Complete mtDNA of *Ciona intestinalis* Reveals Extensive Gene Rearrangement and the Presence of an *atp8* and an Extra *trnM* Gene in Ascidians

Carmela Gissi, Fabio Iannelli, Graziano Pesole

Dipartimento di Scienze Biomolecolari e Biotecnologie, Università di Milano, Via Celoria 26, 20133 Milano, Italy

Abstract. The complete mitochondrial genome (mtDNA) of the model organism *Ciona intestinalis* (Urochordata, Ascidiacea) has been amplified by long-PCR using specific primers designed on putative mitochondrial transcripts identified from publicly available mitochondrial-like expressed sequence tags. The *C. intestinalis* mtDNA encodes 39 genes: 2 rRNAs, 13 subunits of the respiratory complexes, including ATPase subunit 8 (*atp8*), and 24 tRNAs, including 2 tRNA-Met with anticodons 5′-UAU-3′and 5′-CAU-3′, respectively. All genes are transcribed from the same strand. This gene content seems to be a common feature of ascidian mtDNAs, as we have verified the presence of a previously undetected *atp8* and of two *trnM* genes in the two other sequenced ascidian mtDNAs. Extensive gene rearrangement has been found in *C. intestinalis* with respect not only to the common Vertebrata/Cephalochordata/Hemichordata gene organization but also to other ascidian mtDNAs, including the cogeneric *Ciona savignyi*. Other features such as the absence of long noncoding regions, the shortness of rRNA genes, the low GC content (21.4%), and the absence of asymmetric base distribution between the two strands suggest that this genome is more similar to those of some protostomes than to deuterostomes.

Key words: ATP synthase subunit 8 — Gene duplication — Genome organization — Gene rearrangement — Mitochondrial genome — Tunicates — tRNA

Introduction

The phylum Chordata is divided into the three subphyla Urochordates, Cephalochordates, and Vertebrates, with Urochordata, the most basal lineage, made up of the classes Thaliacea, Appendicularia, and Ascidiacea. The subphylum Urochordata groups marine organisms with very different lifestyles and developmental programs. It includes sessil or planktonic, solitary or colonial, sexual or asexual reproducing animals, sharing the presence of a notochord either throughout the entire life cycle or only during the larval stage. Moreover, such organisms are unsegmented and headless. An understanding of the evolution and phylogenetic relationships between these ancient organisms—whose fossil records date back to the Lower Cambrian (550 million years ago) (Shu et al. 2001; Zhang 1987)—is crucial to provide insight into the early stages of chordata evolution, and into the phylogeny of deuterostomes (Cameron et al. 2000; Stach and Turbeville 2002; Swalla et al. 2000; Turbeville et al. 1994). Moreover, the ascidian *Ciona intestinalis* is widely used as model system for chordate developmental biology and for comparative genomic studies. The tadpole larva of ascidians represents one of the most simple and primitive chordate body plans. Furthermore, *C. intestinalis* embryos develop rapidly and can be easy manipulated (Satoh 1994). The small nuclear genome of *C. intestinalis*, about 160 Mbp, is used as a model to

Correspondence to: Graziano Pesole; *email:* graziano.pesole@unimi.it

investigate the nature of the primitive nonduplicated chordate nuclear genome (Holland et al. 1994; Ohno 1970). The sequencing and annotation of the coding portion of this genome have recently been completed and confirm that the hypothesized large-scale gene duplication event occurred in the vertebrate lineage after its divergence from cephalochordates and urochordates (Dehal et al. 2002). An ongoing cDNA project will complement the genome sequencing project, allowing an easier functional interpretation of the genomic sequences and providing valuable information on chordata gene regulation (Satou et al. 2002).

The animal mitochondrial genome (mtDNA) is a good model system to study evolutionary mechanisms at the genome level. This molecule is quite small (15–20 kb), easy to sequence, extremely compact, and shows a relatively constant gene content (Saccone et al. 1999). Complete mtDNA sequences of a large number of metazoans are available for comparative analysis. However, these are mostly from vertebrate taxa, and until the submission of our data, the complete mtDNA sequence was available for only one urochordate, the ascidian *Halocynthia roretzi* (order Pleurogona, suborder Stolidobranchiata, family Pyuridae) (Yokobori et al. 1999). The mtDNA sequence of *Ciona savignyi* (Yokobori, Watanabe, and Oshima, unpublished) was made available shortly after the submission of our sequence. There have long been indications that ascidians belonging to the same suborder as *Halocynthia* have a different mitochondrial gene organization (Durrheim et al. 1993; Jacobs et al. 1988) and that urochordate mtDNA evolves faster than that of other metazoans (Kakuda 2001; Stach and Turbeville 2002; Yokobori et al. 1999). Even at the nuclear level, a urochordate phylogeny reconstructed using a conserved portion of the 18S rRNA gene shows that the molecular divergence within urochordates is larger than in vertebrates and hemichordates, suggesting that Urochordata should be raised to the phylum status (Cameron et al. 2000; Swalla 2001; Swalla et al. 2000). In the light of these considerations, we decided to determine the complete mtDNA sequence of the model organism *C. intestinalis* (order Enterogona, suborder Phlebobranchiata, family Cionidae), an ascidian distantly related to *Halocynthia*, starting from the analysis of the large number of expressed sequence tags (ESTs) available for this organism in the public database (Gissi and Pesole 2003). The availability of the mtDNA sequence of *Ciona savignyi* (Yokobori, Watanabe, and Oshima, unpublished) has allowed a more detailed analysis of the evolution of ascidian mtDNA. The genus *Ciona* includes only solitary species, with *C. intestinalis* a native of northwestern Europe now showing a global distribution and *C. savignyi* a native of Japan which has

recently been found in southern California and the Pacific coast of America (Hoshino and Nishikawa 1985; Lambert 2001). The classification and number of species that constitute the genus *Ciona* remain quite controversial (Copello et al. 1981; Hoshino and Nishikawa 1985; Lambert et al. 1990).

In this paper, we describe the structural and compositional features of the *C. intestinalis* mtDNA, stressing the evolutionary trends of ascidian mtDNA compared to those of other metazoans, especially Chordata.

Materials and Methods

The *Ciona intestinalis* sample was caught in the Tirrenian Sea, Naples Gulf. Total DNA was extracted from the ovary of a single individual using the Puregene Tissue kit (Gentra Systems). The mtDNA was amplified by long-PCR (TaKaRa LA-Taq) in five overlapping fragments using specific primers designed according to the transcribed mitochondrial sequences of *Ciona intestinalis* inferred from mitochondrial-like ESTs (Gissi and Pesole 2003).

- Fragment sco (2.7 kb): colF1 5′-ATAGTTTGAGCAATGA GGGG 3′, smR 5′-TAACCGYGRTARCTGGCAC 3′
- Fragment n2-5 (3.6 kb): nd5F1 5′-GAAGAGTATATTAGGT GAAG-3′, nd2F 5′CTTGGGATATCATTCGAGAG-3′
- Fragment pg1 (2.6 kb): ap6R 5′-TTTCAACAGTTTTAC GAGG-3′, glR1 5′-AAATGAACTACTGA AAGGGC-3′
- Fragment cl-2 (2.7 kb); lg1F 5′-TTCAATTAGATTACCTT CAT-3′, lg2R 5′-CAATACCAAATCGCACACTC-3′
- Fragment cog (6.5 kb): g2F1 5′-ATGATAGTACGTTTTA CCCC-3′, co1R 5′-CTATATGCTCATAAAAGTGG-3′

The amplified fragments were directly cloned in the pCR 2.1-TOPO plasmid using the TOPO-TA Cloning kit (Invitrogen) following the manufacturer's protocol. Recombinant plasmids were extracted from positive clones using the QIAprep Spin Miniprep kit (Qiagen) and the inserts were completely sequenced by primer walking. Sequencing was performed on an ABI Prism 3100 sequencer (Perkin Elmer). The overlaps between amplified fragments range from 170 bp (overlap between fragment cog and sco) to 1.2 kb (overlap between fragment n2-5 and pg1). In case of sequence uncertainties, the corresponding region of the mtDNA was reamplified from the total DNA and directly sequenced without cloning.

Transfer RNA genes were identified using tRNAscan-SE (version 1.21) with parameters specific for organellar tRNA structures (Lowe and Eddy 1997), by searching for canonical anticodon arm sequences (Kumazawa and Nishida 1993) using the PatSearch program (Pesole et al. 2000), and by sequence potential to be folded into cloverleaf structures. Small and large subunit ribosomal rRNA genes (*rrnS* and *rrnL*) were identified by sequence similarity to homologous animal genes. Ends of rRNA genes were not determined experimentally and were assumed to be adjacent to the ends of bordering tRNA genes. As an exception, the 5′ end of *rrnL* was tentatively defined by sequence similarity to homologous genes of the ascidians *Halocynthia roretzi* (AB024528) and *Ciona savignyi* (AB079784), as described in Results and Discussion. Open reading frames of protein genes were defined using the ascidian mitochondrial translation code (Durrheim et al. 1993; Yokobori et al. 1993) and identified by similarity to the homologous mitochondrial-encoded proteins of the following 16 species representative of deuterostome organisms: *Homo sapiens* (AF347015), *Oryctolagus cuniculus* (AJ001588), *Gallus gallus* (X52392), *Alligator mississip-*

piensis (Y13113), *Chrysemys picta* (AF069423), *Dinodon semicarinatus* (AB008539), *Rana nigromaculata* (AB043889), *Squalus acanthias* (Y18134), *Gadus morhua* (X99772), *Latimeria chalumnae* (U82228), *Petromyzon marinus* (U11880), *Myxine glutinosa* (AJ404477), *Branchiostoma lanceolatum* (Y16474), *Halocynthia roretzi* (AB024528), *Ciona savignyi* (AB079784), *Balanoglossus carnosus* (AF051097), and *Paracentrotus lividus* (J04815). The possibility of unusual initiation and termination codons was considered (Wolstenholme 1992). The gene encoding ATPase subunit 8 (*atp8*) was identified by comparison against homologous sequences of a large number of species, including representatives of protostomes, cnidarians, porifers, and the yeast *Saccharomyces cerevisiae* (alignment available on request). Moreover, the presence and location of a transmembrane hydrophobic helix in ATP8 protein were investigated using the DAS software (Cserzo et al. 1997). The definition of gene boundaries was guided by the mapping of mitochondrial-like *Ciona* ESTs, as described in Gissi and Pesole (2003).

Specific nucleotide patterns, such as the rRNA transcription termination box, were searched using the PatSearch program (Pesole et al. 2000). Repeated sequences were identified using RepFinder (Benson 1999) and CleanUP (Grillo et al. 1996) programs.

Base compositional analyses were carried out using the Codontree algorithm (Pesole et al. 1996). The GC- and AT-skews, which indicate compositional differences between the two strands, were calculated according to the formulae of Perna and Kocher (1995).

Mitochondrial genome rearrangements within deuterostomes and in all one-strand transcribed mtDNAs were investigated using the minimal breakpoint method, which reconstructs the most parsimonious phylogeny from breakpoint distances (Blanchette et al. 1999).

Blast search analyses were carried out at the NCBI Web site (http://www.ncbi.nlm.nih.gov/BLAST/).

The complete mtDNA sequence of *Ciona intestinalis* is available at the Embl database under accession number AJ517314.

Results and Discussion

Amplification Strategy

The mtDNA of *C. intestinalis* has been completely amplified from total DNA in five overlapping fragments by taking advantage of the numerous ESTs of this organism available in the Embl database to design PCR primers. As described in Gissi and Pesole (2003), mitochondrial-like ESTs of *C. intestinalis* have been assembled based on the presence of overlapping regions and on clone information. Eight distinct nonoverlapping mitochondrial transcripts have been inferred, containing typical mitochondrial genes and covering on the whole about 12 kb. Their sequences have been used to design several specific primers that have been used in different combinations to amplify the complete mtDNA. The locations of inferred mitochondrial transcripts and amplified fragments on the mtDNA are reported in Fig. 1. All inferred mitochondrial transcripts have been confirmed at the level of both gene order and sequence (>99% identity).

General mtDNA Features

The *Ciona intestinalis* mtDNA is 14,790 bp long, very similar in size to other ascidian mtDNAs. Among animals, the smallest sequenced mtDNA belongs to *Taenia crassiceps* (Nematoda, 13.5 kb) and the largest to *Venerupis philippinarum* (Mollusca, 22.6 kb) (Embl database, May 2003). The ascidians genomes are the smallest among the completely sequenced mtDNAs of Chordata, with a size closer to that of Echinodermata and Hemichordata (range from 15.6 to 16.2 kb) than to other Chordata, which range from the 15 kb of Cephalochordata to the 18.9 kb of the Agnatha *Myxine glutinosa*. The absence of a long noncoding region homologous to the mammalian D-loop-containing region, the presence of small intergenic regions, and the relative shortness of some genes (i.e., rRNAs) make the *Ciona* mtDNA more similar in size to those of invertebrates. Moreover, the genome is highly compact; indeed the coding portion forms 96.9% of the *C. intestinalis* mtDNA.

We were surprised to observe that the *C. intestinalis* mtDNA encodes for 39 genes: the standard set of 37 metazoan mitochondrial genes and 2 additional tRNA genes. The 13 genes coding for proteins of the respiratory complexes have all been identified, including a gene coding for a short form of the F0 ATPase subunit 8 not previously annotated in the closely related *C. savignyi* and *H. roretzi* mtDNAs (Yokobori et al. 1999). Indeed, we found this gene in both *H. roretzi* and *C. savignyi* mtDNAs (see below). In addition to the two genes coding for ribosomal RNAs (*rrnS* and *rrnL*), there are 24 tRNA genes: 2 tRNA genes for decoding Leu, Ser, Gly, and Met codons and 1 tRNA for decoding the remaining amino acids. The existence of two tRNA genes for Leu, Ser, and Gly codons is in accordance with the alteration of the mitochondrial genetic code in ascidians compared to the universal one (Durrheim et al. 1993; Yokobori et al. 1999). Furthermore, two distinct tRNA-Met genes can be identified in *C. intestinalis*, the first possesses the common 5′-CAU-3′ anticodon, *trnM(AUG)*, while the second exhibits the unusual 5′-UAU-3′ anticodon, *trnM(AUA)*. The same situation has been reported for *C. savignyi* (AB079784). In *H. roretzi* only the unusual *trnM(AUA)* has been reported (Yokobori et al. 1999), but an unannotated *trnM(AUG)* overlapping the end of the *nad4l* gene can easily be identified (see below).

Genome Organization

All mitochondrial genes of *C. intestinalis* are encoded on the same strand, as in *H. roretzi* and *C. savignyi*. In the deuterostome mtDNAs completely sequenced, most of the genes are transcribed from one strand,
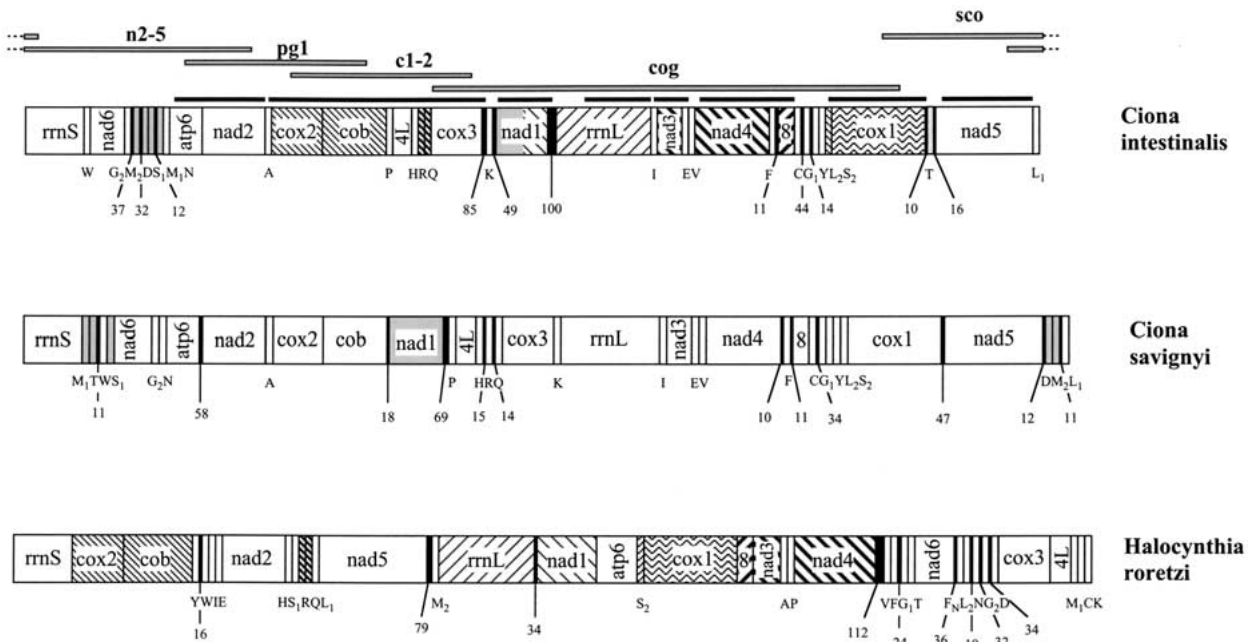
Fig. 1. Amplification strategy and genome structure of *C. intestinalis* mtDNA, compared to those of *Ciona savignyi* (AB079784) and *Halocynthia roretzi* (AB024528). Putative *C. intestinalis* transcripts inferred from EST data (Gissi and Pesole 2003) are indicated by black lines and amplified segments as dark gray boxes. Noncoding regions longer than 9 bp are indicated in black with numbers corresponding to their size (bp). Genes blocks with different positions in *C. intestinalis* and *C. savignyi* are indicated as light gray boxes; gene blocks conserved in *C. intestinalis* and *H.* *roretzi* are indicated by hatched boxes, with reverse hatch for inverted gene order. Feature table of *H. roretzi* has been modified as reported in Gissi and Pesole (2003). Abbreviations for proteins and rRNA genes are as in the text, except for 4L (*nad4L*) and 8 (*atp8*). tRNA genes are indicated according to the transported amino acid and, in addition: F_N, additional tRNA-Phe; G1, Gly(AGR); G2, Gly(GGN); L1, Leu(UUR); L2, Leu(CUN); M1, Met(AUG); M2, Met(AUA); S1, Ser(AGY); S2, Ser(UCN).

whereas a small number of tRNA genes and at least *nad6* are encoded by the complementary strand (Boore 1999; Saccone et al. 1999). The transcription of all mitochondrial genes from the same strand has been found only in ascidians and lower animals, that is, in Cnidaria Zoantharia and some protostomes (Nematoda, except for *Trichinella spiralis*, Platyhelminthes, Brachiopoda, Annelida, and a few representatives of Crustacea and Mollusca) (see Appendix, Table A1). Gene order similarities between these one-strand transcribed mtDNAs have been calculated as breakpoint distances with and without tRNA genes (data not shown), showing that the genome organization is quite conserved only in phylogenetically related organisms with no trace of evolutionary convergence. Thus, no peculiar constraints seem to act on the one-strand transcribed mtDNAs forcing them to adopt a specific gene order.

The gene arrangement of *C. intestinalis* is markedly different from that reported for *H. roretzi*, with which it shares only three conserved gene blocks (Fig. 1): *cox2–cob*, *trnR–trnQ*, and *trnS(UCN)–cox1*. The *nad1* and *rrnL* genes are adjacent in both species, however, the gene order is *nad1–rrnL* in *C. intestinalis* and *rrnL–nad1* in *H. roretzi*. Excluding tRNA genes, a fourth gene block appears to be conserved between the two mtDNAs, that is, the *nad3–nad4*
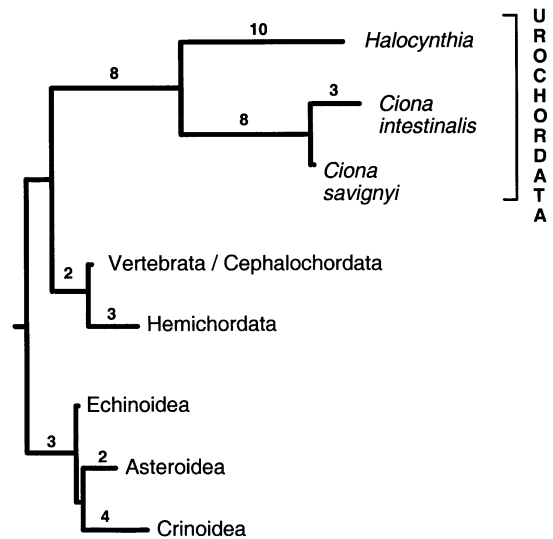
Fig. 2. Minimal breakpoint tree reconstructed from mitochondrial arrangement of Deuterostomes, excluding tRNA genes. Numbers indicate breakpoint distance values. Data as follows: Echinoidea—*Paracentrotus lividus* (J04815), *Strongylocentrotus purpuratus* (X12631), *Arbacia lixula* (X80396); Crinoidea—*Florometra serratissima* (AF049132); Asteroidea—*Asterina pectinifera* (D16387); *Ciona savignyi* (AB079784); *Halocynthia roretzi* (AB024528); Vertebrata/Cephalochordata—general Vertebrata and Cephalochordata mtDNA gene organization (data from Embl, May 2003); Hemichordata—gene organization of hemichordate *Balanoglossus carnosus* (AF051097) and vertebrate Teleostei *Myctophum affine* (AP002922) and *Conger miriaster* (AB038381).

Table 1. Base composition of ascidian mtDNAs calculated separately for the whole genome (Tot), protein-coding genes (CDS), first (P1), second (P2) and third (P3) codon positions, third position of quartet codons (P3Q), and noncoding regions (NC)

	Ciona intestinalis						Ciona savignyi						Halocynthia roretzi					
	%A	%C	%G	%T	AT-skew	GC-skew	%A	%C	%G	%T	AT-skew	GC-skew	%A	%C	%G	%T	AT-skew	GC-skew
Tot	34.2	9.5	11.9	44.4	−0.13	0.12	32.0	8.7	14.1	45.3	−0.17	0.24	24.3	8.5	23.2	44.0	−0.29	0.46
CDS	32.0	9.9	12.1	46.0	−0.18	0.10	29.6	9.0	14.4	47.0	−0.23	0.23	22.4	8.4	23.4	45.7	−0.34	0.47
P1	36.3	10.1	14.9	38.7	−0.03	0.19	34.6	9.2	17.3	38.9	−0.06	0.30	24.7	9.6	26.9	38.8	−0.22	0.47
P2	19.4	15.2	15.4	50.1	−0.44	0.01	19.3	14.9	15.5	50.3	−0.45	0.02	16.6	12.5	22.4	48.5	−0.49	0.28
P3	40.2	4.4	6.0	49.3	−0.10	0.16	35.1	2.8	10.5	51.6	−0.19	0.58	26.0	3.1	21.0	49.9	−0.31	0.74
P3Q	33.4	4.9	5.4	56.3	−0.26	0.06	30.5	2.4	8.9	58.2	−0.31	0.57	26.0	3.6	17.3	53.1	−0.34	0.66
NC	43.6	4.5	9.4	42.5	0.01	0.35	32.5	5.9	12.6	49.0	−0.20	0.36	24.3	4.7	24.9	46.0	−0.31	0.68

gene pair, with the *atp8* located upstream in *H. roretzi* and downstream in *C. intestinalis*. Furthermore, gene order is not conserved even within the genus *Ciona* (Fig. 1). Comparison of *C. intestinalis* and *C. savignyi* mtDNA organization reveals the transposition of *nad1* and five tRNA genes—*trnD*, *trnM(AUA)*, *trnM(AUG)*, *trnS(AGY)*, and *trnT*—with *trnD* and *trnM(AUA)* adjacent but inverted in the two species (Fig. 1). There are few cases of mtDNA completely sequenced in different species of the same genus and the only other reported case of intragenus variability in mt gene order has been found in the Platyhelminthes genus *Schistosoma*, which shows a high rate of gene rearrangement (Le et al. 2000). Hence, at least for invertebrates, gene rearrangements within genus may be more common than expected.

In addition to intraclass genome rearrangements, the ascidian gene order is extremely divergent from that of other Chordata. Indeed, Vertebrata, Cephalochordata, and even the hemichordate *Balanoglossus carnosus* share an almost-identical basal mt genome organization, with only a few transpositions tolerated, mostly involving tRNA genes and noncoding control region (Saccone et al. 1999). The only gene pair conserved among all Chordata is *rrnL–nad1* (not taking into account the tRNA genes) but is missing in Cionidae. Indeed, this gene pair is inverted in *C. intestinalis* and absent in *C. savignyi* (Fig. 1). Figure 2 reports the minimal breakpoint tree reconstructed from deuterostome mtDNA gene order data without considering tRNA genes, owing to their high mobility (Blanchette et al. 1999). The breakpoint distances within the ascidian clade are very high compared to those within the Echinodermata and Vertebrata–Cephalochordata–Hemichordata clades. The total breakpoint distance from the common deuterostome ancestor is 29 for ascidians and 5 for the clade including Vertebrata–Cephalochordata–Hemichordata. This implies a fivefold higher gene rearrangement rate in the former with respect to the latter organisms. Such observations confirm that the ascidian mtDNA has a very high gene rearrangement

rate, previously supposed to be a feature of only protostome mt genomes (Boore 1999; Saccone et al. 1999). A more surprising feature of the ascidian mtDNAs is that even genes usually adjacent in metazoan mtDNAs, such as *rrnS–rrnL* and *atp8–atp6*, are rearranged and distantly located. The *rrnS* and *rrnL* are very close and separated by few tRNA genes in most animal mtDNAs, whereas they are 7 and 5 kb distant in *Ciona* and *Halocynthia*, respectively. Ascidiacea share this gene organization only with Echinoidea (Echinodermata) among deuterostomes, most Cnidaria, and some protostomes (see Appendix, Table A2). This peculiar gene arrangement is interesting because it prevents a coordinated expression of both rRNAs by the synthesis of a single transcript. Indeed, in mammals the rRNA genes are transcribed at a higher level compared to other mt genes through the usage of a transcription termination signal located downstream from the two adjacent rRNA genes (Christianson and Clayton 1986). Similarly, *atp8* immediately precedes and sometimes overlaps *atp6* in most metazoans and even in *S. cerevisiae* (Foury et al. 1998). In mammals the two proteins are translated from the same mature bicistronic transcript, allowing a translational regulation (Fearnley and Walker 1986). In Ascidiacea these genes are 3 or 8.4 kb distant, a situation found in only a few other species (see Appendix, Table A3). It is possible that the different gene order could have important implications for the regulation of gene expression.

Base Composition and Codon Usage

The G + C content of the *C. intestinalis* mtDNA sense strand is 21.4%. According to their percentage, the relative usage of bases is T > A > G > C, with very low use of C (about 9.5%). This trend is conserved both in highly constrained sites (first and second codon positions, P1 and P2, respectively) and in less constrained sites (third codon positions, P3; third position of quartet codons, P3Q; non-

Fig. 3. Secondary structure of the hypothesized tRNA-Met(AUG) of *H. roretzi* (locations 5838–5913 on AB024528 sequence) and *C. savignyi* (AB079784). Canonical and G–T base pairs are differently indicated.

coding sequences, NC) (see Table 1). A similar compositional pattern is present in the *C. savignyi* and *Halocynthia* mt genomes. The asymmetric distribution of bases between the two strands (AT- and GC-skew) is very low in *C. intestinalis* mtDNA (Table 1). Indeed, although C is the lowest represented nucleotide in most sites, G is almost as rare as C in P2, P3, and P3Q. However, an asymmetric base distribution can be observed in *C. savignyi* and, particularly, in *Halocynthia*, which show a positive GC-skew and a negative AT-skew (Table 1).

The bias in base composition toward AT and against GC affects both the codon usage pattern and the amino acid composition of proteins. Indeed, codons GGC (Gly) and CGC (Arg) are not used at all in *C. intestinalis* mtDNA and are among the lowest used codons in other ascidians (data not shown).

Overall, the compositional features of ascidian mtDNA reflect the basal positions of these organisms in the phylum, being closer to those of Cephalochordata than to Vertebrata. The GC% is the lowest among all sequenced mtDNAs of Chordata. The sense strand is C-poor as in Cephalochordata and unlike in Vertebrata, where the sense strand is G-poor (Saccone et al. 1999; Yokobori et al. 1999). AT- and GC-skew are quite variable, being very low in some species and remarkable in others. Moreover, in the case of an asymmetric base distribution between the two strands, the ascidian GC-skew is positive as in Cephalochordata, therefore it is inverted compared to mammals and other vertebrates (negative GC-skew) (Reyes et al. 1998; Saccone et al. 1999).

Protein Genes

The short open reading frame located between *trnF* and *trnC* in *C. intestinalis* mtDNA has been identified as encoding the F0-ATPase subunit 8 (35 amino acids long). A similar ORF is present in the same position in *C. savignyi* mtDNA (46% amino acid similarity) and has been found in the region between *cox1* and *nad3* in *H. roretzi*, which was originally reported as noncoding (33% amino acid similarity) (Yokobori et al. 1999). The ATP8 protein is quite heterogeneous in length between organisms; it is 48 aa long in *S. cerevisiae* and longer in *Drosophila* (52 aa), human (68 aa), and cnidarian (72 aa). The *orfB* gene of plants and protists, which is the *atp8* homologue, is even longer (Gray et al. 1998). A peculiarity of this protein is its conservation more at the level of secondary structure and chemical character of amino acids than in the specific amino acid identity (Papakonstantinou et al. 1996b). In yeast, where it has been widely studied, ATP8 is an intrinsic membrane protein composed of three domains: an N-terminal domain located in the intermembrane space and showing a conserved MPQL motif, a central membrane-spanning hydrophobic domain, and a C-terminal positively charged region exposed in the matrix space. The N-terminal motif plays a role in ATPase activity (Devenish et al. 1992), whereas the positively charged amino acids at the C terminal are involved in both assembly and function of the F0 sector (Papakonstantinou et al. 1993, 1996a). The short ORFs identified in the three ascidian mtDNAs all have a MPQL motif at the N terminal, a transmembrane-spanning domain, as predicted by the DAS software, and one (in *C. intestinalis*) or two (in *C. savignyi* and *H. roretzi*) positively charged amino acids in the C-terminal domain. Thus, despite the low global similarity to other ATP8, such ORFs can be unambiguously regarded as an ATP8 form with short peripheral domains, as reported in the alignment in Fig. 4. In yeast, several lines of evidence indicate that ATP8 is an integral component of the ATPase stator stalk, interacting by means of both peripheral domains with remaining F0 sector subunits and performing a primarily structural role (Stephens et al. 2003). The different number and organization of the F0 subunits in the ATPase of different organisms might explain the high divergence of this protein in various species and thus its peculiarity in ascidians.

ATG is the most common initiation codon, with only *nad5* and *nad4L* genes starting with ATT, *nad2* with ATC, and *nad4* with ATA. In accordance with the low G% of the genome, codons ending with G might be expected to be avoided. However, this rule appears not to be valid for start codons, suggesting that there is a positive selection for the choice of ATG

```
            . . . .10 . . . .20 . . . .30 . . . .40 . . . .50 . . . .60 . . . .70 . . . .80
S_cerevisiae   ...MPQLVPFYFMNQLT.YGFLLMILLLILFSQFFLPMI...LRLYVSRLFISKL................................ 48
C_intestinalis ...MPQLNL.........FSFFYISVVSFSLLMTSFSY....EPILVPVSR................................... 35
C_savignyi     ...MPQLNM.........FSFFLYCVNSFDYVNYFFSI....RPNHLSLN.................................... 34
Halocynthia    MSSVPQING........MSFFFFFLLFWVIYLFLVSS....GGDMLGRKDL.................................. 39
Homo           ...MPQLNT.........TVWPTMITPMLLTLFLITQL....KMLNTNYHLPPSPKPMKMKNYNKPWEPKWTKICSLHSLPPQS 68
Balanoglossus  ...MPQLNPDFWALN...FVTVWLLLAIALATTTSLQLPLQ.PSPSSLQHHTLSTPQWLWT....................... 54
Paracentrotus  ...MPQLD.........FTWWIINFFIIWTAILLTLVILV.NNKTAQNLTTTDSLQIEKNSTNWQWL.................. 54
Drosophila     ...MPQMAP.........ISWLLLFIIFSITFILFCSINYY.SYMPNSPKSNELKNINLNSMNWKW.................. 53
Metridium      ..MMPQLETATYLTQ...YSWTLIALFLLFSFLVVSV.....LPAVKTNFLISSSIGAGWTGAPKTSDLNKGPASLWSWDKI.. 72
Tetilla        ...MPQLEAVTFLCQ...YMWKLVILFVLFSMLVNVMLPS.LQWQIVTSNQVNSMEMKKESIKLETILMM.............. 63
```

Trans-membrane domain

Fig. 4. Alignment of ATP8 amino acid sequences from yeast *Saccharomyces cerevisiae* (AJ011856), mammal *Homo sapiens* (V00662), *Ciona savignyi* (AB079784), *Halocynthia roretzi* (AB024528), hemichordate *Balanoglossus carnosus* (AF051097), echinoderm *Paracentrotus lividus* (J04815), insect *Drosophila melanogaster* (U37541), cnidarian *Metridium senile* (AF000023), and sponge *Tetilla* sp. (AF035265). Highly conserved amino acids are indicated in bold face. Positively charged amino acids are shaded gray.

as initiator codon. Nine protein genes terminate with a TAA stop codon, and only one (*nad4L*) with a TAG codon. Partial stop codons that are completed by transcript polyadenylation, as reported in mammals, have been hypothesized for three genes: *cob*, *atp8* (TA incomplete stop codon), and *nad5* (T incomplete stop codon). The identification of stop codons has been assisted by mapping mt-like ESTs of *C. intestinalis* onto the mtDNA; indeed EST data allowed the identification of poly(A) stretches representing the poly(A) tails of the primary or mature mitochondrial transcripts (Gissi and Pesole 2003). Based on genomic sequence data, the *cob* frame stops with a TAG codon inside the downstream *trnP* gene, however, based on EST inferred transcripts a poly(A) tail starts 9 bp upstream of the TAG codon (Gissi and Pesole 2003). Thus a TAA stop codon should be created from an incomplete TA codon by polyadenylation of the transcript, excluding the existence of an overlap between the *cob* and *trnP* genes. A similar situation is found in *nad5*, where polyadenylation eliminates the possibility of a 5-nt overlap with the downstream *trnL(UUR)*, and in *atp8* (avoided overlap of one nucleotide with *trnC*). In addition, EST mapping suggests that *nad4L* could use the complete TAG stop codon or an incomplete T/TA codon (Gissi and Pesole 2003). As suggested by Boore (2001), the coupling of incomplete stop codons to complete in-frame stop codons requiring a few-nucleotide overlap to the downstream gene could prevent translational readthrough in case of incorrect transcript maturation.

The size of the identified protein genes is similar to those of their homologues in *H. roretzi* and *C. savignyi*, with few exceptions (data not shown). A 25-amino acid extension at the C-terminal of the *Halocynthia* NAD4L protein has not been found in any of the aligned deuterostome homologues (see Materials and Methods), and a detailed analysis revealed the presence of a tRNA-Met(AUG) sec-ondary structure entirely included in this extra region. Such putative tRNA exhibits a normal cloverleaf secondary structure except for an extremely long TΨC stem (7 bp) (Fig. 3). An overlap of 13 bp is present between these two genes even assuming a *nad4L* incomplete termination codon. It is unclear whether this tRNA gene is functional or is a pseudogene of recent origin that has become part of a protein reading frame. A similar situation is found in the sea urchin mtDNA, where a *trnL(CUN)* pseudogene forms the amino end of the NAD5 protein (Cantatore et al. 1987). It is striking that the *trnM(AUG)* gene is among the genes transposed between *Ciona* mtDNAs.

Gene overlaps were found in all cases of adjacent protein-coding genes with no tRNA genes in between, that is, for the gene pairs *cox2–cob* and *atp6–nad2*. The estimated overlaps are 26 and 10 nt, respectively. The overlap between *cox2* and *cob* is observed also in *H. roretzi* and *C. savignyi*. The *nad2* reading frame overlaps the *atp6* gene for 46 nt, but protein similarity data support a shorter overlap of only 10 nt (data not shown). The identification of bicistronic transcripts in the putative mitochondrial transcripts of *C. intestinalis* inferred from the EST data further support the existence of both overlaps (Gissi and Pesole 2003).

tRNA Genes

Proposed secondary structures for the 24 identified tRNAs are shown in Fig. 5. The genes vary in size from 52 (*trnC*) to 70 (*trnE* and *trnD*) nt. In all tRNAs a T precedes the anticodon and a purine, mostly A, follows it, except in tRNA-Asp, where the anticodon is followed by a T. Each tRNA has been inferred to have an aminoacyl acceptor stem of 7 nt and an anticodon stem of 5 nt, except for tRNA-Ser(AGY), where the anticodon stem is 7 nt long. There are no unpaired nucleotides in the acceptor
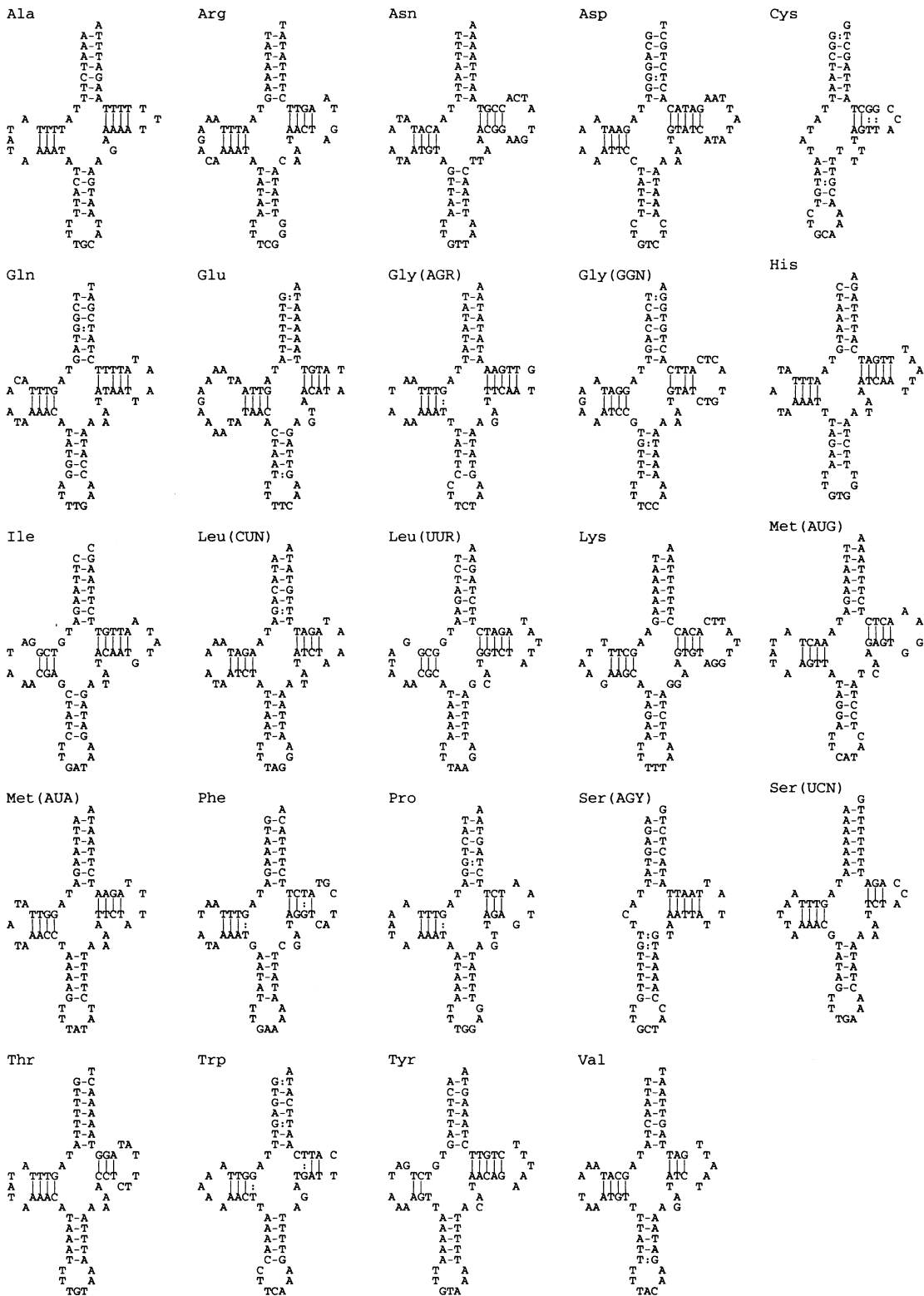
Fig. 5. Putative secondary structures of the 24 tRNAs encoded by the *C. intestinalis* mtDNA. Canonical and G–T base pairs are differently indicated.

stems, although one mismatched base pair is present in the anticodon stems of tRNA-Phe and tRNA-Gly(AGR). The DHU and TΨC arms vary greatly in size and sequence. The stems are 3, 4, or 5 nt long in both arms. The TΨC loop size is highly variable between different tRNAs ranging from 2 to 9 nt, whereas the DHU loop is 5 or 6 nt long in most tRNAs except for tRNA-Glu (13 nt long).

Table 2. Length (bp) of the small (*rrnS*) and large (*rrnL*) subunit rRNA genes of ascidians and a few other metazoan representatives

	rrnS	*rrnL*
M. musculus	955	1581
Agnatha	859–903	1541–1621
Cephalochordata	844–846	1367–1371
C. intestinalis	711	1100
C. savignyi	687	1097
H. roretzi	706	1113
Hemichordata	815	1498
Echinodermata	848–897	1511–1554
D. melanogaster	789	1325
Nematoda	672–700	947–987
Platyhelminthes	704–766	963–1055

Note. Species list is reported in Table A4 (Appendix).

Only tRNA-Ser(AGY) and tRNA-Cys lack the DHU arm, which is replaced by loops 4 and 5 nt long, respectively. The variable loop is 2, 3, or 4 nt long and the size of spacers between arms is in accordance with the model of animal mitochondrial tRNAs reported by Kumazawa and Nishida (1993). The tRNA secondary structure lacking the DHU arm is typical of metazoan tRNA-Ser(AGY) (Kumazawa and Nishida 1993), whereas in tRNA-Cys this structure has been found only in Cephalochordata (Boore et al. 1999; Spruyt et al. 1998) and in some vertebrates, where it has been correlated with the loss of the immediately adjacent stem–loop structure of the L-strand replication origin (Macey et al. 1999; Seutin et al. 1994; Zardoya and Meyer 1996). Among ascidians, the tRNA-Cys has a normal cloverleaf secondary structure in *Halocynthia* (Yokobori et al. 1999) and possesses a DHU replacement loop 13 bp long in *C. savignyi*.

The secondary structures of the two tRNA-Met identified in *C. intestinalis* are compatible with their functionality (Fig. 5). Most metazoan mtDNAs possess a single *trnM* gene with a 5′-CAU-3′ anticodon able to decode methionine codons, ATG or ATR, depending on the genetic code, without discriminating initial and internal methionines (Yokobori et al. 2001). Only in the mollusks *Mytilus edulis* and *M. californianus* has an additional gene for *trnM*(AUA) been found and the transcription product identified (Beagley et al. 1999; Hoffmann et al. 1992). In addition, a tRNA-like structure with an anticodon 5′-UAU-3′ has been found in the nematode *Trichinella spiralis*, but its unusual secondary structure suggests that it may be not functional (Lavrov and Brown 2001). In *Mytilus* it has been suggested that tRNA-Met(AUG) may be used as initiator, preferentially recognizing the ATG codon, whereas tRNA-Met(AUA) may be used as an elongator, although there is no definitive evidence to support this hypothesis (Beagley et al. 1999). This hypothesis could also be valid for ascidians. Indeed, in *C. intestinalis*, the initiator codon is ATG in 9 of the 13 mt-protein genes. Moreover, tRNA-Met(AUG) presents two consecutive G–C base pairs close to the bottom of the anticodon. Three consecutive G–C base pairs at the base of the anticodon stem are essential for the function of the *E. coli* tRNA-Met as initiator (Rajbhandary and Chow 1995) and are conserved in the mitochondrial tRNA-Met(AUG) of most Metazoa (Sprinzl et al. 1998). Considering the low GC% in *C. intestinalis* mitochondrial tRNAs, the conservation of such feature is even more significant. To shed light on the role performed by the two tRNA-Met, it will be important to define their decoding capacity. The two tRNA-Met could potentially recognize both ATR codons. Indeed, the tRNA-Met(AUA) of *H. roretzi* has U-derivative at the wobble position, allowing the decoding of both ATR codons (Kondow et al. 1998). In metazoans, tRNA-Met(AUG) decodes both ATR codons when there is a 5-formylcytidine at the wobble position of the anticodon or an A-derivative at the position 3′ adjacent to the anticodon, while it recognizes only the AUG codon when an unmodified nucleotide is present at both positions (Tomita et al. 1999). Also *C. savignyi* and *Halocynthia* tRNA-Met (AUG) possess two consecutive G–C base pairs in the anticodon stem (Fig. 3). It would be interesting to verify the existence and functionality of both tRNA-Met in a larger number of ascidian species.

It is striking that among deuterostome mt genetic codes, AUA codes for Met in Chordata but for Ile in Echinodermata and Hemichordata, and at least in the genus *Ciona* there is an additional tRNA able to decode such codon. To verify the meaning of AUA codons in the ascidian mt genetic code, we looked for sites with 100% conserved Met or Ile in the alignments of the most conserved mitochondrial genes (*cox1*, *cox2*, *cox3*, and *cob*), which include species as distant as *C. elegans* and human (20 representative species plus ascidians). There are 10 Met conserved sites in the analyzed alignments, and in all of them at least one ascidian species presents an AUA codon, demonstrating that the AUA codon is used in sites where methionine is conserved over long evolutionary distances. On the contrary, there are no Ile conserved sites with an AUA codon in ascidians. These observations unambiguously demonstrate that AUA codes for Met in ascidians.

In *M. edulis*, similarities at the 5′ end of the two *trnM* genes suggest that they arose by gene duplication (Hoffmann et al. 1992). In ascidians, the low similarity between the two *trnM* genes of the same species seems to exclude the origin of the two genes by recent gene duplication of a common ancestor.

rrnS

```
C_intestinalis (AJ517314)    152  AACGTGCCAGCTACCGCGGTTATAC   176
C_savignyi (AB079784)       3703  AAGGTGCCAGCTACCGCGGTTATAC  3727
Halocynthia (AB024528)      6208  TGGGTGCCAGTAATCACGGTTAGAC  6232
H_sapiens (J01415)           884  TTCGTGCCAGCCACCGGGTCACAC   908
C_elegans (X54252)          1024  CTTGTTCCAGAATAATCGGCTAGAC 1048
                                   ** ****    ***  *  **
```

rrnL

```
C_intestinalis (AJ517314)   8647  ATTA-GAAAAATTAACCTTAGGGATAACAGCGCTAAT   8682
C_savignyi (AB079784)      12051  GTATTGAA-AATTAACCTTAGGGATAACAGCGCTAAT  12086
Halocynthia (AB024528)     12848  GAAATGAAGAACCAACCCCAGGGATAACAGCGC-AAT  12883
H_sapiens (J01415)          2912  CAACGGAACAAGTTACCCTAGGGATAACAGCGCAATC   2948
C_elegans (X54252)         11085  AATTTAAAGAATTTACTCCGGAGTTAACAGAAAATCA  11120
                                   **  **     **     * * ******
```

Fig. 6. Alignment of the short conserved sequence pattern of small (*rrnS*) and large (*rrnL*) subunit ribosomal RNAs. Positions involved in tertiary interactions are underlined; positions involved in stem formation are shaded gray. Asterisks mark conserved positions. Numbers refer to location on the reported mtDNA entry.

rRNA Genes

The 5′ and 3′ ends of the small subunit rRNA gene (*rrnS*) have been tentatively defined as immediately adjacent to the ends of the flanking genes *trnL(UUR)* and *trnW* (Fig. 1). An alignment of ascidian *rrnS* shows the remarkable similarity at both ends of the predicted gene, supporting the validity of these boundaries.

The *C. intestinalis* large subunit rRNA gene (*rrnL*) is located between the *nad1* and *trnI* genes (Fig. 1). Boundaries have been identified by sequence similarity to other ascidian *rrnL* and by searching for the TRGCAKAn$_5$G mt rRNA transcription termination signal, which is well conserved near the 3′ end of *rrnL* in a wide range of organisms (Valverde et al. 1994). In *H. roretzi*, EST data demonstrate that *rrnL* transcripts stop with a poly(A) tail added at multiple sites located in an AT-rich region immediately upstream of a typical rRNA termination box, TGGCATAtaaacG (Gissi and Pesole 2003). Searches using the *Halocynthia* termination box as a probe revealed a degenerate version of the rRNA termination signal (TGTCGGAtaaacG) in the mtDNA of *C. intestinalis*. This motif is present in the whole genome only in the gene *trnI*, located immediately downstream of the *rrnL*. The pattern TGKCRKAtaaayG might be interpreted as an rRNA termination box specific for ascidians. The 5′ end of *C. intestinalis rrnL* has been tentatively located 100 bp downstream of *nad1*, that is, after a repeated sequence found in this 100 bp-long region (see below) and starting from the region of high similarity to *C. savignyi rrnL*.

Both ascidian rRNA genes are the shortest among deuterostomes (Table 2), moreover, they are comparable in length to those of nematodes and platyhelminthes, which possess the shortest rRNAs among metazoa (Hu et al. 2002).

Both rRNA genes of *Ciona intestinalis* show very low global similarity to the mitochondrial rRNAs of other metazoan species. Using such genes as queries in a Blastn search against the nonredundant nucleotide database (nr), the only statistically significant matches to other organisms are with very short regions of mt rRNA genes (data not shown). Indeed, the only element which identifies *rrnS* is a 23-bp-long sequence. In the human rRNA secondary structure (Cannone et al. 2002) this motif is reported as part of a small stem–loop structure, with the loop implicated in tertiary interactions. In *C. elegans rrnS*, this structure is conserved, although the primary sequence is not identical. Figure 6 reports the alignment of these rRNA regions in ascidians, human, and *C. elegans*, together with data on secondary structure. As for *rrnS*, a short region, 35–18 bp long, allows unambiguous identification of the *rrnL* gene. This region has been mapped on the secondary structure model of human and *C. elegans* large subunit mt rRNA (Cannone et al. 2002), showing that the ends are included in two distinct stem structures, whereas the internal part is unpaired and highly conserved at the level of the primary sequence (Fig. 6). The high degree of conservation suggests that both regions represent important signatures of mt rRNAs.

The low global similarities to homologous rRNAs suggest that the ascidian rRNAs are subject to a fast evolutionary rate and/or to taxon-specific functional constraints.

Noncoding Regions

There are 447 noncoding bp in the mtDNA of *C. intestinalis*, representing 3% of the whole genome, a portion comparable to that observed in other ascidian genomes (2.4% in *C. savignyi* and 3.3% in *H. roretzi*). Sixteen noncoding regions, ranging in size from 1 to 20 bp, have been considered intergenic spacers. Four additional noncoding regions range from 30 to 50 bp. The two longest noncoding regions

are an 85-bp-long region located between *cox3* and *trnK* and a tentatively 100-bp-long noncoding region located between *nad1* and *rrnS* (Fig. 1). These two regions share a sequence 30 bp long, identical except for a single T/A substitution, which is located 11 bp downstream of a further conserved sequence, 13 bp long, that includes the stop codon of the upstream protein-coding gene. Overall, there is an imperfect repeated sequence 54 bp long showing 87% identity and mostly contained in the two longest noncoding regions. As inferred from mitochondrial-like EST analysis (Gissi and Pesole 2003), both such long noncoding regions are transcribed and seem to form a 3′ untranslated region in the *cox3* and *nad1* transcripts. There are no similar sequences in other regions of *C. intestinalis* mtDNA or in the mt genome of other ascidians. The functional meaning of this repeated sequence is obscure but it is peculiar that the two copies of this repeat are close to *nad1*, which is the only protein-gene transposed in both *Ciona* mtDNAs. In *C. savignyi* *nad1* is flanked by short noncoding regions (18 and 69 bp long upstream and downstream, respectively) with no similarity to those of *C. intestinalis*. The *C. savignyi* *nad1* has an extra 30-aa-long C-terminal region, absent in homologous deuterostome proteins. This region contains a 9-bp-long motif which is identically repeated in the noncoding region upstream of *nad1*. Thus, the *nad1* C-terminal region could have originated by the extension of the original frame into a secondarily acquired downstream region. Only one other repeated sequence has been found in ascidian mtDNAs: in *H. roretzi* there are two copies of *trnF* with a 94% identity and conservation of the tRNA cloverleaf secondary structure, clearly suggesting a recent duplication of the *trnF* gene (Gissi and Pesole 2003). The imperfect repeats found in *C. intestinalis* mtDNA could represent traces of the *nad1* transposition event—remnants either of the ancestral target sequence or of a tRNA gene transposed together with *nad1*.

No similarity has been found between the noncoding regions of the three completely sequenced mtDNAs, thus sequence conservation does not help the identification of a putative control region. Only weak speculations can be made as to the existence of a control region in ascidian mtDNAs. Considering that high AT content is a feature of some invertebrate control regions (Saccone and Sbisà 1994), the 37-bp-long noncoding region located between *nad6* and *trnG(GGN)* (97% AT) is a good candidate in *C. intestinalis*. Using the same criterion, in other ascidian mtDNAs the best candidates should be the 112-bp-long sequence of *H. roretzi* (79% AT) and the 47-bp-long sequence of *C. savignyi* (87% AT). However, an alternative hypothesis is that the *C. intestinalis* 85- and 100-bp-long noncoding regions

close to *nad1* represent a multipartite control region. Indeed, such regions are characterized by the presence of an imperfect repeat and proximity to a protein gene transposed between two closely related species. The presence of repeats is a peculiar feature of vertebrate and invertebrate control regions and in Chordata the few tolerated gene transpositions mostly involve the control region (Saccone et al. 1999).

No secondary structures similar to the typical stem–loop of vertebrate L-strand replication origins (Hixson et al. 1986) have been found in *C. intestinalis* or other ascidian mtDNAs.

Conclusions

The *C. intestinalis* and, more generally, the ascidian mtDNA features are very peculiar and suggest that the evolutionary dynamics of these genomes are very different from those of other Chordata. The mt genomes of deuterostomes are believed to be subject to less frequent gene arrangement than protostomes, with Chordata gene arrangement almost frozen (Boore 1999; Saccone et al. 1999). Ascidians should be considered an exception to this tendency because of their high gene rearrangement rate. The extreme reduction of genome size, due to short noncoding regions and smaller rRNA genes, the low GC%, the strong intraclass gene rearrangement, and the transcription of all genes by the same strand are features shared together only by Ascidiacea, Platyhelmintes, and Nematoda. To date an intragenus rearrangement has been found only in the genera *Ciona* and *Schistosoma* (Platyhelminthes). Interestingly, all these taxa show a high evolutionary rate in the nuclear rRNA genes, which leads to long branches in phylogenetic reconstruction (Cameron et al. 2000; Winchell et al. 2002; Aguinaldo et al. 1997; Peterson et al. 2001). Rapid evolution (highly rearranged mt genomes and high substitution rate of nuclear genes) could be the common factor among the above-mentioned taxa. Nevertheless, ascidians still conserve some traits of Chordata mtDNA, particularly of Cephalochordata. Indeed, the absence of a typical vertebrate control region, a C-poor sense strand, and a tRNA-Cys lacking the DHU arm are features common to Cephalochordata and Ascidiacea. Other features, such as the variation of the genetic code with the consequent acquisition of an additional *trnG* gene (Yokobori et al. 2001) and the presence of two *trnM* genes whose functionality and decoding ability have to be demonstrated, are almost exclusive of ascidians. The *atp8* gene, which was previously undetected in ascidian mtDNAs, is conservatively present in all Chordata mt genomes.

Appendix

Table A1. List of completely sequenced metazoan mtDNAs showing all genes transcribed from one strand

Taxon	Species	Accession no.
Annelida	*Lumbricus terrestris*	U24570
	Platynereis dumerilii	AF178678
Arthropoda	*Tigriopus japonicus*	AB060648
Brachiopoda	*Laqueus rubellus*	AB035869
	Terebratalia transversa	AF331161
	Terebratulina retusa	AJ245743
Cnidaria	*Acropora tenuis*	AF338425
Zoantharia		
	Metridium senile	AF000023
Mollusca	*Crassostrea gigas*	AF177226
	Mytilus edulis	M83756-62
	Venerupis philippinarum female	AB065375
	Venerupis philippinarum male	AB065374
Nematoda	*Ancylostoma duodenale*	AJ417718
	Ascaris suum	X54253
	Brugia malayi	AF538716
	Caenorhabditis elegans	X54252
	Necator americanus	AJ417719
	Onchocerca volvulus	AF015193
Platyhelminthes	*Echinococcus granulosus* G1	AF297617
	Echinococcus granulosus G4	AF346403
	Echinococcus multilocularis	AB018440
	Fasciola hepatica	AF216697
	Paragonimus westermani	AF540958
	Schistosoma japonicum	AF215860
	Schistosoma mansoni	AF216698
	Schistosoma mekongi	AF217449
	Taenia crassiceps	AF216699
	Taenia solium	AB086256

Table A2. List of completely sequenced metazoan mtDNAs showing a gene order with the two rRNA genes separated by more than three genes

Taxon	Species	Accession no.
Brachiopoda	*Terebratalia transversa*	AF331161
Cnidaria	*Acropora tenuis*	AF338425
	Metridium senile	AF000023
	Sarcophyton glaucum	AF064823
Echinodermata	*Arbacia lixula*	X80396
Echinoidea		
	Paracentrotus lividus	J04815
	Strongylocentrotus purpuratus	X12631
Mollusca	*Albinaria coerulea*	X83390
	Cepaea nemoralis	U23045
	Euhadra herklotsi	Z71693-701
	Mytilus edulis	M83756-62
	Pupa strigosa	AB028237
	Roboastra europaea	AY083457
	Venerupis philippinarum female	AB065375
	Venerupis philippinarum male	AB065374
Nematoda	*Ancylostoma duodenale*	AJ417718
	Ascaris suum	X54253
	Brugia malayi	AF538716
	Caenorhabditis elegans	X54252
	Necator americanus	AJ417719
	Onchocerca volvulus	AF015193

Table A3. List of completely sequenced metazoan mtDNAs showing a gene order with *atp6* and *atp8* genes distantly located

Taxon	Species	Accession no.
Annelida	*Lumbricus terrestris*	U24570
	Platynereis dumerilii	AF178678
Cnidaria	*Acropora tenuis*	AF338425
Brachiopoda	*Laqueus rubellus*	AB035869
	Terebratalia transversa	AF331161
Mollusca Gastropoda	*Albinaria caerulea*	X83390
	Cepaea nemoralis	U23045
	Euhadra herklotsi	Z71693-701
	Pupa strigosa	AB028237
	Roboastra europaea	AY083457
Nematoda	*Trichinella spiralis*	AF293969

Table A4. Mitochondrial sequences analyzed for the calculation of *rrnL* and *rrnS* gene length

Taxon	Species	Accession no.
Mammalia	*Mus musculus*	V00711
Agnatha	*Eptatretus burgeri*	AJ278504
	Lampetra fluviatilis	Y18683
	Myxine glutinosa	AJ404477
	Petromyzon marinus	U11880
Cephalochordata	*Branchiostoma floridae*	AF098298
	Branchiostoma lanceolatus	Y16474
Urochordata	*Halocynthia rorretzi*	AB024528
	Ciona savignyi	AB079784
	Ciona intestinalis	AJ517314
Hemichordata	*Balanoglossus carnosus*	AF051097
Echinodermata	*Arbacia lixula*	X80396
	Asterina pectinifera	D16387
	Florometra serratissima	AF049132
	Paracentrotus lividus	J04815
	Strongylocentrotus purpuratus	X12631
Arthropoda	*Drosophila melanogaster*	U37541
Nematoda	*Ancylostoma duodenale*	AJ417718
	Ascaris suum	X54253
	Brugia malayi	AF538716
	Caenorhabditis elegans	X54252
	Necator americanus	AJ417719
	Onchocerca volvulus	AF015193
	Trichinella spiralis	AF293969
Platyhelminthes	*Echinococcus granulosus* G1	AF297617
	Echinococcus granulosus G4	AF346403
	Echinococcus multilocularis	AB018440
	Fasciola hepatica	AF216697
	Paragonimus westermani	AF219379
	Schistosoma japonicum	AF215860
	Schistosoma mansoni	AF216698
	Schistosoma mekongi	AF217449
	Taenia crassiceps	AF216699
	Taenia solium	AB086256

References

Aguinaldo AM, Turbeville JM, Linford LS, Rivera MC, Garey JR, Raff RA, Lake JA (1997) Evidence for a clade of nematodes, arthropods and other moulting animals. Nature 387:489–493

Beagley CT, Okimoto R, Wolstenholme DR (1999) Mytilus mitochondrial DNA contains a functional gene for a tRNA-Ser(UCN) with a dihydrouridine arm-replacement loop and a pseudo-tRNASer(UCN) gene. Genetics 152:641–652

Benson G (1999) Tandem repeats finder: A program to analyze DNA sequences. Nucleic Acids Res 27:573–580

Blanchette M, Kunisawa T, Sankoff D (1999) Gene order breakpoint evidence in animal mitochondrial phylogeny. J Mol Evol 49:193–203

Boore JL (1999) Animal mitochondrial genomes. Nucleic Acids Res 27:1767–1780

Boore JL (2001) Complete mitochondrial genome sequence of the polychaete annelid Platynereis dumerilii. Mol Biol Evol 18:1413–1416

Boore JL, Daehler LL, Brown WM (1999) Complete sequence, gene arrangement, and genetic code of mitochondrial DNA of the cephalochordate Branchiostoma floridae (Amphioxus). Mol Biol Evol 16:410–418

Cameron CB, Garey JR, Swalla BJ (2000) Evolution of the chordate body plan: new insights from phylogenetic analyses of deuterostome phyla. Proc Natl Acad Sci USA 97:4469–4474

Cannone JJ, Subramanian S, Schnare MN, Collett JR, D'Souza LM, Du Y, Feng B, Lin N, Madabusi LV, Muller KM, Pande N, Shang Z, Yu N, Gutell RR (2002) The Comparative RNA Web (CRW) site: An online database of comparative sequence and structure information for ribosomal, intron, and other RNAs: Correction. BMC Bioinform 3:15

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

Christianson TW, Clayton DA (1986) In vitro transcription of human mitochondrial DNA: Accurate termination. Proc Natl Acad Sci USA 83:6277–6281

Copello M, Devos L, Lafargue F (1981) Ciona edwardsi (Roule, 1886) espèce littorale de méditerranée distincte de Ciona intestinalis Linné, 1767. Vie Milieu 31:243–253

Cserzo M, Wallin E, Simon I, von Heijne G, Elofsson A (1997) Prediction of transmembrane alpha-helices in procariotic membrane proteins: The dense alignment surface method. Prot Eng 10:673–676

Dehal P, Satou Y, Campbell RK, Chapman J, Degnan B, De Tomaso A, Davidson B, Di Gregorio A, Gelpke M, Goodstein DM, Harafuji N, Hastings KE, Ho I, Hotta K, Huang W, Kawashima T, Lemaire P, Martinez D, Meinertzhagen IA, Necula S, Nonaka M, Putnam N, Rash S, Saiga H, Satake M, Terry A, Yamada L, Wang HG, Awazu S, Azumi K, Boore J, Branno M, Chin-Bow S, DeSantis R, Doyle S, Francino P, Keys DN, Haga S, Hayashi H, Hino K, Imai KS, Inaba K, Kano S, Kobayashi K, Kobayashi M, Lee BI, Makabe KW, Manohar C, Matassi G, Medina M, Mochizuki Y, Mount S, Morishita T, Miura S, Nakayama A, Nishizaka S, Nomoto H, Ohta F, Oishi K, Rigoutsos I, Sano M, Sasaki A, Sasakura Y, Shoguchi E, Shin-i T, Spagnuolo A, Stainier D, Suzuki MM, Tassy O, Takatori N, Tokuoka M, Yagi K, Yoshizaki F, Wada S, Zhang C, Hyatt PD, Larimer F, Detter C, Doggett N, Glavina T, Hawkins T, Richardson P, Lucas S, Kohara Y, Levine M, Satoh N, Rokhsar DS (2002) The draft genome of Ciona intestinalis: insights into chordate and vertebrate origins. Science 298:2157–2167

Devenish RJ, Papakonstantinou T, Galanis M, Law RH, Linnane AW, Nagley P (1992) Structure/function analysis of yeast mitochondrial ATP synthase subunit 8. Ann NY Acad Sci 671:403–414

Durrheim GA, Corfield VA, Harley EH, Ricketts MH (1993) Nucleotide sequence of cytochrome oxidase (subunit III) from the mitochondrion of the tunicate Pyura stolonifera: Evidence that AGR encodes glycine. Nucleic Acids Res 21:3587–3588

Fearnley IM, Walker JE (1986) Two overlapping genes in bovine mitochondrial DNA encode membrane components of ATP synthase. EMBO J 5:2003–2008

Foury F, Roganti T, Lecrenier N, Purnelle B (1998) The complete sequence of the mitochondrial genome of Saccharomyces cerevisiae. FEBS Lett 440:325–313

Gissi C, Pesole G (2003) Transcript mapping and genome annotation of ascidian mtDNA using EST data. Genome Res 13:2203–2212

Gray MW, Lang BF, Cedergren R, Golding GB, Lemieux C, Sankoff D, Turmel M, Brossard N, Delage E, Littlejohn TG, Plante I, Rioux P, Saint-Louis D, Zhu Y, Burger G (1998) Genome structure and gene content in protist mitochondrial DNAs. Nucleic Acids Res 26:865–878

Grillo G, Attimonelli M, Liuni S, Pesole G (1996) CLEANUP: A fast computer program for removing redundancies from nucleotide. Comput Appl Biosci 12:1–8

Hixson JE, Wong TW, Clayton DA (1986) Both the conserved stem-loop and divergent 5′-flanking sequences are required for initiation at the human mitochondrial origin of light- strand DNA replication. J Biol Chem 261:2384–2093

Hoffmann RJ, Boore JL, Brown WM (1992) A novel mitochondrial genome organization for the blue mussel, Mytilus edulis. Genetics 131:397–412

Holland PWH, Garcia-Fernàndez J, Williams NA, Sidow A (1994) Gene duplications and the origins of vertebrate development. Development 1994:125–133

Hoshino Z, Nishikawa T (1985) Taxonomic studies of Ciona intestinalis (L.) and its allies. Publ Seto Mar Biol Lab 10:61–79

Hu M, Chilton NB, Gasser RB (2002) The mitochondrial genomes of the human hookworms, Ancylostoma duodenale and Necator americanus (Nematoda: Secernentea). Int J Parasitol 32:145–158

Jacobs HT, Balfe P, Cohen B, Farquharson A, Comito L (1988) Phylogenetic implications of genome rearrangement and sequence evolution in echinoderm mitochondrial DNA. In: Paul CRC, Smith AB (eds) Echinoderm phylogeny and evolutionary biology. Clarendon Press, Oxford, pp 121–137

Kakuda T (2001) Mitochondrial DNA analysis of Boltenia echinata ibuki (OKA, 1934). In: Sawada H, Yokosawa H, Lambert CC (eds) The biology of ascidians. Springer-Verlag, Tokyo, pp 283–289

Kondow A, Yokobori S, Ueda T, Watanabe K (1998) Ascidian mitochondrial tRNA(Met) possessing unique structural characteristics. Nucleosides Nucleotides 17:531–539

Kumazawa Y, Nishida M (1993) Sequence evolution of mitochondrial tRNA genes and deep-branch animal phylogenetics. J Mol Evol 37:380–398

Lambert C, Lafargue F, Lambert G (1990) Preliminary note on the genetic isolation of Ciona species (Ascidiacea, Urochordata). Vie Milieu 40:293–295

Lambert G (2001) A global overview os ascidian introductions and their possible impact on the endemic fauna. In: Sawada H, Yokosawa H, Lambert CC (eds) The biology of ascidians. Springer-Verlag, Tokyo, pp 249–257

Lavrov DV, Brown WM (2001) Trichinella spiralis mtDNA: A nematode mitochondrial genome that encodes a putative ATP8 and normally structured tRNAS and has a gene arrangement relatable to those of coelomate metazoans. Genetics 157:621–637

Le TH, Blair D, Agatsuma T, Humair P-F, Campbell NJH, Iwagami M, Littlewood DT, Peacock B, Johnston DA, Bartley J, Rollinson D, Herniou EA, Zarlenga DS, McManus DP (2000) Phylogenies inferred from mitochondrial gene orders-a cautionary tale from. Mol Biol Evol 17:1123–1125

Lowe TM, Eddy SR (1997) tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res 25:955–964

Macey JR, Schulte JA, 2nd, Larson A, Tuniyev BS, Orlov N, Papenfuss TJ (1999) Molecular phylogenetics, tRNA evolution, and historical biogeography in anguid lizards and related taxonomic families. Mol Phylogenet Evol 12:250–227

Ohno S (1970) Evolution by gene duplication. George Allen and Unwin, London

Papakonstantinou T, Galanis M, Nagley P, Devenish RJ (1993) Each of three positively-charged amino acids in the C-terminal region of yeast mitochondrial ATP synthase subunit 8 is required for assembly. Biochim Biophys Acta 1144:22–32

Papakonstantinou T, Law RH, Nagley P, Devenish RJ (1996a) Non-functional variants of yeast mitochondrial ATP synthase subunit 8 that assemble into the complex. Biochem Mol Biol Int 39:253–260

Papakonstantinou T, Law RH, Nesbitt WS, Nagley P, Devenish RJ (1996b) Molecular genetic analysis of the central hydrophobic domain of subunit 8 of yeast mitochondrial ATP synthase. Curr Genet 30:12–18

Perna NT, Kocher TD (1995) Patterns of nucleotide composition at fourfold degenerate sites of animal mitochondrial genomes. J Mol Evol 41:353–358

Pesole G, Attimonelli M, Saccone C (1996) Linguistic analysis of nucleotide sequences: Algorithms for pattern recognition and analysis of codon strategy. Methods Enzymol 266:281–294

Pesole G, Liuni S, D'Souza M (2000) PatSearch: A pattern matcher software that finds functional elements in nucleotide and protein sequences and assesses their statistical significance. Bioinformatics 16:439–450

Peterson KJ, Eernisse DJ (2001) Animal phylogeny and the ancestry of bilaterians: Inferences from morphology and 18S rDNA gene sequences. Evol Dev 3:170–205

Rajbhandary U, Chow MC (1995) Initiator tRNAs and initiation of protein synthesis. In: Soll D, Rajbhandary U (eds) tRNA structure, biosynthesis and function. American Society of Microbiology, Washington, DC, pp 511–528

Reyes A, Gissi C, Pesole G, Saccone C (1998) Evolution of asymmetric base composition in the mitochondrial genome of mammals. Mol Biol Evol 15:957–966

Saccone C, Sbisà E (1994) The evolution of the mitochondrial genome. In: Bittar EE (ed) Principles of medical biology. JAI Press, Greenwich, CT, pp 39–72

Saccone C, De Giorgi C, Gissi C, Pesole G, Reyes A (1999) Evolutionary genomics in Metazoa: The mitochondrial DNA as model system. Gene 238:195–209

Satoh N (1994) Developmental biology of ascidians. Cambridge University Press, New York

Satou Y, Takatori N, Fujiwara S, Nishikata T, Saiga H, Kusakabe T, Shin-i T, Kohara Y, Satoh N (2002) Ciona intestinalis cDNA projects: Expressed sequence tag analyses and gene expression profiles during embryogenesis. Gene 287:83–96

Seutin G, Lang BF, Mindell DP, Morais R (1994) Evolution of the WANCY region in amniote mitochondrial DNA. Mol Biol Evol 11:329–340

Shu DG, Chen L, Han J, Zhang XL (2001) An Early Cambrian tunicate from China. Nature 411:472–473

Sprinzl M, Horn C, Brown M, Ioudovitch A, Steinberg S (1998) Compilation of tRNA sequences and sequences of tRNA genes. Nucleic Acids Res 26:148–153

Spruyt N, Delarbre C, Gachelin G, Laudet V (1998) Complete sequence of the amphioxus (Branchiostoma lanceolatum) mitochondrial genome: relations to vertebrates. Nucleic Acids Res 26:3279–3285

Stach T, Turbeville JM (2002) Phylogeny of Tunicata inferred from molecular and morphological characters. Mol Phylogenet Evol 25:408–428

Stephens AN, Khan MA, Roucou X, Nagley P, Devenish RJ (2003) The molecular neighborhood of subunit 8 of yeast mitochondrial F1F0-ATP synthase probed by cysteine scanning mutagenesis and chemical modification. J Biol Chem 278:17867–17875

Swalla B (2001) Phylogeny of the Urochordates: Implications for Chordata evolution. In: Sawada H, Yokosawa H, Lambert CC (eds) The biology of ascidians. Springer-Verlag, Tokyo, pp 219–224

Swalla BJ, Cameron CB, Corley LS, Garey JR (2000) Urochordates are monophyletic within the deuterostomes. Syst Biol 49:52–64

Tomita K, Ueda T, Ishiwa S, Crain PF, McCloskey JA, Watanabe K (1999) Codon reading patterns in Drosophila melanogaster mitochondria based on their tRNA sequences: A unique wobble rule in animal mitochondria. Nucleic Acids Res 27:4291–4297

Turbeville JM, Schulz JR, Raff RA (1994) Deuterostome phylogeny and the sister group of the chordates: Evidence from molecules and morphology. Mol Biol Evol 11:648–655

Valverde JR, Marco R, Garesse R (1994) A conserved heptamer motif for ribosomal RNA transcription termination in animal mitochondria. Proc Natl Acad Sci USA 91:5368–5371

Wolstenholme DR (1992) Animal mitochondrial DNA: structure and evolution. Int Rev Cytol 141:173–216

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

Yokobori S, Ueda T, Feldmaier-Fuchs G, Paabo S, Ueshima R, Kondow A, Nishikawa K, Watanabe K (1999) Complete DNA sequence of the mitochondrial genome of the ascidian Halocynthia roretzi (Chordata, Urochodata). Genetics 153:1851–1862

Yokobori S, Suzuki T, Watanabe K (2001) Genetic code variations in mitochondria: tRNA as a major determinant of genetic code plasticity. J Mol Evol 53:314–326

Zardoya R, Meyer A (1996) The complete nucleotide sequence of the mitochondrial genome of the lungfish (Protopterus dolloi) supports its phylogenetic position as a close relative of land vertebrates. Genetics 142:1249–1263

Zhang A (1987) Fossil appendicularians in the early Cambrian. Sci Sinica (B) 30:888–896

Winchell CJ, Sullivan J, Cameron CB, Swalla BJ, Mallatt J (2002) Evaluating hypotheses of deuterostome phylogeny and chordate evolution with new LSU and SSU ribosomal DNA data. Mol Biol Evol 19:762–776