATPase8* were used to estimate the phylogenetic relationship among Halocynthia, Xenopus (Vertebrata; Roe *et al.* 1985), Asterina (Echinodermata; Asakawa *et al.* 1995), *D. yakuba* (Arthropoda; Clary and Wolstenholme 1985), *Apis mellifera* (Arthropoda; Crozier and Crozier 1993), *Artemia franciscana* (Arthropoda; Valverde *et al.* 1994), *Lumbricus terrestris* (Annelida; Boore and Brown 1995), *Loligo bleekeri* (Mollusca; Sasuga *et al.* 1999; K. Tomita, S. Yokobori, T. Ueda and K. Watanabe, unpublished results), Albinaria (Mollusca; Hatzoglou *et al.* 1995), Caenorhabditis (Nematoda; Okimoto *et al.* 1992), and *Metriduim senile* (Cnidaria; Beagley *et al.* 1998). Each protein was aligned with Clustal X (Thompson *et al.* 1997), ambiguously aligned regions were removed, and all the sites were then combined. PROTML of MOLPHY ver. 2.3b3 (Adachi and Hasegawa 1996) was used for ML analysis with the mtREV-F and exhaustive search options. The bootstrap probability of each tree was estimated by the RELL method (Adachi and Hasegawa 1996).

RESULTS AND DISCUSSION

General features of Halocynthia mtDNA: The nucleotide sequence of Halocynthia mtDNA has been deposited in the DDBJ/GenBank/EMBL DNA databases under the accession no. AB024528. The complete length (14,771 bp) of the mtDNA is shorter than those of vertebrate mtDNAs [*e.g.*, the 16,201 bp-long Petromyzon mtDNA is the shortest among vertebrate mtDNAs completely sequenced (Lee and Kocher 1995)], echinoderm [*e.g.*, Strongylocentrotus: 15,650 bp (Jacobs *et al.* 1988a)], arthropod [*e.g.*, Locusta: 15,722 bp (Flook *et al.* 1995)], and an annelid [14,998 bp in Lumbricus (Boore and Brown 1995)], but longer than that of nematode [13,794 bp in Caenorhabditis and 14,284 bp in Ascaris (Okimoto *et al.* 1992)] and pulmonate [14,100 bp in Cepaea (Terrett *et al.* 1996) and 14,130 bp in Albinaria (Hatzoglou *et al.* 1995)] mtDNAs.

The genes of 12 proteins, two rRNAs, and 23 tRNAs were identified in the Halocynthia mt-genome (Figure 1). Like the mtDNAs of nematodes (Okimoto *et al.* 1992) and Mytilus (Hoffmann *et al.* 1992), Halocynthia mtDNA encodes no *ATPase8*; also, there is not any noncoding region capable of encoding. From this, it is assumed that the loss of *ATPase8* from mt-genomes occurred at least three times, independently, during the evolution of metazoans. The Halocynthia mtDNA is also unique in that it encodes an additional $tRNA^{Gly}$ (discussed below).

All the genes encoded by Halocynthia mtDNA are located on the same strand. This is similar to some mtDNAs such as nematodes (Okimoto *et al.* 1992), but different from the case of other deuterostomes (*e.g.*, Anderson *et al.* 1981; Asakawa *et al.* 1995).

The nucleotide composition of the coding strand of Halocynthia mtDNA is shown in Table 1 along with those of various other metazoan mtDNAs, from which it can be seen that the Halocynthia mtDNA has a nucleotide composition similar to that of Ascaris rather than to vertebrate and echinoderm mtDNAs, with T being

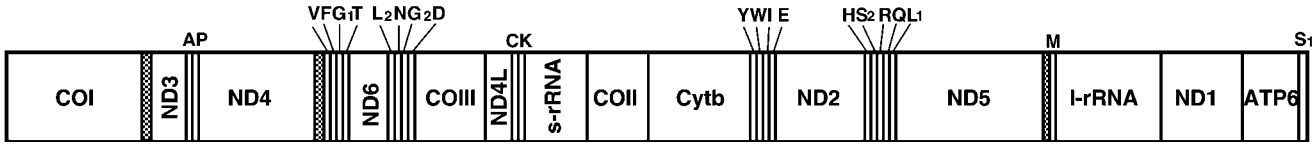

Figure 1.—Gene organization of the Halocynthia mt-genome. All the genes are transcribed in the same direction (left to right). The abbreviations for protein and rRNA genes are as used in the text. tRNA genes are indicated by the one-letter code of the corresponding amino acid. In addition, L1, L2, S1, S2, G1, and G2 stand for $tRNA^{Leu}_{UAA}$, $tRNA^{Leu}_{UAG}$, $tRNA^{Ser}_{UGA}$, $tRNA^{Ser}_{GCU}$, $tRNA^{Gly}_{UCU}$, and $tRNA^{Gly}_{UCC}$, respectively. Noncoding regions are shaded.

used most frequently and C the least frequently. An asymmetric nucleotide composition (in which the A content is unequal to the T content and the G content is unequal to the C content in the same strand of the DNA duplex) has been reported for metazoan mtDNAs (Anderson *et al.* 1981; Asakawa *et al.* 1991; W. K. Thomas and A. C. Wilson, unpublished results). In vertebrate and starfish mtDNAs, the major coding strand is rich in A and C rather than G and T (strand-specific nucleotide composition bias; Asakawa *et al.* 1991; W. K. Thomas and A. C. Wilson, unpublished results). It has been postulated that the displacement-type replication unique to metazoan mtDNAs (Attardi 1985; Clayton 1992) causes the asymmetric nucleotide composition. In the case of vertebrate mtDNAs, the AC-rich strand is the second strand for replication (see W. K. Thomas and A. C. Wilson, unpublished results). Accordingly, the major coding strand of Halocynthia mtDNA may be the first strand for replication; however, its mode of replication is as yet unknown. Elucidation of this would facilitate our understanding of the substitution pattern of metazoan mtDNAs.

Genetic code and tRNA genes: *Genetic code:* The previously determined partial sequence of Halocynthia *COI* (Yokobori *et al.* 1993) indicated that Halocynthia mtDNA uses the following genetic code: AUA specifies Met, UGA specifies Trp, and AGA and AGG specify Gly. This is now confirmed by the analysis of the complete mtDNA sequence presented here. The same codon table variations have been reported for another ascidian, *Pyura stolonifera* (Durrheim *et al.* 1993).

General features of tRNA genes: The cloverleaf structures of the 23 tDNAs in the Halocynthia mtDNA are presented in Figure 2A. The set includes an additional tRNA gene, $tRNA^{Gly}$ for AGR codons (see below), in addition to the 22 known standard tRNA genes.

Among these 23 tRNA genes, only $tRNA^{Ser}_{GCU}$ has both the "conserved" GG and TTC sequences in the D and T loops, respectively (Figure 2A). The $tRNA^{Tyr}$ has the GG sequence in the D loop but no TTC sequence in the T loop. It has been suggested that in some metazoan mt-tRNAs, there is no direct contact between the D and T loops (Wakita *et al.* 1994; Watanabe *et al.* 1994), and this seems to hold for the Halocynthia mt-tRNA genes. However, like bovine mt-$tRNA^{Phe}$ (Wakita *et al.* 1994), 21 out of the 23 Halocynthia mt-tRNA genes may preserve interaction between the D and extra loops, which could play an important role in the formation of

TABLE 1

Nucleotide composition (in percentage terms) of various metazoan mtDNAs whose complete sequences have been determined

Organism	A	G	C	T	A + T	G + T	Reference
Halocynthia	24.25	23.21	8.52	**44.00**	68.25	**67.21**	This work
Human	**30.91**	13.13	31.23	*24.70*	*55.61*	*37.83*	Anderson *et al.* (1981)
Xenopus	**33.05**	13.49	23.48	*29.96*	*63.01*	*43.45*	Roe *et al.* (1985)
Strongylocentrotus	28.74	18.35	22.66	***30.23***	*58.97*	*48.58*	Jacobs *et al.* (1988a)
Asterina	**32.40**	14.12	24.60	*28.86*	*61.26*	*42.98*	Asakawa *et al.* (1995)
Drosophila	**39.49**	9.24	12.16	*39.09*	**78.58**	*48.33*	Clary and Wolstenholme (1985)
Apis	**43.20**	5.53	9.60	*41.65*	**84.85**	*47.18*	Crozier and Crozier (1993)
Artemia	30.95	17.68	17.87	***33.48***	*64.43*	***51.16***	Valverde *et al.* (1994)
Katharina	31.39	*18.61*	11.89	***38.01***	69.40	***56.62***	Boore and Brown (1994)
Cepaea	26.16	*21.26*	18.94	***33.63***	*59.79*	***54.89***	Terrett *et al.* (1996)
Lumbricus	29.84	15.80	22.54	***31.77***	*61.61*	*47.57*	Boore and Brown (1995)
Caenorhabditis	31.42	*14.89*	8.88	**44.79**	**76.21**	***59.68***	Okimoto *et al.* (1992)
Ascaris	*22.17*	*20.36*	7.66	**49.78**	**71.95**	**70.14**	Okimoto *et al.* (1992)

The nucleotide compositions were calculated from the major coding strand. The frequency of the nucleotide used most frequently in each genome is indicated in boldface type, and that of the nucleotide used least frequently is underlined. Frequencies of A and T (A + T) of >70% and of G and T (G + T) of >60% are indicated in boldface type. In all cases, values lower than the Halocynthia counterparts are italicized.

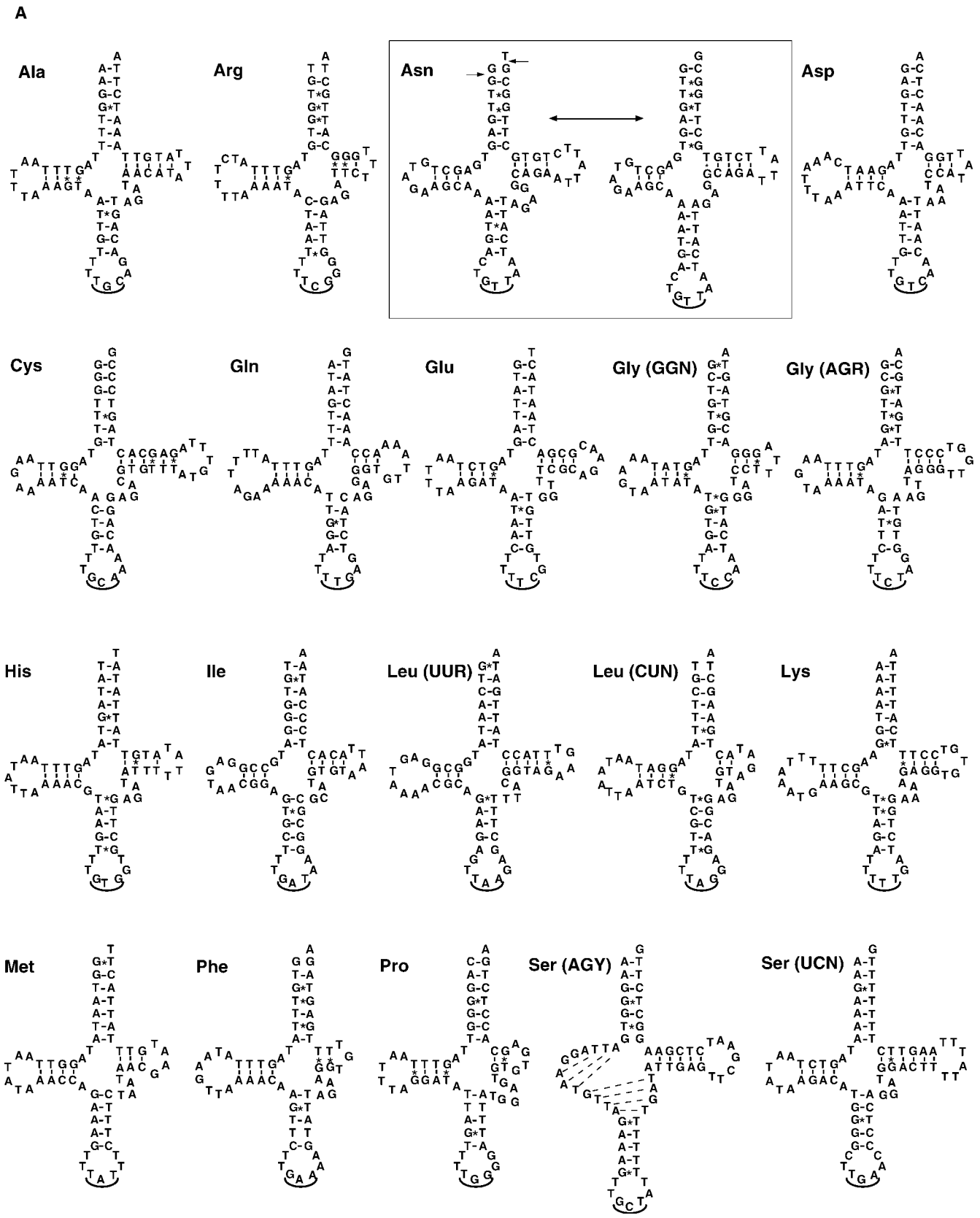

Figure 2.—(A) Putative secondary structures of Halocynthia mt-tRNA genes. (B) Putative Pyura mt-tRNA genes for Glu, Tyr, and His. The sequences were taken from GenBank (accession nos. X74513 and X75386). Nucleotides in the putative Pyura mt-tRNA genes shared with their Halocynthia counterparts are printed in outline style.

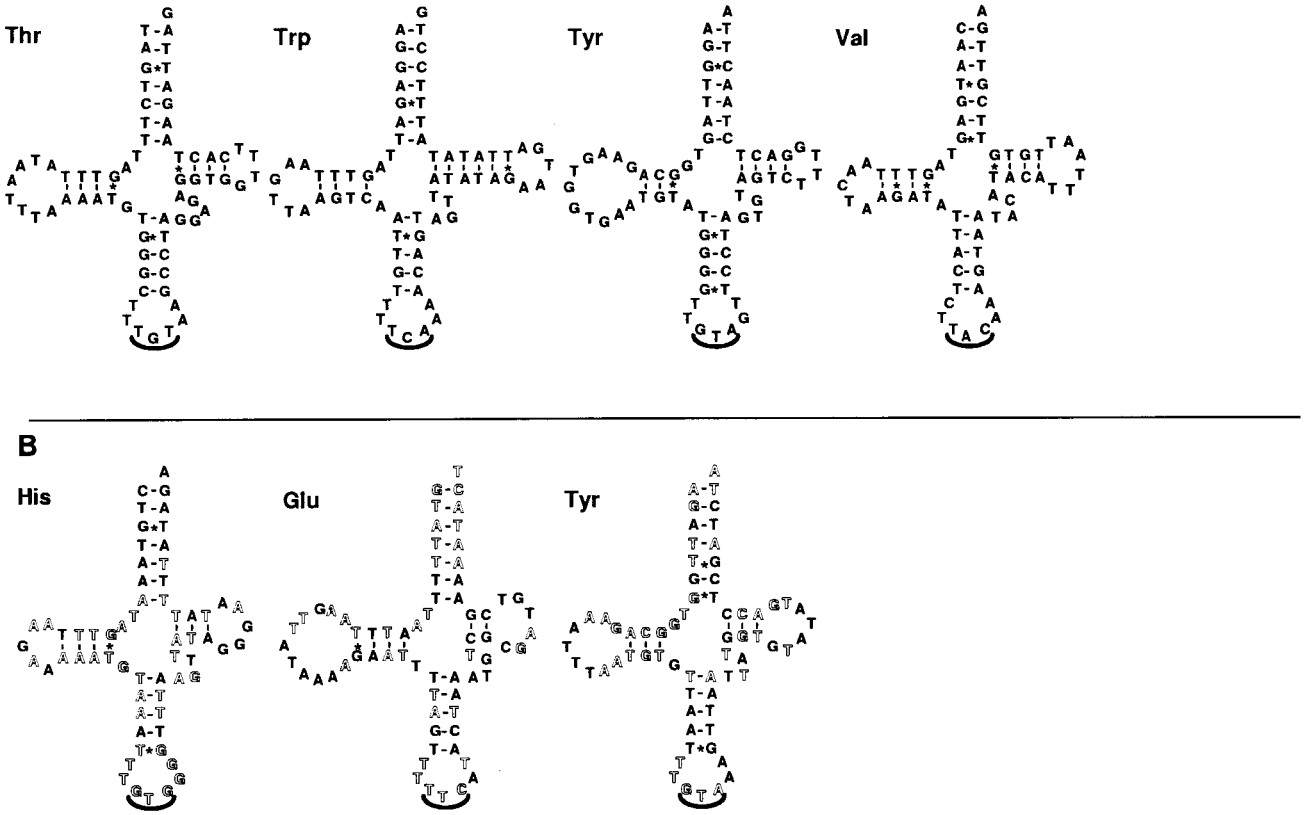

Figure 2.—*Continued.*

the L-shape-like structure of the tRNA. The length of the T stem of the Halocynthia mt-tRNA genes varies from 2 to 6 bp. As in other metazoan mt-tRNA^{Ser}s for AGY/AGN (Y = U/T or C, and N = U/T, C, A, or G) codons (Steinberg *et al.* 1994), the Halocynthia mt-*tRNA*$^{Ser}_{GCU}$ gene has a quite unusual secondary structure in that up to 9 bp can be formed in the anticodon stem.

The *tRNA*^{Arg}, *tRNA*^{Asn} (the left of the two boxed structures in Figure 2A) and *tRNA*^{Phe} genes have a mismatch at the top base pair of the acceptor stem (Figure 2A). An alternative cloverleaf configuration can be formed for the *tRNA*^{Asn} gene (the right boxed structure), which is unusual in that there are two nucleotides between the D and anticodon stems. The acceptor and T stems seem to be more stable in this configuration than in the left-hand one, and this alternative tRNA^{Asn} structure is similar to that of the echinoderm mt-tRNA^{Ile} genes proposed by Jacobs *et al.* (1988a) and De Giorgi *et al.* (1996). To confirm the secondary structure of each tRNA, direct analysis of the tRNA sequence as well as of the higher order structure is required (see Yokogawa *et al.* 1991).

tRNA^{Gly} *gene for AGR codons and its origin:* As already noted, a tRNA gene with the anticodon TCT that can recognize either AGN or AGR codons was found in the Halocynthia mt-genome (Figure 2A). Sequence analysis of this tRNA molecule revealed that the anticodon first position was modified to an unknown U derivative (U*). The amino acid attached to this tRNA *in vivo* was glycine. From these results it is almost certain that this tRNA decodes only AGR codons but not AGY codons as glycine (Kondo *et al.* 1996; Kondow *et al.* 1999).

There may be some possible explanations for origin of the *tRNA*$^{Gly}_{UCU}$. First, the *tRNA*$^{Gly}_{UCU}$ gene would have originated from duplication of *tRNA*$^{Gly}_{UCC}$ gene. Second, the *tRNA*$^{Gly}_{UCU}$ gene would have originated from duplication of a certain tRNA gene that does not correspond to Gly. The former possibility may be most plausible because only one base replacement at the anticodon third position should be adequate for the tRNA to become recognizable toward AGR codons as Gly.

However, at present, the latter possibility could not be excluded either. The identity elements determining the recognition mechanism of tRNAs by their cognate aminoacyl tRNA synthetases (ARSs) seem to be smaller in their size and number in metazoan mitochondria than in cytoplasm (Kumazawa *et al.* 1989, 1991). Therefore, a few substitutions on the mt-tRNA genes may potentially be able to change the tRNA identity. Indeed, there is one apparent example that only one base substitution can change the tRNA identity. Opossum mt-*tRNA*^{Asp} gene has the GCC anticodon, and the tRNA^{Asp} carrying this GCC anticodon is aminoacylated with glycine but not with aspartate; only after the RNA editing that causes the anticodon to change from GCC to GUC occurs, the tRNA^{Asp} becomes chargeable with aspartate

(Janke and Pääbo 1993; Mörl *et al.* 1995; Börner *et al.* 1996). It should be noted that, when two $tRNA^{Gly}$ ($tRNA^{Gly}_{UCU}$ and $tRNA^{Gly}_{UCC}$) genes are compared, only 33 out of 61 nucleotides are common to these two tRNA genes (54% homology; Figure 2A).

In both cases above, the gene shuffling might have been needed for duplication of the tRNA gene. Judging from a suggestion that recruitment of a tRNA gene as a part of a protein gene can occur in echinoderm mtDNA (Cantatore *et al.* 1987), it is also possible to imagine that a tRNA gene corresponding to a certain amino acid may have been recruited as another tRNA gene corresponding to another amino acid. The gene organization of the Halocynthia mt-genome differs so much from those of other metazoan mitochondria that it might have resulted from the frequent gene rearrangement in Halocynthia mt-genome. It might, in turn, have borne an extra tRNA gene in the Halocynthia mt-genome in addition to the standard set of metazoan mt-tRNA genes.

It should be noted that the Mytilus mtDNA possessing an additional $tRNA^{Met}$ gene ($tRNA^{Met}_{UAU}$) has a very different gene arrangement from those of other molluscan mtDNAs (Hoffmann *et al.* 1992; Boore and Brown 1994).

$tRNA^{Met}$: We identified a $tRNA^{Met}$ gene with the anticodon TAT, which can potentially recognize AUR codons (Figure 2A), instead of the typical $tRNA^{Met}$ gene with the anticodon CAT. Direct sequencing of the tRNA showed that this tRNA has the anticodon U*AU, which is capable of decoding the codon AUA as well as AUG (Kondow *et al.* 1998). The Mytilus mt-genome is known to possess the same type of $tRNA^{Met}_{TAT}$ gene, but it also encodes another $tRNA^{Met}_{CAT}$ gene. There have been no reports of any other $tRNA^{Met}_{TAT}$ genes except for those in Halocynthia and Mytilus mt-genetic systems.

Protein genes: The 5′-end of the Halocynthia *ND3* seems to be longer by ~60 bp (20 amino acids) than *ND3*s of vertebrates, echinoderms, and arthropods. However, there is a possibility that the actual *ND3* starts from the 21st ATG codon, which would give a gene size similar to those of other metazoans.

As shown in Table 2, for all the protein genes the amino acid sequence similarities are much lower between Halocynthia and Xenopus, and between Halocynthia and Asterina than between Xenopus and Asterina. The *COIII* of Pyura (Durrheim *et al.* 1993) is the only available representative of urochordate mt-protein genes apart from the Halocynthia mt-protein genes sequenced here. The amino acid sequence similarity between Pyura and Halocynthia is almost the same as that between Xenopus and Drosophila and that between Asterina and Drosophila, but rather lower than that between Xenopus and Asterina (Table 2). This might in part be a reflection of the difference in the nucleotide composition of the mtDNA encoding *COIII* (discussed in later section). However, the fact that the primary

TABLE 2

Pairwise similarities of amino acid sequences for 12 protein genes encoded by Halocynthia mtDNA to several metazoan counterparts

Protein	(Hro/Pst)	Hro/Xla	Hro/Ape	Hro/Dya	Hro/Cel	Xla/Ape	Xla/Dya	Xla/Cel	Ape/Dya	Ape/Cel	Dya/Cel
ATPase6		35.1	27.6	23.2	23.3	46.0	34.1	24.9	37.1	21.4	25.6
COI		66.0	66.0	68.5	57.9	73.5	74.4	57.8	74.6	61.9	63.1
COII		42.7	47.6	40.0	39.6	61.6	58.8	39.5	62.7	41.2	40.4
COIII	(67.8)	54.8	54.2	53.5	40.8	71.8	65.0	48.0	68.1	44.9	46.7
		(53.8)	(49.6)	(50.2)	(39.2)						
Cytb		50.4	51.7	51.4	40.7	68.0	66.8	41.9	66.4	43.5	44.1
ND1		37.7	38.6	37.4	36.0	53.7	47.9	39.5	44.7	37.1	39.9
ND2		19.7	23.3	22.5	17.0	39.4	35.2	19.9	33.5	15.6	19.3
ND3		24.8	26.9	29.4	28.4	53.3	40.7	23.9	48.1	27.4	28.8
ND4		27.2	28.7	28.5	27.3	45.2	38.1	25.2	39.8	26.2	32.8
ND4L		13.3	12.2	14.9	18.2	31.6	30.9	16.9	33.0	20.8	27.3
ND5		32.6	32.5	27.9	28.9	45.8	30.6	23.2	32.6	28.0	30.8
ND6		21.9	23.4	18.5	16.9	33.3	19.9	14.9	18.9	15.9	20.1

The percentages of shared amino acids in pairwise comparisions of two sequences are shown. All gaps are excluded from the analysis. Hro, Halocynthia; Pst, Pyura (Durrheim *et al.* 1993); Xla, Xenopus (Roe *et al.* 1985); Ape, Asterina (Asakawa *et al.* 1995); Dya, Drosophila (Clary and Wolstenholme 1985); Cel, Caenorhabditis (Okimoto *et al.* 1992). All values between Pyura *COIII* and other *COIII* are placed in parentheses.

sequence of *COIII* is so different in Halocynthia and Pyura makes it likely that the substitution rate of the amino acid sequence, and probably also of the nucleotide sequence, is accelerated in urochordate mtDNA.

Initiation and termination codons: Only three protein genes, *COI*, *ND2*, and *ND4L*, start with the ATG codon; *ATPase6*, *COIII*, *ND4*, and *ND6* start with GTG, and *COII*, *ND1*, and *ND3* with ATA. It can be predicted that *Cytb* starts with ATT and *ND5* with TTG. Seven protein genes terminate with the TAA codon, and three with TAG. *ATPase6* seems to terminate with T followed directly by the $tRNA^{Ser}_{TGA}$ gene. Hence, the complete termination codon for the *ATPase6* mRNA may be UAA, which can be created by polyadenylation, as is found in several mammalian mt-mRNAs (Anderson *et al.* 1981; Ojala *et al.* 1981). *ND4L* does not have a complete termination codon if overlap with the downstream $tRNA^{Cys}$ gene is not allowed. A possible termination codon for *ND4L*, TAG, appears at the 8th to 10th position of the $tRNA^{Cys}$ gene. Alternatively, *ND4L* could terminate with an incomplete termination codon, T (Ojala *et al.* 1981); however, there is an additional T between the incomplete termination codon and the $tRNA^{Cys}$ gene. Similar situations have been observed in some metazoan mtDNA, including Cepaea (Terrett *et al.* 1996; Yamazaki *et al.* 1997) and *Ornithorhynchus anatinus* (platypus; Janke *et al.* 1996).

Codon usage and deduced amino acid composition in protein genes: The frequency of the nucleotide at the third codon position of Halocynthia mt-protein genes reflects the overall nucleotide composition of the mtDNA: T is used most frequently and C the least frequently at the codon third position. The five most frequently used codons are TTT-Phe, TTA-Leu, TTG-Leu, GTT-Val, and ATT-Ile, all of which are rich in T. In contrast, several codons ending with C, for example, TGC-Cys, scarcely appear in Halocynthia mt-protein genes.

In the mt-protein genes of Halocynthia, Drosophila, and Caenorhabditis, TTR-Leu codons are preferred over CTN-Leu, whereas, in those of Xenopus and Asterina that are encoded by the major coding strand, CTN-Leu codons are preferred over TTR-Leu (Table 3). However, among these five animal mt-protein genes, the frequency of occurrence of Leu is not very different. AAY-Asn, ATY-Ile, and ACN-Thr are not used as much in Halocynthia as in the mt-protein genes of Xenopus, Asterina, Drosophila, and Caenorhabditis. In contrast, TGY-Cys, AGR-Gly, GGN-Gly, and GTN-Val are utilized in Halocynthia much more frequently than in Xenopus, Asterina, Drosophila, and Caenorhabditis. The excessive use of Gly in Halocynthia mt-protein genes can be explained in part by the number of Gly codons: in Halocynthia mitochondria, Gly is assigned to a six-codon family, whereas it is assigned to a four-codon family in other mitochondria (Osawa 1995; Watanabe and Osawa 1996). Since the coding strand of the Halocynthia mt-genome is rich in T, rather than A, and G, rather than C, ACN-Thr would be used less frequently, and GTN-Val and TGY-Cys more frequently, than in the mt-genomes rich in A, rather than T, and C, rather than G—such as vertebrate and echinoderm mt-genomes are (see Asakawa *et al.* 1991; Perna and Kocher 1995; W. K. Thomas and A. C. Wilson, unpublished results). Indeed, the Caenorhabditis mt-genome prefers T to A, and G to C, and the Caenorhabditis and Halocynthia mt-protein genes share an abundance of TTY-Phe and a dearth of CAY-His, CCN-Pro, and TCN-Ser, compared with those of Xenopus, Asterina, and Drosophila.

Ribosomal RNA genes: The Halocynthia mt-*srRNA* is estimated to be at most 706 bp long. This is as short as those of nematodes [*e.g.*, 697 bp in Caenorhabditis (Okimoto *et al.* 1992)], but much shorter than those of vertebrates and echinoderms [*e.g.*, 954 bp in human (Anderson *et al.* 1981); 878 bp in Strongylocentrotus (Jacobs *et al.* 1988a)].

The Halocynthia mt-*lrRNA* is estimated to be at most 1178 bp long. This is also much shorter than those of vertebrates and echinoderms [*e.g.*, 1559 bp in human (Anderson *et al.* 1981); 1530 bp in Strongylocentrotus (Jacobs *et al.* 1988a)], but longer than those of nematodes [*e.g.*, 953 bp in Caenorhabditis (Okimoto *et al.* 1992)]. A presumed secondary structure model of the Halocynthia mt-*lrRNA* was constructed (Y. Muramatsu, S. Yokobori and T. Oshima, unpublished results). This was found to retain the secondary structure of the peptidyltransferase center that is apparently essential for ribosomal function and protein synthesis (see Green and Noller 1997), and to be more similar to the predicted nematode mt-lrRNA secondary model (Okimoto *et al.* 1994) than to those of vertebrates.

Unassigned sequences: The longest noncoding sequence in Halocynthia mtDNA consists of 115 bp and is located between *COI* and *ND3*. Other noncoding sequences of 112 bp and 79 bp occur between *ND4* and $tRNA^{Val}$, and $tRNA^{Leu}_{TAA}$ and *ND5*, respectively. As shown in Figure 3, the noncoding regions between *COI* and *ND3* and between *ND4* and $tRNA^{Val}$ can be folded into reasonable secondary structures. There can be a stem-and-loop structure in the noncoding region between *COI* and *ND3*. The 20-nucleotide-long loop of this structure is rich in thymine residues (T_{18} and flanking Gs exist at both ends). In the case of vertebrate mt-genomes, a stem-and-loop structure is found as the replication origin of the L strand (Clayton 1992). The loop sequence is generally rich in adenosine (or purine) residues in the L strand (the latter strand in the replication process), and in its antisense strand, the T (or pyrimidine)-rich sequence in the loop region is known to work as the template for RNA primer synthesis in L strand replication (Clayton 1992). Nothing is yet known about the replication process of Halocynthia mtDNA, but it can be supposed that the putative stem-and-loop structure in the noncoding region between *COI* and *ND3* is somehow involved in the initiation of replication.

TABLE 3

Comparison of amino acid frequency (percentage) in mitochondrially encoded protein genes

Amino acid	(Codon)	Halocynthia[a]	Xenopus[b]	Asterina[c]	Drosophila[d]	Caenorhabditis[e]
Ala	**(GCN)**	**3.99**	**7.27 (1.8)**	**6.37 (1.6)**	**4.64 (1.2)**	**3.33 (0.8)**
Arg	**(CGN)**	**1.92**	**1.77 (0.9)**	**1.94 (1.0)**	**1.58 (0.8)**	**0.90 (0.5)**
Asn	**(AAY/Y + A)[f]**	**2.32**	**4.04 (1.7)**	**5.80 (2.5)**	**5.52 (2.4)**	**4.41 (1.9)**
Asp	**(GAY)**	**1.90**	**1.99 (1.0)**	**1.97 (1.0)**	**1.71 (0.9)**	**1.84 (1.0)**
Cys	**(TGY)**	**2.12**	**0.77 (0.4)**	**0.80 (0.4)**	**1.12 (0.5)**	**1.37 (0.6)**
Gln	**(CAR)**	**1.17**	**2.74 (2.3)**	**2.24 (1.9)**	**1.87 (1.6)**	**1.28 (1.1)**
Glu	**(GAR)**	**2.46**	**2.57 (1.0)**	**2.21 (0.9)**	**2.22 (0.9)**	**2.28 (0.9)**
Gly	(GGN)	7.16	5.64 (0.8)	5.46 (0.8)	5.90 (0.8)	5.46 (0.8)
Gly	(AGR)[g]	4.76	—	—	—	—
Gly	**(all)**	**11.92**	**5.64 (0.5)**	**5.46 (0.5)**	**5.90 (0.5)**	**5.46 (0.5)**
His	**(CAY)**	**1.78**	**2.65 (1.5)**	**2.58 (1.4)**	**2.06 (1.2)**	**1.69 (0.9)**
Ile	**(ATY/Y + A)[h]**	**5.08**	**9.05 (1.8)**	**10.66 (2.1)**	**9.65 (1.9)**	**8.18 (1.6)**
Leu	(TTR)	13.45	6.09 (0.5)	3.62 (0.3)	15.21 (1.1)	12.77 (0.9)
Leu	(CTN)	1.98	10.02 (5.1)	12.01 (6.1)	1.58 (0.8)	2.72 (1.4)
Leu	**(all)**	**15.43**	**16.11 (1.0)**	**15.63 (1.0)**	**16.79 (1.1)**	**15.49 (1.0)**
Lys	**(AAR/G)[i]**	**1.98**	**2.26 (1.1)**	**1.03 (0.5)**	**2.28 (1.2)**	**3.18 (1.6)**
Met	**(ATR/G)[j]**	**4.80**	**5.17 (1.1)**	**1.97 (0.4)**	**5.68 (1.2)**	**5.20 (1.1)**
Phe	**(TTY)**	**11.65**	**6.08 (0.5)**	**8.48 (0.7)**	**8.85 (0.8)**	**13.21 (1.1)**
Pro	**(CCN)**	**2.73**	**5.53 (2.0)**	**4.83 (1.8)**	**3.48 (1.3)**	**2.36 (0.9)**
Ser	(TCN)	3.44	6.34 (1.8)	7.08 (2.1)	6.17 (1.8)	4.24 (1.2)
Ser	(AGN/Y + A/Y)[k]	3.61	1.44 (0.4)	2.58 (0.7)	2.90 (0.8)	6.78 (1.9)
Ser	**(all)**	**7.05**	**7.78 (1.1)**	**9.66 (1.4)**	**9.07 (1.3)**	**11.02 (1.6)**
Thr	**(ACN)**	**2.48**	**8.49 (3.4)**	**7.91 (3.2)**	**5.01 (2.0)**	**4.20 (1.7)**
Trp	**(TGR)**	**3.29**	**2.93 (0.9)**	**2.78 (0.8)**	**2.73 (0.8)**	**2.07 (0.6)**
Tyr	**(TAY)**	**4.66**	**2.93 (0.6)**	**2.85 (0.6)**	**4.53 (1.0)**	**5.05 (1.1)**
Val	**(GTN)**	**11.57**	**4.15 (0.4)**	**4.72 (0.4)**	**4.20 (0.4)**	**7.39 (0.6)**

Total amino acid frequencies (%) are shown. The relative frequency of each amino acid with respect to Halocynthia is shown in parentheses. The definitions of GT-, AC-, AT-, and GC-rich amino acids are given in the text. Sums for each amino acid are shown in boldface type.

[a] All the protein genes are used.
[b] Roe *et al.* (1985); all the protein genes except *ND6* are used.
[c] Asakawa *et al.* (1995); all the protein genes except *ND1*, *ND2*, and *ND6* are used.
[d] Clary and Wolstenholme (1985); all the protein genes are used.
[e] Okimoto *et al.* (1992); all the protein genes are used.
[f] AAY for Halocynthia, Xenopus, Drosophila, and Caenorhabditis; AAY/AAA for Asterina.
[g] GGN for Xenopus, Asterina, Drosophila, and Caenorhabditis; AGR/GGN for Halocynthia.
[h] ATY for Halocynthia, Xenopus, Drosophila, and Caenorhabditis; ATY/ATA for Asterina.
[i] AAR for Halocynthia, Xenopus, Drosophila, and Caenorhabditis; AAG for Asterina.
[j] ATR for Halocynthia, Xenopus, Drosophila, and Caenorhabditis; ATG for Asterina.
[k] TCN/AGY for Halocynthia and Xenopus; TCN/AGN for Asterina and Caenorhabditis; TCN/AGY/AGA for Drosophila.

The boundaries between the following genes—*lrRNA* and *ND1*, *ND1* and *ATPase6*, *COIII* and *ND4L*, *srRNA* and *COII*, and *COII* and *Cytb*—are not interrupted by tRNA genes nor by noncoding sequences longer than 20 bp. In general, the tRNA parts in nascent multicistronic transcripts are thought to be processing signals, resulting in mRNAs and rRNAs (Ojala *et al.* 1981), and protein–protein (or rRNA) gene boundaries without the interruption of a tRNA gene have potential stem-and-loop structures, which have been proposed to act as processing signals. This is also likely to be the case for the Halocynthia mt-genetic system, although no data on the mode of the transcription is yet available. Stem-and-loop structures can also be formed at the boundary regions between *ND1* and *ATPase6*, *COIII* and *ND4L*, and *COII* and *Cytb* (data not shown).

Gene arrangement: The gene arrangement of the Halocynthia mt-genome is quite different from those of other metazoan mt-genomes, including vertebrates and echinoderms. The order and direction of $tRNA^{His}$–$tRNA^{Ser}_{GCU}$ is common among vertebrate, echinoderm, and Halocynthia mtDNAs, but echinoderm mtDNAs share many additional gene boundaries with those of vertebrates. Even if tRNA genes are excluded from the comparison, only the order and direction of *lrRNA–ND1* is shared by vertebrate and Halocynthia mtDNAs, and only the *ND3–ND4* boundary by echinoderm and Halocynthia mtDNAs (Figure 4). It can be seen that while there

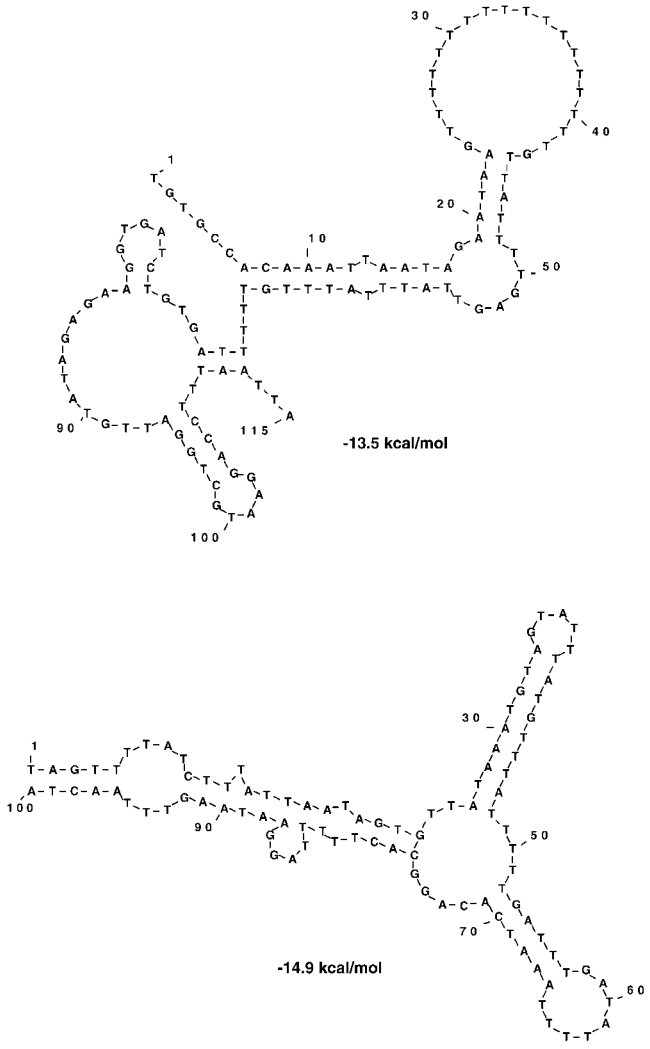

Figure 3.—Putative secondary structures with calculated energies of the noncoding regions between *COI* and *ND3* (top) and *ND4* and $tRNA^{Val}$ (bottom).

are two major gene transpositions between the human and sea urchin mt-genomes, there are many transpositions between the Holocynthia and human mt-genomes and between the Halocynthia and sea urchin mt-genomes.

The only other urochordate mtDNA sequences so far available are three fragments from Pyura: *COIII* (Durrheim *et al.* 1993), *lrRNA* (EMBL/GenBank/DDBJ DNA databank accession no. X74513), and $tRNA^{His}$ (X75386, the cloverleaf structure of which is shown in Figure 2B). The Pyura "mt-*lrRNA*" is reported to be ~400 bp longer than that of Halocynthia. However, since two tRNA cloverleaf structures ($tRNA^{Glu}$ and $tRNA^{Tyr}$) can be formed in the Pyura mt-*lrRNA* (Figure 2B), it may not be longer than 1108 bp. The pairwise similarity of this "shortened" Pyura mt-*lrRNA* with Halocynthia mt-*lrRNA* is 57.7%. In addition, the region upstream of $tRNA^{Glu}$ shows sequence similarity with the 3′ end of the Halocynthia *srRNA* (the similarity is 65.9%). The pairwise similarities of mt-$tRNA^{Glu}$ and mt-$tRNA^{Tyr}$ between Pyura and Halocynthia are 54.8% and 57.8%, respectively. However, the similarity of mt-$tRNA^{His}$ between Halocynthia and Pyura is 61.3%, suggesting that the mt-$tRNA^{Glu}$ and mt-$tRNA^{Tyr}$ similarities between Halocynthia and Pyura are not so low. If our prediction is correct, the partial gene order in the Pyura mt-genome would be *srRNA*–$tRNA^{Glu}$–$tRNA^{Tyr}$–*lrRNA*. This is different from Halocynthia, in which mt-*lrRNA* follows $tRNA^{Met}_{TAT}$, and is also different from the gene order of vertebrates and echinoderms.

Jacobs *et al.* (1988b) reported that in another ascidian, *Styela clava*, *COI* is ~7 kbp downstream of *lrRNA*, although in the Halocynthia mtDNA *lrRNA* is 1.5 kbp away from *COI*. Like Halocynthia and Pyura, Styela belongs to the suborder Stolidobranchia (Pearse *et al.* 1987), but the mt-gene arrangement in Halocynthia is different from those in both Styela and Pyura.

It can be concluded that the mt-gene arrangement is more changeable in ascidians than in vertebrates, although both groups belong to Chordata. Additional data on the mt-gene arrangement of other urochordates are needed to elucidate the evolution of metazoan mt-gene structures.

Evolutionary considerations: A phylogenetic analysis based on the first and second codon positions of *COIII* was performed by the maximum likelihood (ML) method using DNAML in PHYLIP (Felsenstein 1995). Halocynthia (Hro), Pyura (Pst), Homo (Hsa), Xenopus (Xla), Strongylocentrotus (Spu), and Asterina (Ape) were used for the analysis and *D. yakuba* (Dya) was used as an outgroup. The log likelihood (lnL) of the ML tree obtained (Figure 5) is −2534.00, and its topology matches the traditional view: [(Arthropods (Echinoderms (Urochordates, Vertebrates)))]. A neighbor-joining tree (Saitou and Nei 1987) was also constructed. It gave the same topology as the ML tree (data not shown). However, since the differences of the lnL of the ML tree from the trees [(Arthropods (Urochordates (Echinoderms, Vertebrates)))] and [(Arthropods (Vertebrates (Echinoderms, Urochordates)))] are only −10.55 (SD ± 7.51) and −12.28 (SD ± 6.89), respectively, supporting the monophyly of deuterostomes in this analysis is not strong. It is apparent that the branches of urochordates are more than twice as long as those of vertebrates after the separation of vertebrates and urochordates. As shown in Figure 5, both the AT and GT contents of the codon first and second positions of *COIII* are higher in the urochordate line than in the echinoderm and vertebrate lines. This may explain in part why the urochordate branches are longer than the echinoderm and vertebrate ones; under the different nucleotide composition bias the nucleotide substitution pattern may differ, and hence the "observed" substitution rate.

An expanded phylogenetic analysis of metazoans (Xenopus, Halocynthia, Asterina, *D. yakuba*, Apis, Lumbricus, Loligo, Albinaria, and Caenorhabditis) was car-

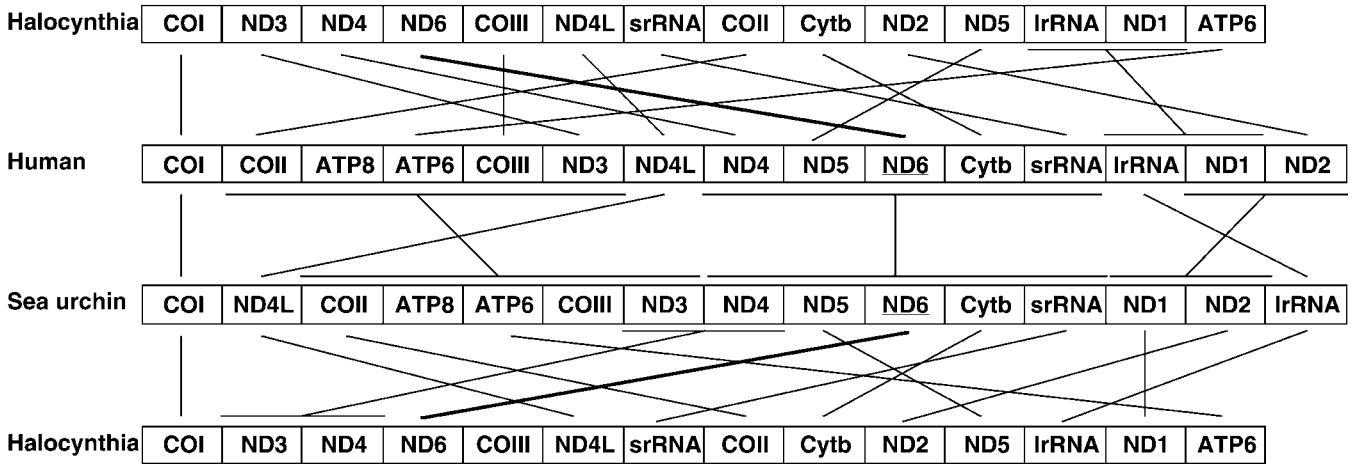

Figure 4.—Comparison of gene organization among Halocynthia, human (Anderson *et al.* 1981), and sea urchin (Jacobs *et al.* 1988a; Cantatore *et al.* 1989) mt-genomes. tRNA genes and noncoding regions are excluded from the comparison. Relocation of genes without inversion is shown by thin lines and with inversion by thick lines.

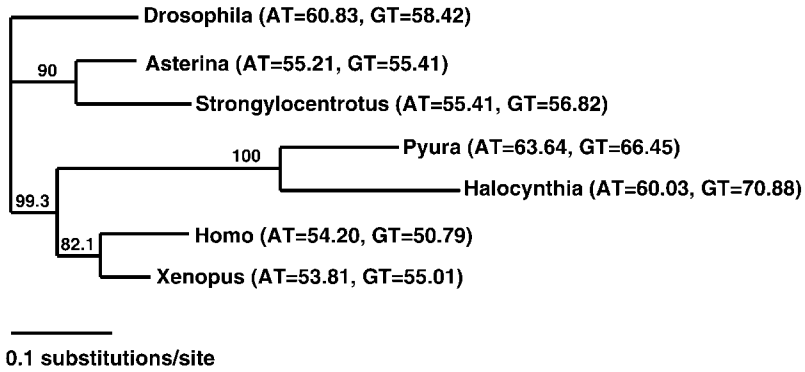

Figure 5.—Maximum likelihood (ML) tree based on the first and second codon positions of the *COIII* nucleotide sequence (498 sites). DNAML in PHYLIP Ver. 3.572c (Felsenstein 1995) was used. The bootstrap probability of each branch is indicated as the percentage of the occurrent trees supporting the branch in 1000 replicates.

ried out using amino acid sequences inferred from mt-protein genes. An ML tree based on inferred amino acid sequences (2056 sites in total) of all the mt-protein genes except for *ATPase8* was reconstructed by PROTML in MOLPHY (Adachi and Hasegawa 1996; Figure 6). The cnidarian Metridium (Mse) was used as an outgroup. The lnL of the ML tree is −31209.8. The tree does not support the monophyly of chordates, since Halocynthia is the sister taxon of other triploblasticans (Figure 6). In addition, another 36 tree topologies in the analysis could not be rejected since their lnL differences from the ML tree were smaller than 2 SE.

The bootstrap probability (BP) of the ML tree presented in Figure 6 estimated by the RELL method (Adachi and Hasegawa 1996) is only 0.3543. The sum of the BPs for the trees [Mse, (Hro, others)] is 0.6969, and for [Mse, ((Hro, Protostomes) (Xla, Ape))] it is 0.2575, whereas that for the trees in which deuterostomes (Halocynthia, Xenopus, and Asterina) appear as a monophyletic group is only 0.0086. It is concluded

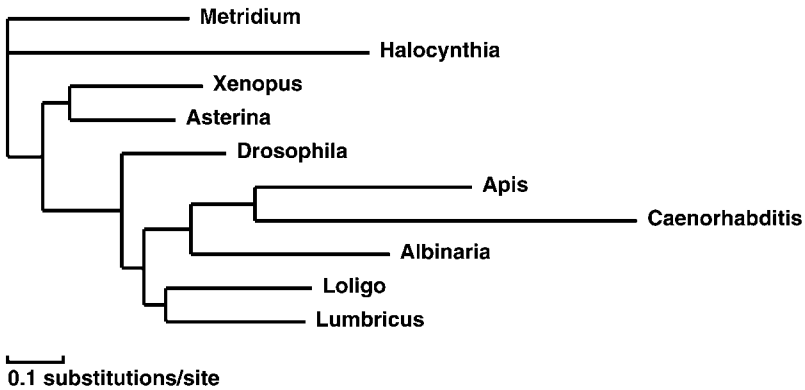

Figure 6.—ML tree based on the inferred amino acid sequences (2056 sites in total) from the nucleotide sequences of 12 mt-protein genes (all the mt-protein genes other than *ATPase8*). The tree was reconstructed by PROTML in MOLPHY (Adachi and Hasegawa 1996) with the mtREV-F and exhaustive search options.

from this analysis that the data support an unusual phylogenetic position for Halocynthia—namely, it is the most deeply branched species among all the triploblasticans analyzed in this work. However, the possibilities that Halocynthia is a sister taxon of protostomes (this is also very unusual in terms of the traditional view) or that it is in the deuterostome clade (this appears to fit with the widely accepted view of metazoan evolution) are not rejected by the analysis.

A phylogenetic analysis based on 18S rRNA failed to show the monophyly of chordates, since ascidians formed a group with hemichordates and echinoderms, but not with cephalochordates and vertebrates (Wada and Satoh 1994). Interestingly, the branch length of the ascidian line was much longer than those of the cephalochordate and vertebrate lines. Our analysis showed a similar higher substitution rate in urochordates than in vertebrates, in both the COIII tree (Figure 5) and the tree based on the combined data of 12 mt-protein genes (Figure 6). A comparison of the evolutionary patterns of both nuclear and mt-genes in vertebrates and urochordates (ascidians) could be an interesting line of investigation.

We are grateful to T. Oshima of Tokyo University of Pharmacy and Life Science for helpful comments. This work was supported by a Grant-in-Aid for Scientific Research on Priority Areas from the Ministry Education, Science, Sports and Culture of Japan, by a grant from the Human Frontier Science Program to K.W., and by grants from the Deutsche Forschungsgemeinschaft to S.P. S.Y. was supported by a fellowship from the Alexander-von-Humboldt Foundation.

LITERATURE CITED

Adachi, J., and M. Hasegawa, 1996 MOLPHY Version 2.3b3. Institute of Statistical Mathematics, Tokyo.

Anderson, S., A. T. Bankier, B. G. Barrell, M. H. L. de Bruijn, A. R. Coulson *et al.*, 1981 Sequence and organization of the human mitochondrial genome. Nature **290:** 457–465.

Asakawa, S., Y. Kumazawa, T. Araki, H. Himeno, K. Miura *et al.*, 1991 Strand-specific nucleotide composition bias in echinoderm and vertebrate mitochondrial genomes. J. Mol. Evol. **32:** 511–520.

Asakawa, S., H. Himeno, K. Miura and K. Watanabe, 1995 Nucleotide sequence and gene organization of the starfish *Asterina pectinifera* mitochondrial genome. Genetics **140:** 1047–1060.

Attardi, G., 1985 Animal mitochondrial DNA: an extreme example of genetic economy. Int. Rev. Cytol. **93:** 93–145.

Avise, J. H., 1994 *Molecular Markers, Natural History and Evolution*, Chapman & Hall, New York.

Beagley, C. T., R. Okimoto and D. R. Wolstenholme, 1998 The mitochondrial genome of the sea anemone *Metridium senile* (Cnidaria): introns, a paucity of tRNA genes, and a near-standard genetic code. Genetics **148:** 1091–1108.

Boore, J. L., and W. M. Brown, 1994 Complete DNA sequence of the mitochondrial genome of the black chiton, *Katharina tunicata*. Genetics **138:** 423–443.

Boore, J. L., and W. M. Brown, 1995 Complete sequence of the mitochondrial DNA of the annelid worm *Lumbricus terrertris*. Genetics **141:** 305–319.

Boore, J. L., T. M. Collins, D. Stanton, L. L. Daehler and W. M. Brown, 1995 Deducing the pattern of arthropod phylogeny from mitochondrial DNA rearrangements. Nature **376:** 163–165.

Börner, G. V., M. Mörl, A. Janke and S. Pääbo, 1996 RNA editing changes the identity of a mitochondrial tRNA in marsupials. EMBO J. **15:** 5949–5957.

Cantatore, P., M. N. Gadaleta, M. Roberti, C. Saccone and A. C. Wilson, 1987 Duplication and remoulding of tRNA genes during the evolutionary rearrangement of mitochondrial genomes. Nature **329:** 853–855.

Cantatore, P., M. Roberti, G. Rainardi, M. N. Gadaleta and C. Saccone, 1989 The complete nucleotide sequence, gene organization and genetic code of mitochondrial genome of *Paracentroutus lividus*. J. Biol. Chem. **264:** 10965–10975.

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.

Clayton, D. A., 1992 Replication and transcription of animal mitochondrial DNA. Int. Rev. Cytol. **141:** 217–232.

Crozier, R. H., and Y. C. Crozier, 1993 The mitochondrial genome of the honeybee *Apis mellifera*: complete sequence and genome organization. Genetics **133:** 97–117.

De Giorgi, C., A. Martiradonna, C. Lanave and C. Saccone, 1996 Complete sequence of the mitochondrial DNA in the sea urchin *Arbacia lixula*: conserved features of the echinoid mitochondrial genome. Mol. Phylogenet. Evol. **5:** 323–332.

Desjardins, P., and R. Morais, 1990 Sequence and organization of the chicken mitochondrial genome: a novel gene order in higher vertebrates. J. Mol. Biol. **212:** 599–634.

Durrheim, G. A., V. A. Corfield, E. H. Harley and M. H. Ricketts, 1993 Nucleotide sequence of cytochrome oxidase (subunit III) from the mitochondrion of the tunicate *Pyura stlonifera*: evidence that AGR encode glycine. Nucleic Acids Res. **21:** 3587–3588.

Felsenstein, J., 1995 PHYLIP (Phylogeny Inference Package) Version 3.57c. University of Washington, Seattle.

Flook, P. K., C. H. F. Rowell and G. Gellissen, 1995 The sequence, organisation and evolution of the *Locusta migratoria* mitochondrial genome. J. Mol. Evol. **41:** 928–941.

Green, R., and H. F. Noller, 1997 Ribosomes and translation. Annu. Rev. Biochem. **66:** 679–716.

Hatzoglou, E., G. C. Rodakis and R. Lecanidou, 1995 Complete sequence and gene organization of the mitochondrial genome of the land snail *Albinaria coerulea*. Genetics **140:** 1353–1366.

Himeno, H., H. Masaki, T. Kawai, T. Ohta, I. Kumagai *et al.*, 1987 Unusual genetic code and a novel gene structure for $tRNA^{Ser}_{AGY}$ in starfish mitochondrial DNA. Gene **56:** 219–230.

Hoffmann, R. J., J. L. Boore and W. M. Brown, 1992 A novel mitochondrial genome organization for the blue mussel, *Mytilus edulis*. Genetics **131:** 397–412.

Jacobs, H. T., D. J. Elliot, V. B. Math and A. Farquharson, 1988a Nucleotide sequence and gene organization of sea urchin mitochondrial DNA. J. Mol. Biol. **202:** 185–217.

Jacobs, H. T., P. Balfe, B. Cohen, A. Farquharson and L. Comito, 1988b Phylogenetic implications of genome rearrangement and sequence evolution in echinoderm mitochondrial DNA, pp. 121–137 in *Echinoderm Phylogeny and Evolutionary Biology*, edited by C. R. C. Paul and A. B. Smith. Clarendon Press, Oxford.

Janke, A., and S. Pääbo, 1993 Editing of a tRNA anticodon in marsupial mitochondria changes its codon recognition. Nucleic Acids Res. **21:** 1523–1525.

Janke, A., G. Feldmaier-Fuchs, W. K. Thomas, A. von Haeseler and S. Pääbo, 1994 The marsupial mitochondrial genome and the evolution of placental mammals. Genetics **137:** 243–256.

Janke, A., N. J. Gemmell, G. Feldmaier-Fuchs, A. von Haeseler and S. Pääbo, 1996 The mitochondrial genome of a monotreme—the platypus (*Ornithorhynchus anatinus*). J. Mol. Evol. **42:** 153–159.

Kyte, J., and R. F. Doolittle, 1982 A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. **157:** 105–132.

Kondo, A., S. Yokobori, T. Ueda and K. Watanabe, 1996 Primary sequence of ascidian mitochondrial glycine tRNA translating non-universal codon AGR (R:A,G). Nucleic Acids Res. Symp. Ser. **35:** 279–280.

Kondow, A., S. Yokobori, T. Ueda and K. Watanabe, 1998 Ascidian mitochondrial $tRNA^{Met}$ possessing unique structural characteristics. Nucleosides and Nucleotides **17:** 531–539.

Kondow, A., T. Suzuki, S. Yokobori, T. Ueda and K. Watanabe, 1999 An extra $tRNA^{Gly}$(U*CU) found in ascidian mitochondria responsible for decoding non-universal codons AGA / AGG as glycine. Nucleic Acids Res. **27:** 2554–2559.

Kumazawa, Y., T. Yokogawa, E. Hasegawa, K. Miura and K. Watanabe, 1989 The aminoacylation of structurally variant phenylal-

anine tRNAs from mitochondria and various nonmitochondrial sources by bovine mitochondrial phenylalanyl-tRNA synthetase. J. Biol. Chem. **264:** 13005–13011.

Kumazawa, Y., H. Himeno, K. Miura and K. Watanabe, 1991 Unilateral aminoacylation specificity between bovine mitochondria and eubacteria. J. Biochem. (Tokyo) **109:** 421–427.

Lee, W.-J., and T. D. Kocher, 1995 Complete sequence of a sea lamprey (*Petromyzon marinus*) mitochondrial genome: early establishment of the vertebrate genome organization. Genetics **139:** 873–887.

Mörl, M., M. Dörner and S. Pääbo, 1995 C to U editing and modifications during the maturation of the mitochondrial $tRNA^{Asp}$ in marsupials. Nucleic Acids Res. **23:** 3380–3384.

Ojala, D., J. Montoya and G. Attardi, 1981 tRNA punctuation model of RNA processing in human mitochondria. Nature **290:** 470–474.

Okimoto, R., H. M. Camberlin, J. L. MacFarlane and D. R. Wolstenholme, 1991 Repeated sequence sets in mitochondrial DNA molecules of root knot nematodes (*Meloidogyne*): nucleotide sequences, genome location and potential for host-race identification. Nucleic Acids Res. **19:** 1619–1626.

Okimoto, R., J. L. MacFarlane, D. O. Clary and D. R. Wolstenholme, 1992 The mitochondrial genomes of two nematodes, *Caenorhabditis elegans* and *Ascaris suum*. Genetics **130:** 471–498.

Okimoto, R., J. L. MacFarlane and D. R. Wolstenholme, 1994 The mitochondrial ribosomal RNA genes of the nematodes *Caenorhabditis elegans* and *Ascaris suum*: consensus secondary-structure models and conserved nucleotide sets for phylogenetic analysis. J. Mol. Evol. **39:** 598–613.

Osawa, S., 1995 *Evolution of the Genetic Code*. Academic Press, New York.

Pearse, V., J. Pearse, M. Buchsbaum and R. Buchsbaum, 1987 *Living Invertebrates*. Blackwell Scientific Publications, Palo Alto, CA.

Perna, N. T., and T. D. Kocher, 1995 Patterns of nucleotide composition at fourfold degenerate sites of animal mitochondrial genomes. J. Mol. Evol. **41:** 353–358.

Roe, B. A., D. P. Ma, R. K. Wilson and J. F.-H. Wong, 1985 The complete sequence of the *Xenopus laevis* mitochondrial genome. J. Biol. Chem. **260:** 9759–9774.

Saiki, R. K., D. H. Gelfand, S. Stoffel, S. J. Schorf, R. Higuchi *et al.*, 1988 Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science **239:** 487–491.

Saitou, N., and M. Nei, 1987 The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. **4:** 406–425.

Sambrook, J., E. F. Fritsch and T. Maniatis, 1989 *Molecular Cloning: A Laboratory Manual*, Ed. 2. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Sasuga, J., S. Yokobori, M. Kaifu, T. Ueda, K. Nishikawa *et al.*, 1999 The gene structure of mitochondrial DNA segment of the squid *Loligo bleekeri*. J. Mol. Evol. **48:** 692–702.

Steinberg, S., D. Gautheret and R. Cedergren, 1994 Fitting the structurally diverse animal mitochondrial $tRNAs^{Ser}$ to common three-dimensional constraints. J. Mol. Biol. **236:** 982–989.

Terrett, J. A., S. Miles and R. H. Thomas, 1996 Complete DNA sequence of the mitochondrial genome of *Cepaea nemoralis* (Gastropoda: Pulmonata). J. Mol. Evol. **42:** 160–168.

Thompson, J. D., D. G. Higgins and T. J. Gibson, 1994 CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. **22:** 4673–4680.

Thompson, J. D., T. J. Gibson, F. Plewniak, F. Jeanmougin and D. G. Higgins, 1997 The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. **25:** 4876–4882.

Valverde, J., B. Batuecas, C. Moratilla, R. Marco and R. Garesse, 1994 The complete mitochondrial DNA sequence of the crustacean *Artemia franciscana*. J. Mol. Evol. **39:** 400–408.

Wada, H., and N. Satoh, 1994 Details of the evolutionary history from invertebrates to vertebrates, as deduced from the sequences of 18S rDNA. Proc. Natl. Acad. Sci. USA. **91:** 1801–1804.

Wakita, K., Y. Watanabe, T. Yokogawa, Y. Kumazawa, S. Nakamura *et al.*, 1994 Higher structure of bovine mitochondrial $tRNA^{Phe}$ lacking the 'conserved' GG and TΨCG sequences as inferred by enzymatic and chemical probing. Nucleic Acids Res. **22:** 347–353.

Warrior, R., and J. Gall, 1985 The mitochondrial DNA of *Hydra attenuata* and *Hydra littoralis* consists of two linear molecules. Arch. Sc. Geneve **38:** 439–445.

Watanabe, K., and S. Osawa, 1996 tRNA sequence variations in the genetic code, pp. 225–250 in *tRNA: Structure, Biosynthesis and Function*, edited by D. Söll and U. L. RajBahandary. ASM Press, Washington DC.

Watanabe, Y., H. Tsurui, T. Ueda, R. Furushima, S. Tamiya *et al.*, 1994 Primary and higher-order structures of nematode (*Ascaris suum*) mt tRNAs lacking either the T or D stem. J. Biol. Chem. **269:** 22902–22906.

Wolstenholme, D. R., 1992 Animal mitochondrial DNA: structure and evolution. Int. Rev. Cytol. **141:** 173–216.

Yamazaki, N., R. Ueshima, J. A. Terrett, S. Yokobori, M. Kaifu *et al.*, 1997 Evolution of pulmonate gastropod mitochondrial genomes: comparisons of complete gene organizations of Euhadra, Cepaea and Albinaria and implications of unusual tRNA secondary structures. Genetics **145:** 749–758.

Yokobori, S. T. Ueda and K. Watanabe, 1993 Codons AGA and AGG are read as glycine in ascidian mitochondria. J. Mol. Evol. **36:** 1–8.

Yokogawa, T., Y. Watanabe, Y. Kumazawa, T. Ueda, I. Hirao *et al.*, 1991 A novel cloverleaf structure found in mammalian mitochondrial $tRNA^{Ser}$(UCN). Nucleic Acids Res. **19:** 6101–6105.

Communicating editor: N. Takahata