The Mitochondrial Genome of the Hemichordate *Balanoglossus carnosus* and the Evolution of Deuterostome Mitochondria

Jose Castresana,*,[1] Gertraud Feldmaier-Fuchs,* Shin-ichi Yokobori,*,[2] Noriyuki Satoh† and Svante Pääbo*

**Institute of Zoology, University of Munich, D-80333 Munich, Germany and †Department of Zoology, Faculty of Science, Kyoto University, Kyoto 606-01, Japan*

Manuscript received February 20, 1998
Accepted for publication August 3, 1998

ABSTRACT

The complete nucleotide sequence of the mitochondrial genome of the hemichordate *Balanoglossus carnosus* (acorn worm) was determined. The arrangement of the genes encoding 13 protein, 22 tRNA, and 2 rRNA genes is essentially the same as in vertebrates, indicating that the vertebrate and hemichordate mitochondrial gene arrangement is close to that of their common ancestor, and, thus, that it has been conserved for more than 600 million years, whereas that of echinoderms has been rearranged extensively. The genetic code of hemichordate mitochondria is similar to that of echinoderms in that ATA encodes isoleucine and AGA serine, whereas the codons AAA and AGG, whose amino acid assignments also differ between echinoderms and vertebrates, are absent from the *B. carnosus* mitochondrial genome. There are three noncoding regions of length 277, 41, and 32 bp: the larger one is likely to be equivalent to the control region of other deuterostomes, while the two others may contain transcriptional promoters for genes encoded on the minor coding strand. Phylogenetic trees estimated from the inferred protein sequences indicate that hemichordates are a sister group of echinoderms.

THE deuterostomes are a major group of metazoans composed of chordates (vertebrates, cephalochordates, and urochordates), hemichordates, echinoderms, and possibly other phyla. Among deuterostomes, complete mitochondrial DNA sequences (mtDNAs) have been determined in several vertebrate groups, including mammals (*e.g.*, Anderson *et al.* 1981), birds (*e.g.*, Desjardins and Morais 1990), amphibians (Roe *et al.* 1985), lobe-finned fishes (*e.g.*, Zardoya and Meyer 1996), ray-finned fishes (*e.g.*, Tzeng *et al.* 1992), and lampreys (Lee and Kocher 1995). In echinoderms, mtDNAs of three sea urchins and a starfish have been described (Jacobs *et al.* 1988; Cantatore *et al.* 1989; Asakawa *et al.* 1995; De Giorgi *et al.* 1996). Whereas the mitochondrial genome is very similar among vertebrates, the mtDNAs of echinoderms differ from the vertebrates in gene arrangement, genetic code, and nucleotide signals involved in transcription and DNA replication.

The hemichordates are a phylum represented by around 85 extant species (Brusca and Brusca 1990). They are divided into enteropneusts (or acorn worms), which are worm-like, marine animals that live buried under soft sediments, and pterobranchs, which are sessile, colonial animals with tentaculate arms. Despite the low number of hemichordate species, the group has a central role in understanding the evolution of deuterostomes because they possess some chordate characters in the adult form, whereas their embryology provides a link with the echinoderms (Hyman 1959). Furthermore, because of this peculiar amalgam of morphological characters, the phylogenetic position of hemichordates within deuterostomes is controversial (Willmer 1990; Nielsen 1995; Gee 1996).

To better understand the evolution of the deuterostome mitochondrial genome, as well as the deuterostome phylogeny, we have determined the complete mtDNA sequence of a hemichordate, the acorn worm *Balanoglossus carnosus.* The phylogeny estimated from the inferred *B. carnosus* mitochondrial proteins agrees with results from 18S RNA sequences (Turbeville *et al.* 1994; Wada and Satoh 1994; Halanych 1995), indicating that hemichordates are a sister group of echinoderms. However, as is the case with many morphological features of hemichordates, their "mtDNA plan" exhibits a mixture of chordate and echinoderm characters.

MATERIALS AND METHODS

The complete mtDNA was amplified from total DNA of a *B. carnosus* captured near Seto Marine Biological Laboratory of Kyoto University (Shirahama, Wakayama, Japan) by means of two overlapping long PCR reactions. To accomplish this,

Corresponding author: Jose Castresana, European Molecular Biology Laboratory, Meyerhofstrasse 1, 69117 Heidelberg, Germany. E-mail: castresa@embl-heidelberg.de

[1]*Present address:* European Molecular Biology Laboratory, Meyerhofstrasse 1, 69117 Heidelberg, Germany.

[2]*Present address:* Department of Molecular Biology, Faculty of Life Science, Tokyo University of Pharmacy and Life Science, Tokyo 192-03, Japan.

two short segments of the 12S rRNA and cytochrome *c* oxidase subunit 1 (*CO1*) genes were obtained initially by standard PCR using versatile primers (Kocher *et al.* 1989; Yokobori *et al.* 1993). Using the *CO1* gene sequence, we designed the primers 5′-CCTTTGGGATGATCTCTCATGTCA-3′ (CO1-1) and 5′-CAAAGAAGAGGGGCTTCTCATTGGA-3′ (CO1-2), and using the 12S rRNA gene, we designed the primers 5′-ACAAAACTCAAAGGATTTGGCGGTATC-3′ (12S-1) and 5′-TATAACTGTTGCAAGTAATAATTCGTC-3′ (12S-2). Both pairs were first tested in amplifications of their respective gene fragments before being used in the following four combinations: CO1-1 with 12S-1, CO1-1 with 12S-2, CO1-2 with 12S-1, and CO1-2 with 12S-2 because we did not know the relative orientation of these two genes in the *B. carnosus* mitochondrial genome. A mixture of the *Pfu* and *Taq* polymerases with the buffers supplied in the TaqPlus system (Stratagene, La Jolla, CA) was used for the long PCR amplifications. The second and third primer pairs yielded 10.5- and 5.7-kb products, respectively. Both amplification products were digested with *Bam*HI, generating a total of five fragments with lengths of 2–3.8 kb. These fragments were cloned into a pBluescript vector (Stratagene). To exclude PCR and cloning errors in the sequences, three clones for each fragment were completely sequenced, covering both strands, by primer walking using the dideoxy chain-termination technique. In all cases, the three clones represented independent cloning events because they differed in a few substitutions from each other. Sequences surrounding the *Bam*HI sites were confirmed by sequencing of cloned PCR products encompassing the restriction sites. The sequence of the *B. carnosus* mtDNA has been deposited in GenBank under the accession no. AF051097.

Genes were detected by comparison with known mitochondrial sequences. Using the Megalign program (Lasergene), unassigned regions were investigated by dot plots to search for similarities with other sequences. Amino acid alignments of several deuterostome protein-coding gene sequences extracted from GenBank (Benson *et al.* 1998) were performed with ClustalW (version 1.7, Thompson *et al.* 1994) using default parameters. The ND6 subunit of NADH dehydrogenase was excluded from the analyses because its sequence was not available for one of the species considered (hagfish). Regions of problematic alignment were detected and removed with a program (J. Castresana, unpublished results) in which homologous segments were defined according to the following criteria: (1) they contain less than eight (or four) contiguous positions where $<50\%$ of the taxa analyzed have an identical residue, (2) they are flanked by positions where $>85\%$ of the taxa have identical residues, and (3) they have a total length >15 positions. Furthermore, positions with gaps and nonconserved positions adjacent to such gaps were removed. The resulting alignments were concatenated for the phylogenetic analysis. Two further datasets were generated by removing positions containing any Met or Lys residue because the frequency of these two amino acids differs in echinoderm and vertebrate mitochondrial proteins (Castresana *et al.* 1998). Phylogenetic trees were estimated by the neighbor-joining algorithm using the Dayhoff matrix of amino acid substitution (Dayhoff *et al.* 1978) from 100 bootstrap replicates, as implemented in the PHYLIP package (version 3.572c, Felsenstein 1993) and by maximum likelihood using the mtREV (Adachi and Hasegawa 1996a), JTT (Jones *et al.* 1992), and Dayhoff (Dayhoff *et al.* 1978) models of amino acid substitution with amino acid frequencies from the respective datasets using the MOLPHY package (version 2.3, Adachi and Hasegawa 1996b). Robustness of the phylogenetic position of *B. carnosus* was tested by comparing the maximum-likelihood trees with alternative tree topologies by means of the Kishino-Hasegawa test (Kishino and Hasegawa 1989) and bootstrap analysis in 10,000 replicates by the RELL (resampling of the estimated log likelihood) method (Kishino *et al.* 1990), as implemented in MOLPHY. A difference in log likelihood of >1.96 times the standard error of that difference ($P < 0.05$) was considered significant.

RESULTS AND DISCUSSION

Nucleotide sequence and gene organization: The complete mtDNA sequence of the *B. carnosus* mtDNA is 15,708 bp long and has an A + T content of 51.4%. As in most other metazoan mtDNAs, the genome contains the genes for 13 proteins (three subunits of cytochrome *c* oxidase, CO1, CO2, and CO3; one subunit of cytochrome *c*-ubiquinol oxidoreductase, CYTb; seven subunits of NADH dehydrogenase, ND1, ND2, ND3, ND4L, ND4, ND5, and ND6; and two subunits of ATP synthase, ATP6 and ATP8), 22 tRNAs, and 2 rRNAs (12S and 16S).

The genes are arranged in an extremely economical fashion such that most genes have no bases between their coding bodies, and 11 pairs of genes appear to overlap by 1–10 bases. Of these overlaps, 5 occur between genes encoded on different strands. The remaining 6 cases involve genes encoded on the same strand. In 3 instances, the transcript might be processed such that the upstream protein-coding gene (*ATP6*, *ND5*, and *CYTb*, respectively) gets a UAA stop codon by polyadenylation. A fourth case involves the *ND4* gene and the downstream $tRNA^{His}$ gene. Provided that the extra guanylate residue that $tRNA^{His}$ carries at its 5′ end (Sprinzl *et al.* 1989) and that is added post-transcriptionally in other metazoans (L'Abbe *et al.* 1990) is encoded by the tRNA gene in the hemichordate, a UAA stop codon can be created on the *ND4* transcript by polyadenylation. A fifth case involves the genes for ATP8 and ATP6, which overlap by 10 bases. A similar overlap between these genes is seen in most other metazoans, and it has been shown that the transcription of *ATP6* is initiated within the ATP8-coding sequence in bovine mitochondria (Fearnley and Walker 1986). The sixth and final case involves the 3′ end of the gene for $tRNA^{Glu}$ and the 5′ end of the gene for $tRNA^{Thr}$. Such overlaps involving mitochondrial tRNAs encoded on the same strand occur in molluscs, annelids, echinoderms, and vertebrates (Börner *et al.* 1997). It has been shown that they are processed such that the downstream tRNA is released in its intact form, whereas the upstream tRNA is truncated and gets filled in by an RNA-editing activity (Yokobori and Pääbo 1995a,b) that may involve polyadenylation (Yokobori and Pääbo 1997).

The organization of the genes and noncoding regions is presented in Figure 1 with the situations in a vertebrate, human (Anderson *et al.* 1981), and in an echinoderm, the sea urchin *Arbacia lixula* (De Giorgi *et al.* 1996). It can be observed that the gene order of the *B. carnosus* mtDNA is very similar to that in humans and,

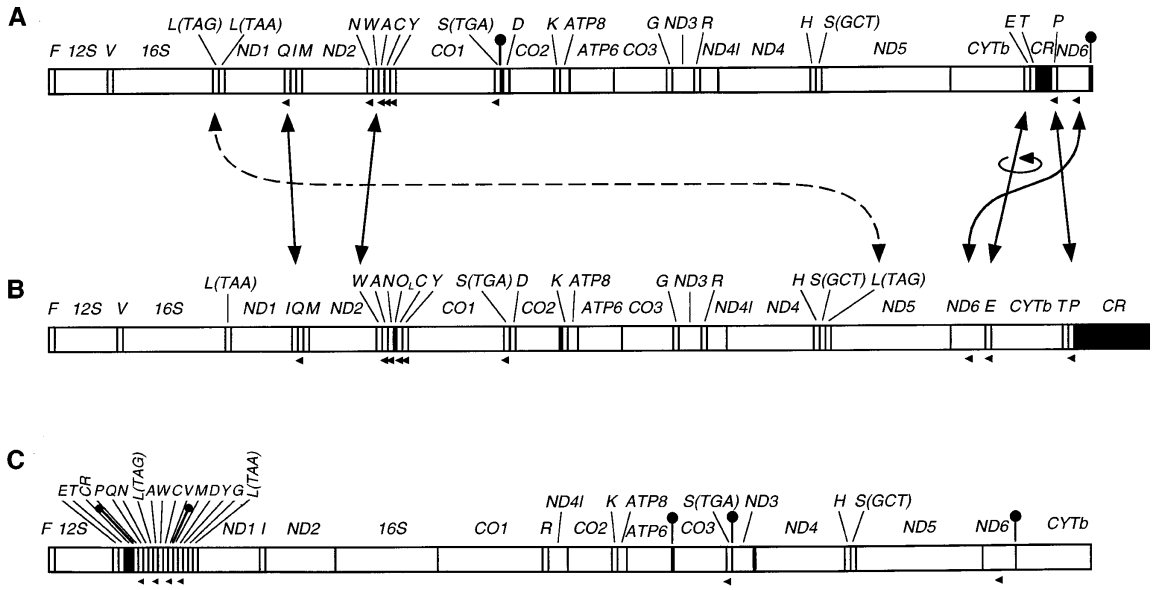

Figure 1.—Gene organization of (A) the *B. carnosus* (hemichordate) mitochondrial genome compared to that of (B) human (vertebrate) and (C) the sea urchin *Arbacia lixula* (echinoderm). Protein genes are designated using the abbreviations given in the text. Transfer RNA genes are indicated by the single-letter abbreviations. The isoacceptors of serine and leucine tRNAs are given with the corresponding anticodons in parentheses. Genes coded in the opposite strand are shown with an arrowhead. Noncoding regions are shown as black boxes. CR, control region; O_L (in the human mtDNA), the L-strand origin of replication; solid circles, putative signals of transcription outside the control region in the hemichordate and the echinoderm mitochondrial genomes. The minimum number of translocations necessary to interconvert the *B. carnosus* and the human mitochondrial genomes are indicated with double arrows. An inversion is indicated with an additional circular arrow. The differences with the echinoderm genome are too numerous to be depicted. The dashed line indicates that this translocation may have arisen by a duplication of the $tRNA^{Leu}$(TAA) gene.

therefore, to other vertebrates, requiring a total of only six rearrangements to convert one gene order into the other. Only the gene encoding $tRNA^{Glu}$ has changed from one strand to the other. In contrast, the order of genes in the echinoderm is remarkably different. One striking feature of the latter genome is the cluster of 13 tRNA genes that also occurs in the two other sea urchins (Jacobs *et al.* 1988; Cantatore *et al.* 1989) and in the starfish (Asakawa *et al.* 1995), although in the latter species, an inversion with respect to the sea urchins of a 4.6-kb region encompassing the tRNA gene cluster is seen.

Given the very different gene arrangements between vertebrates and echinoderms, it has been debated which of these might represent the ancestral gene order of these phyla (Cantatore *et al.* 1987; Jacobs *et al.* 1989a). Because the earliest group to branch off within vertebrates, the lampreys (Cyclostomata), have a gene order similar to that of other vertebrates (Lee and Kocher 1995), and partial sequences of a cephalochordate (Delarbre *et al.* 1997) appear to show the same arrangement as vertebrates, it is likely that the chordate gene order was established more than 400 mya. Because the hemichordate has virtually the same gene order as vertebrates, this pushes the date for the establishment of this organization of the mtDNA back to at least the split between vertebrates and hemichordates, more than 600 mya (Ayala *et al.* 1998). Furthermore, because hemichordates are probably more related to echinoderms than to vertebrates (Turbeville *et al.* 1994; Wada and Satoh 1994; Halanych 1995; see Figure 4), the vertebrate/hemichordate gene arrangement is likely to be the ancestral, and the echinoderm is likely to be derived.

The similarity between the hemichordate and vertebrate gene arrangements allows more detailed questions regarding translocations of specific genes to be addressed. Taking the high number of identities between the two $tRNA^{Leu}$ genes in the echinoderm *Paracentrotus lividus* into account, it has been proposed that the $tRNA^{Leu}$(TAG) gene arose from a duplication of the $tRNA^{Leu}$(TAA) gene (Cantatore *et al.* 1987). Echinoderms show a close similarity between both $tRNA^{Leu}$ genes, ranging from 67 to 81% identity (compared to 38–48% identity between vertebrate $tRNA^{Leu}$ genes, Sprinzl *et al.* 1989), pointing to a gene duplication that preceded the split of sea urchins and starfish. In the *B. carnosus* mitochondrial genome, both $tRNA^{Leu}$ genes are also very similar (63% identity), indicating that the duplication took place before the hemichordate/echinoderm split. Furthermore, in the hemichordate mitochondrial genome, the $tRNA^{Leu}$(TAG) gene is located immediately upstream of the $tRNA^{Leu}$(TAA) gene. Because the latter gene is located upstream of the *ND1* gene in echinoderms, vertebrates, and the hemichordate, it is probable that the hemichordate mtDNA represents the situation that existed in an ancestral species

after the duplication and subsequent anticodon change in one of the tRNA^{Leu}(TAA) genes took place. We also note that the 5′ end of the hemichordate *ND5* gene, which is adjacent to the tRNA^{Leu}(TAG) gene in vertebrates (Figure 1), is ~80 bases longer than in vertebrates, having the same length as the homologous gene in echinoderms. This makes it plausible that this extension represents the remnants of a tRNA^{Leu}(TAG) gene that was appended to the *ND5* gene, as suggested by Cantatore *et al.* (1987). Because the tRNA^{Leu}(TAG)-tRNA^{Leu}(TAA)-*ND1* cluster also occurs in some molluscs, such as the blue mussel (Hoffmann *et al.* 1992) and the black chiton (Boore and Brown 1994), and in an arthropod, the horseshoe crab (Boore *et al.* 1995), it is conceivable that this arrangement is ancestral to protostomes and deuterostomes. However, the extension of the *ND5* gene does not occur in protostomes, and the adjacent localization of the two tRNA^{Leu} genes is absent in other molluscs and arthropods. Therefore, an independent gene duplication and change of the tRNA^{Leu}(TAA) gene to a tRNA^{Leu}(TAG) gene, as well as the addition of the vestigial tRNA^{Leu}(TAG) gene to the *ND5* gene, probably occurred in a common ancestor of echinoderms and hemichordates.

The comparison of the mitochondrial gene order among organisms has proved to be a valuable tool to establish phylogenetic relationships when shared derived characters can be used for such a purpose (Boore *et al.* 1995). However, the comparison of the gene order of different deuterostome phyla (Figure 1) shows that whereas vertebrates and hemichordates maintain many features of an ancestral gene order, echinoderms have experienced a large number of rearrangements, leaving no shared derived characters between phyla to establish a reliable phylogeny. Thus, the different rate of gene arrangements between groups makes it hard to identify shared derived characters and, thus, complicates the use of mtDNA gene order as a phylogenetic tool for metazoan phyla.

Protein-coding genes and genetic code: The *B. carnosus* mitochondrial genome encodes the same 13 proteins as other metazoans (Figure 1). While ATG serves as a start codon in most genes, GTG is used in *CO2* and *ATP8*, ATA in *ND6*, and CTG in *CO1*. The CTG initiation codon has not been seen to date in other metazoan mtDNAs (Wolstenholme 1992), but it has been shown to initiate translation in mammalian nuclear genes (Hann *et al.* 1988; Peabody 1989). The occurrence of this unusual mitochondrial initiation codon in *B. carnosus* is likely because the inferred N-terminal part of the hemichordate CO1 protein is well conserved compared to that of other metazoans. TAA or TAG stop codons are used, except in *CYTb* and possibly a few other genes that overlap with downstream genes, where only one T or TA is found (see above). In these cases, a TAA stop codon is probably created post-transcriptionally through polyadenylation, as is the case in other organisms.

There are four mitochondrial codons whose meaning varies among the main deuterostome groups. Thus, ATA codes for Met in vertebrates and urochordates, and for Ile in echinoderms; AAA codes for Lys in vertebrates and urochordates, and for Asn in echinoderms; and AGA and AGG are stop codons in vertebrates, but they encode Ser in echinoderms and Gly in urochordates (Yokobori *et al.* 1993). Table 1 shows the codons and their amino acid assignments in the 13 protein genes of the *B. carnosus* mtDNA. The amino

TABLE 1

Genetic code and codon usage in the *B. carnosus* mitochondrial genome

Codon	Count	Codon	Count	Codon	Count	Codon	Count
TTT Phe	98	TCT Ser	51	TAT Tyr	30	TGT Cys	11
TTC Phe	157	TCC Ser	95	TAC Tyr	77	TGC Cys	21
TTA Leu	59	TCA Ser	79	TAA End	9	TGA Trp	72
TTG Leu	34	TCG Ser	26	TAG End	3	TGG Trp	38
CTT Leu	164	CCT Pro	51	CAT His	26	CGT Arg	12
CTC Leu	212	CCC Pro	91	CAC His	80	CGC Arg	18
CTA Leu	155	CCA Pro	58	CAA Gln	89	CGA Arg	48
CTG Leu	63	CCG Pro	10	CAG Gln	22	CGG Arg	10
ATT Ile	97	ACT Thr	62	AAT Asn	20	AGT Ser	18
ATC Ile	148	ACC Thr	115	AAC Asn	98	AGC Ser	35
ATA Ile	69	ACA Thr	81	AAA	0	AGA Ser	19
ATG Met	77	ACG Thr	20	AAG Lys	45	AGG	0
GTT Val	63	GCT Ala	62	GAT Asp	15	GGT Gly	34
GTC Val	83	GCC Ala	151	GAC Asp	58	GGC Gly	77
GTA Val	41	GCA Ala	108	GAA Glu	56	GGA Gly	78
GTG Val	29	GCG Ala	21	GAG Glu	27	GGG Gly	58

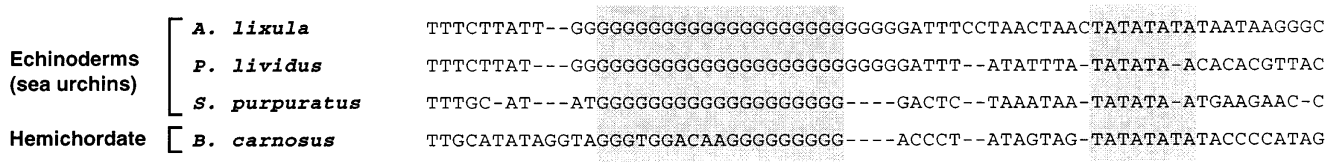

Figure 2.—Alignment of the segments with best similarity between the echinoderms and *B. carnosus* control regions. The *B. carnosus* sequence has been inverted with respect to the one represented in Figure 1. An asterisk indicates that the position is completely conserved in the species considered. The shaded blocks cover the best conserved motifs within this segment.

acid assignments were achieved by multiple alignments of the *B. carnosus* protein-coding genes with proteins from 21 species representing five metazoan phyla. The codons ATA and AGA correspond to Ile and Ser, respectively, as is the case in echinoderms. Specifically, out of the 69 ATA hemichordate codons, 20 occur in positions where Ile is conserved in >55% of the sequences (eight of these positions have Ile in >90% of the sequences), while none of them occur at positions where Met is conserved. Of the 19 AGA codons, 6 occur at positions where Ser is conserved in >55% of the sequences (3 of them with Ser in >90% of the sequences), whereas they are never used as stop codons. The 2 other codons that differ between the deuterostome phyla, AAA and AGG, are absent from the *B. carnosus* mtDNA, while all other 62 codon triplets are used (Table 1).

The overall sequences and predicted secondary structures of the *B. carnosus* tRNAs are similar to those of other deuterostomes (Sprinzl *et al.* 1989). However, one feature is of special interest. In the *B. carnosus* mtDNA, the $tRNA^{Lys}$ gene carries the anticodon CTT, as is the case in echinoderms (Asakawa *et al.* 1995), whereas this tRNA gene carries the anticodon TTT in most other metazoans. This feature, which probably limits the recognition capacity of the $tRNA^{Lys}$ to the codon AAG, together with the absence of AAA codons, could reflect an intermediate stage during the reassignment of the AAA codon from Lys to Asn in the echinoderm lineage (Castresana *et al.* 1998), and it supports the codon capture hypothesis for codon reassignment (Osawa and Jukes 1989).

Noncoding regions and possible transcription and replication signals: Among vertebrates, the main sequences necessary for transcription and initiation of mtDNA replication are situated in the control region, between the $tRNA^{Pro}$ and $tRNA^{Phe}$ genes (Shadel and Clayton 1993, 1997). Two promoters exist in this region, one initiating the transcription of the H strand (the main coding strand) and another initiating the transcription of the opposite strand (L strand). DNA replication is linked to transcription because RNA transcripts initiated at the L-strand promoter serve as primers for replication of the H strand. Many of these initiation products become arrested in a way such that D-loop structures can be observed in many vertebrate mtDNA molecules. The L-strand origin of replication is situated outside the control region, normally between the $tRNA^{Asn}$ and $tRNA^{Cys}$ genes. In echinoderms, it has been shown in the sea urchin *Strongylocentrotus purpuratus* (Jacobs *et al.* 1988, 1989a,b; Elliott and Jacobs 1989) that the mtDNA may contain one major origin of replication where a D loop is formed, albeit smaller than in vertebrates, and that the initiation of DNA replication in this region is also linked to a transcriptional promoter. In addition, the echinoderm mtDNA contains several palindromic decanucleotide motifs scattered throughout the genome (filled circles in Figure 1) that have been proposed to serve as bidirectional transcription initiation signals (Elliott and Jacobs 1989; Jacobs *et al.* 1989a), a situation reminiscent of the multiple promoters in the yeast mitochondrial genome (Christianson and Rabinowitz 1983).

To explore possible models of transcription and replication of the *B. carnosus* mtDNA, we looked for similarities by dot plots between its main noncoding region of 277 bp and other vertebrate and echinoderm control regions. The only segment exhibiting similarity to other taxa was a G tract followed by several AT dinucleotides that occur at a similar region in the three sea urchin sequences (Figure 2). No similarities were discerned in comparisons with the starfish or with vertebrates. This region of similarity has been previously observed by comparison of the control regions among the three sea urchins (De Giorgi *et al.* 1996). Interestingly, this is the region where the beginning of the formation of a D loop has been mapped in *S. purpuratus* (Jacobs *et al.* 1989b). This might indicate that a similar structure could also be formed in other echinoderms and in the hemichordate.

No sequences with 9 or 10 matches to the decanucleotide sequence proposed to be involved in transcription in echinoderm mtDNA can be found in the *B. carnosus* genome. However, upon comparing the two small noncoding regions (of 32 and 41 bp, respectively), a decanucleotide motif T(T/C)ACCTTTTT was found at the beginning of both regions when the main coding strand was considered (Figure 3). This motif is found nowhere else in the *B. carnosus* mtDNA. Strikingly, both of these motifs are facing genes that are encoded on the minor coding strand. It is tempting to speculate that these

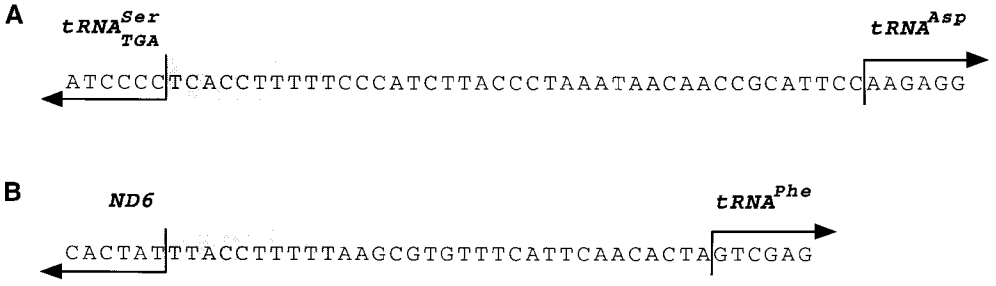

Figure 3.—The two small noncoding regions of the *B. carnosus* mtDNA. The shaded blocks indicate the shared motif found in these regions. Arrows indicate the starting point and direction of transcription of the flanking genes. (A) The region between the $tRNA^{Ser}$(TGA) and $tRNA^{Asp}$ genes. (B) The region between the *ND6* and $tRNA^{Phe}$ genes.

noncoding regions contain initiation signals for two short transcripts that could cover all the genes located on the minor coding strand (Figure 1). If that were true, the hemichordate mtDNA would represent an intermediate stage between the vertebrate (with only two promoters, both located in the control region) and the echinoderm (with multiple promoters scattered around the genome) modes of transcription. Experimental data from these and other animals would be necessary to fully understand the function and evolution of the promoter sequences.

Phylogenetic trees: Hemichordates possess adult characters, such as gill slits or a dorsal hollow nerve cord, that link them to the chordates, while many embryological features place them closer to echinoderms (Hyman 1959; Schaefer 1987). As a consequence, their phylogenetic position within deuterostomes has been widely debated (Willmer 1990; Nielsen 1995; Gee 1996). Complete nuclear 18S RNA sequences from the enteropneusts *B. carnosus* (Wada and Satoh 1994) and *Saccoglossus kowalevskii* (Turbeville *et al.* 1994), as well as partial sequences of *S. cambrensis* (Holland *et al.* 1991) and the pterobranch *Rhabdopleura normani* (Halanych 1995), have been recently used to try to assess the phylogenetic affinities of hemichordates. With the exception of the analysis of the *S. cambrensis* partial sequence, these studies have found the hemichordates to be the sister taxon of echinoderms rather than vertebrates, but without sufficient support to arrive at a firm conclusion.

Alignments of the inferred amino acid sequences of 12 *B. carnosus* mtDNA proteins with the homologous proteins in echinoderms, chordates, and a mollusc used as an outgroup were constructed. Regions of problematic alignment were removed, allowing for less than eight (or four) contiguous, nonconserved positions in the selected segments (see materials and methods), and the resulting alignments were concatenated. Furthermore, because the frequency of the amino acids Lys and Met has changed in echinoderms and hemichordates (Castresana *et al.* 1998), all positions containing Lys or Met were eliminated from each alignment. Thus, a total of four alignments were produced (Table 2).

Neighbor-joining trees derived from all these alignments show the sister group status of hemichordates and echinoderms to be supported by >95% of bootstrap replications (Figure 4, A–D). The neighbor-joining trees were used as starting points to search, by means of a local rearrangement algorithm, for better tree topologies according to the maximum-likelihood criterion using the mtREV model of amino acid substitution based on mitochondrial proteins (Adachi and Hasegawa 1996a). The best topologies found for every alignment

TABLE 2

Phylogenetic analysis of mtDNA-encoded proteins

Type of alignment[b]	No. of sites	No. of constant sites	[(hemi, ech), chor][a]		[(hemi, chor), ech]		[hemi, (chor, ech)]	
			ln L[c]	BP[d]	Δ ln L ± SE	BP	Δ ln L ± SE	BP
<8	2814	852	−54073.4	9012	21.1 ± 16.1	906	30.8 ± 14.5*	82
<4	2520	845	−43011.2	9134	22.6 ± 16.8	863	42.1 ± 13.9*	3
<8 (−KM)	2065	807	−30143.2	6090	6.1 ± 11.7	2413	7.6 ± 11.1	1497
<4 (−KM)	1928	800	−25661.5	4887	0.1 ± 11.4	4831	11.4 ± 8.9	282

[a] Three different topologies relating hemichordates (hemi), echinoderms (ech), and a cephalochordates-vertebrates clade (chor). The first topology is the preferred one by the maximum-likelihood criterion with all the alignments.

[b] The alignments contain either <8 or <4 contiguous nonconserved sites. In the last two alignments (−KM), all positions with Met and Lys residues were removed.

[c] Log-likelihood values of the preferred topology (ln L), and difference in log likelihood and one standard error of that difference (Δ ln L ± SE) relative to the preferred topology. * The difference is significant.

[d] Bootstrap proportions supporting each of the three topologies in 10,000 replicates calculated by the RELL method.

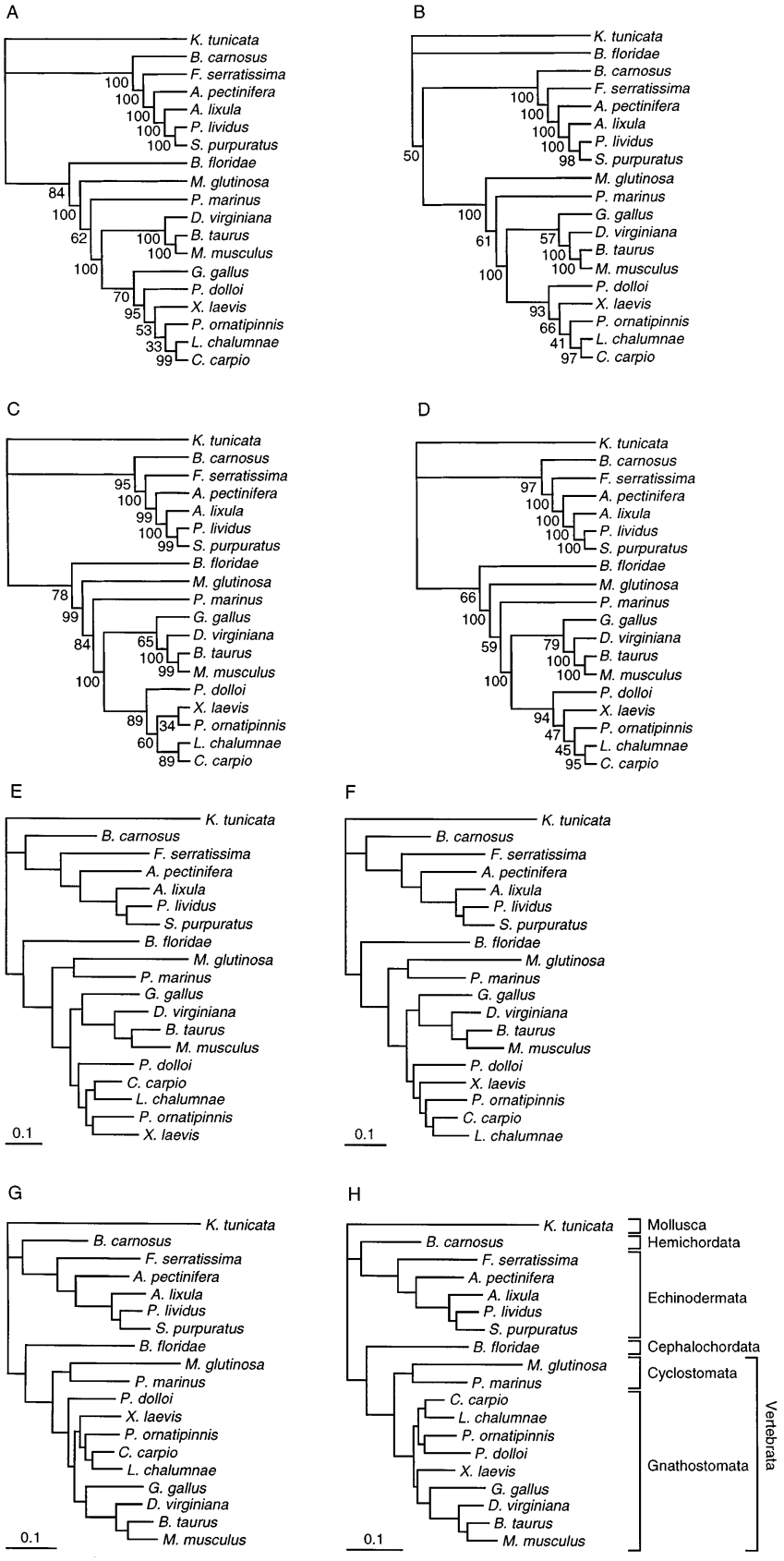

Figure 4.—Phylogenetic trees relating several deuterostomes as estimated from alignments of inferred mtDNA proteins. (A–D) Trees estimated by the neighbor-joining algorithm using Dayhoff distances. Numbers represent local bootstrap values from 100 replicates. (E–H) Trees estimated by maximum likelihood with the mtREV model of amino acid substitution. Horizontal branch lengths are proportional to evolutionary distances. The scale bar represents a distance of 0.1 substitutions per site. When the JTT and Dayhoff models of amino acid substitution were used, the preferred topologies were almost identical to those presented here, with the major exception being that the cephalochordate *B. floridae* was placed with the outgroup in one of the trees using the JTT model. Trees A and E, B and F, C and G, and D and H were calculated from four different alignments, respectively, as explained in the text and in Table 2. Sequences used for constructing the trees (with GenBank accession numbers of the mtDNA sequences in parentheses) are from the following species: the mollusc *Katharina tunicata* (U09810); the cephalochordate *B. floridae* (AF035164–AF035176); the cyclostomes *P. marinus* (U-11880) and *M. glutinosa* (Y15180–Y15191); the ray-finned fishes *Cyprinus carpio* (X61010) and *Polypterus ornatipinnis* (U62532); the lobe-finned fishes *Protopterus dolloi* (L42813) and *Latimeria chalumnae* (U82228); the amphibian *Xenopus laevis* (M10217); the bird *Gallus gallus* (X52392); the mammals *Didelphis virginiana* (Z29573), *Mus musculus* (J01420), and *Bos taurus* (J01394); the sea urchins *P. lividus* (J04815), *S. purpuratus* (X12631), and *A. lixula* (X80396); the starfish *Asterina pectinifera* (D16387); the crinoid *Florometra serratissima* (AF049132); and the hemichordate *B. carnosus* (AF051097).

are shown in Figure 4 (E–H). They all show the hemichordate *B. carnosus* to group with echinoderms. To further investigate the phylogenetic position of hemichordates, the preferred maximum-likelihood topology found for every alignment was compared with two alternative topologies for the interrelationship of the three deuterostome groups. Table 2 shows the differences in log likelihood relative to the preferred topology and the relevant bootstrap proportions supporting each topology. Only two of the alignments allow one of the alternative topologies to be rejected and support the preferred topology by >90% of bootstrap value. Therefore, although the sequence analyses favor the sister group status of hemichordates and echinoderms, alternative phylogenetic positions of the hemichordate cannot be statistically rejected using these alignments.

With respect to other taxa, the neighbor-joining trees (Figure 4, A–D) deviate from the expected morphological trees (*e.g.*, Young 1981) in three main aspects: tetrapods are not monophyletic, cyclostomes (the lamprey, *Petromyzon marinus*, and the hagfish, *Myxine glutinosa*) are always paraphyletic, and the cephalochordate amphioxus (*Branchiostoma floridae*) is not at the base of vertebrates in one of the trees. Furthermore, the relation of tetrapods to fishes is very weakly supported. In most of these aspects, the maximum-likelihood trees are closer to the expected morphological trees. For example, they all have the cephalochordates at the base of vertebrates and cyclostomes as a monophyletic group. However, the tetrapods are monophyletic in only one maximum-likelihood tree, and the relation of tetrapods to different fish groups is unstable.

Previous studies based on mitochondrial protein-coding genes failed to recover the expected position of cephalochordates at the base of the vertebrates (Naylor and Brown 1997, 1998). However, the results presented here may indicate that the use of inferred amino acid sequences and a mitochondrial model of amino acid substitution (Adachi and Hasegawa 1996a) may allow a better estimation of the phylogenetic relationships of these divergent taxa. In another analysis of inferred mitochondrial proteins (Rasmussen *et al.* 1998), cyclostomes were shown to be paraphyletic, in contrast with the prevalent morphological view. The cyclostome paraphyly is seen here in the neighbor-joining trees (Figure 4, A–D), but not in the maximum-likelihood trees (Figure 4, E–H). Further work is obviously necessary to resolve these issues. This is particularly clear in view of the fact that the generally accepted tetrapod monophyly fails to be recovered in most of the trees, something that has also been observed when vertebrate mitochondrial proteins (Nei 1996) and 18S RNA sequences of several animals were considered (Philippe *et al.* 1994). The reason for this is unclear, but it may involve changes in the patterns and rates of substitution in different lineages (Philippe *et al.* 1994; Nei 1996).

In addition to the phylogenetic analyses of the inferred protein sequences, other features of the *B. carnosus* mtDNA strongly support its phylogenetic association with echinoderms. For example, the hemichordate shares aspects of its genetic code, sequence motifs in the control region, the close similarity of the two $tRNA^{Leu}$ genes, and the presence of a N-terminal extension of ND5 with echinoderms. In contrast, the overall order of the mtDNA genes in hemichordates is quite similar to that of vertebrates, while it shows substantial differences to that of echinoderms. Thus, while a large number of gene rearrangements has occurred in echinoderms, both hemichordates and vertebrates have maintained a mitochondrial gene order similar to the one that existed in their common ancestor >600 mya.

We thank Dr. W. Schartau for primer synthesis, Dr. R. Ueshima for his help, and the Deutsche Forschungsgemeinschaft (Pa452/4-1) for financial support.

LITERATURE CITED

Adachi, J., and M. Hasegawa, 1996a Model of amino acid substitution in proteins encoded by mitochondrial DNA. J. Mol. Evol. **42:** 459–468.

Adachi, K., and M. Hasegawa, 1996b MOLPHY version 2.3: programs for molecular phylogenetics based on maximum likelihood. Comput. Sci. Monogr. **28:** 1–150.

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

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.

Ayala, F. J., A. Rzhetsky and F. J. Ayala, 1998 Origin of the metazoan phyla: molecular clocks confirm paleontological estimates. Proc. Natl. Acad. Sci. USA **95:** 606–611.

Benson, D. A., M. S. Boguski, D. J. Lipman, J. Ostell and B. F. Ouellette, 1998 GenBank. Nucleic Acids Res. **26:** 1–7.

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., 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., S. Yokobori, M. Mörl, M. Dörner and S. Pääbo, 1997 RNA editing in metazoan mitochondria: staying fit without sex. FEBS Lett. **409:** 320–324.

Brusca, R. C., and G. J. Brusca, 1990 *Invertebrates*. Sinauer Associates, Sunderland, MA.

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

Castresana, J., G. Feldmaier-Fuchs and S. Pääbo, 1998 Codon reassignment and amino acid composition in hemichordate mitochondria. Proc. Natl. Acad. Sci. USA **95:** 3703–3707.

Christianson, T., and M. Rabinowitz, 1983 Identification of multiple transcriptional initiation sites on the yeast mitochondrial genome by in vitro capping with guanylyltransferase. J. Biol. Chem. **258:** 14025–14033.

Dayhoff, M. O., R. M. Schwartz and B. C. Orcutt, 1978 A model of evolutionary change in proteins, pp. 345–352 in *Atlas of Protein Sequence Structure*, edited by M. O. Dayhoff. National Biomedical Research Foundation, Washington, DC.

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.

Delarbre, C., V. Barriel, S. Tillier, P. Janvier and G. Gachelin, 1997 The main features of the craniate mitochondrial DNA between the *ND1* and the *COI* genes were established in the common ancestor with the lancelet. Mol. Biol. Evol. **14:** 807–813.

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

Elliott, D. J., and H. T. Jacobs, 1989 Mutually exclusive synthetic pathways for sea urchin mitochondrial rRNA and mRNA. Mol. Cell. Biol. **9:** 1069–1082.

Fearnley, I. M., and J. E. Walker, 1986 Two overlapping genes in bovine mitochondrial DNA encode membrane components of ATP synthase. EMBO J. **5:** 2003–2008.

Felsenstein, J., 1993 *PHYLIP (Phylogeny Inference Package), Version 3.5c.* Distributed by the author, Department of Genetics, University of Washington, Seattle.

Gee, H., 1996 *Before the Backbone: Visions on the Origin of Vertebrates.* Chapman & Hall, London.

Halanych, K. M., 1995 The phylogenetic position of the pterobranch hemichordates based on 18S rDNA sequence data. Mol. Phylogenet. Evol. **4:** 72–76.

Hann, S. R., M. W. King, D. L. Bentley, C. W. Anderson and R. N. Eisenman, 1988 A non-AUG translational initiation in c-myc exon 1 generates an N-terminally distinct protein whose synthesis is disrupted in Burkitt's lymphomas. Cell **52:** 185–195.

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.

Holland, P. W., A. M. Hacker and N. A. Williams, 1991 A molecular analysis of the phylogenetic affinities of *Saccoglossus cambrensis* Brambell & Cole (Hemichordata). Philos. Trans. R. Soc. Lond. B Biol. Sci. **332:** 185–189.

Hyman, L. H., 1959 *The Invertebrates: Smaller Coelomate Groups. Chaetognatha, Hemichordata, Pogonophora, Phoronida, Ectoprocta, Brachiopoda, Sipunculida, the Coelomate Bilateralia.* McGraw-Hill, New York.

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

Jacobs, H. T., S. Asakawa, T. Araki, K. Miura, M. J. Smith *et al.*, 1989a Conserved tRNA gene cluster in starfish mitochondrial DNA. Curr. Genet. **15:** 193–206.

Jacobs, H. T., E. R. Herbert and J. Rankine, 1989b Sea urchin egg mitochondrial DNA contains a short displacement loop (D loop) in the replication origin region. Nucleic Acids Res. **17:** 8949–8965.

Jones, D. T., W. R. Taylor and J. M. Thornton, 1992 The rapid generation of mutation data matrices from protein sequences. Comput. Appl. Biosci. **8:** 275–282.

Kishino, H., and M. Hasegawa, 1989 Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in hominoidea. J. Mol. Evol. **29:** 170–179.

Kishino, H., T. Miyata and M. Hasegawa, 1990 Maximum likelihood inference of protein phylogeny and the origin of chloroplasts. J. Mol. Evol. **31:** 151–160.

Kocher, T. D., W. K. Thomas, A. Meyer, S. V. Edwards, S. Pääbo *et al.*, 1989 Dynamics of mitochondrial DNA evolution in animals: amplification and sequencing with conserved primers. Proc. Natl. Acad. Sci. USA **86:** 6196–6200.

L'Abbe, D., B. F. Lang, P. Desjardins and R. Morais, 1990 Histidine tRNA from chicken mitochondria has an uncoded 5′-terminal guanylate residue. J. Biol. Chem. **265:** 2988–2992.

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.

Naylor, G. J., and W. M. Brown, 1997 Structural biology and phylogenetic estimation. Nature **388:** 527–528.

Naylor, G. J., and W. M. Brown, 1998 Amphioxus mitochondrial DNA, chordate phylogeny, and the limits of inference based on comparisons of sequences. Syst. Biol. **47:** 61–76.

Nei, M., 1996 Phylogenetic analysis in molecular evolutionary genetics. Annu. Rev. Genet. **30:** 371–403.

Nielsen, C., 1995 *Animal Evolution.* Oxford University Press, Oxford.

Osawa, S., and T. H. Jukes, 1989 Codon reassignment (codon capture) in evolution. J. Mol. Evol. **28:** 271–278.

Peabody, D. S., 1989 Translation initiation at non-AUG triplets in mammalian cells. J. Biol. Chem. **264:** 5031–5035.

Philippe, H., A. Chenuil and A. Adoutte, 1994 Can the Cambrian explosion be inferred through molecular phylogeny? Development (Suppl.): 15–25.

Rasmussen, A. S., A. Janke and U. Arnason, 1998 The mitochondrial DNA molecule of the hagfish (*Myxine glutinosa*) and vertebrate phylogeny. J. Mol. Evol. **46:** 382–388.

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

Schaefer, B., 1987 Deuterostome monophyly and phylogeny, pp. 179–235 in *Evolutionary Biology*, Vol. 21, edited by M. K. Hecht, B. Wallace and G. T. Prance. Plenum Press, New York.

Shadel, G. S., and D. A. Clayton, 1993 Mitochondrial transcription initiation. Variation and conservation. J. Biol. Chem. **268:** 16083–16086.

Shadel, G. S., and D. A. Clayton, 1997 Mitochondrial DNA maintenance in vertebrates. Annu. Rev. Biochem. **66:** 409–435.

Sprinzl, M., T. Hartmann, J. Weber, J. Blank and R. Zeidler, 1989 Compilation of tRNA sequences and sequences of tRNA genes. Nucleic Acids Res. **17**(Suppl.)**:** 1–172.

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.

Turbeville, J. M., J. R. Schulz and R. A. Raff, 1994 Deuterostome phylogeny and the sister group of the chordates: evidence from molecules and morphology. Mol. Biol. Evol. **11:** 648–655.

Tzeng, C. S., C. F. Hui, S. C. Shen and P. C. Huang, 1992 The complete nucleotide sequence of the *Crossostoma lacustre* mitochondrial genome: conservation and variations among vertebrates. Nucleic Acids Res. **20:** 4853–4858.

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.

Willmer, P., 1990 *Invertebrate Relationships. Patterns in Animal Evolution.* Cambridge University Press, Cambridge.

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

Yokobori, S., and S. Pääbo, 1995a Transfer RNA editing in land snail mitochondria. Proc. Natl. Acad. Sci. USA **92:** 10432–10435.

Yokobori, S., and S. Paabo, 1995b tRNA editing in metazoans. Nature **377:** 490.

Yokobori, S., and S. Pääbo, 1997 Polyadenylation creates the discriminator nucleotide of chicken mitochondrial tRNA(Tyr). J. Mol. Biol. **265:** 95–99.

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.

Young, J. C., 1981 *The Life of Vertebrates.* Oxford University Press, Oxford.

Zardoya, R., and A. Meyer, 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.

Communicating editor: W. Stephan