Complete Sequences of the Highly Rearranged Molluscan Mitochondrial Genomes of the Scaphopod *Graptacme eborea* and the Bivalve *Mytilus edulis*

Jeffrey L. Boore,† Monica Medina,† and Lewis A. Rosenberg*[1]

*Department of Biology, University of Michigan, Ann Arbor; and †DOE Joint Genome Institute and Lawrence Berkeley National Laboratory, Walnut Creek, California

We have determined the complete sequence of the mitochondrial genome of the scaphopod mollusk *Graptacme eborea* (14,492 nts) and completed the sequence of the mitochondrial genome of the bivalve mollusk *Mytilus edulis* (16,740 nts). (The name *Graptacme eborea* is a revision of the species formerly known as *Dentalium eboreum*.) *G. eborea* mtDNA contains the 37 genes that are typically found and has the genes divided about evenly between the two strands, but *M. edulis* contains an extra *trnM* and is missing *atp8*, and it has all genes on the same strand. Each has a highly rearranged gene order relative to each other and to all other studied mtDNAs. *G. eborea* mtDNA has almost no strand skew, but the coding strand of *M. edulis* mtDNA is very rich in G and T. This is reflected in differential codon usage patterns and even in amino acid compositions. *G. eborea* mtDNA has fewer noncoding nucleotides than any other mtDNA studied to date, with the largest noncoding region only 24 nt long. Phylogenetic analysis using 2,420 aligned amino acid positions of concatenated proteins weakly supports an association of the scaphopod with gastropods to the exclusion of Bivalvia, Cephalopoda, and Polyplacophora, but it is generally unable to convincingly resolve the relationships among major groups of the Lophotrochozoa, in contrast to the good resolution seen for several other major metazoan groups.

Introduction

Complete mitochondrial (mt) DNA sequences have been reported for nine mollusks representing four classes: the polyplacophoran *Katharina tunicata* (Boore and Brown 1994*a*, 1994*b*); the cephalopod *Loligo bleekeri* (Tomita et al. 2002); the bivalves *Crassostrea gigas* (Kim, Je, and Park 1999), *Venerupis* (*Ruditapes*) *philippinarum* (Okazaki and Ueshima 2002), and *Inversidens japanensis* (incomplete by only a small fragment of apparently noncoding sequence; Okazaki and Ueshima 2001); and the gastropods *Cepaea nemoralis* (Terrett, Miles, and Thomas 1996), *Albinaria coerulea* (Hatzoglou, Rodakis, and Lecanidou 1995), *Roboastra europaea* (Grande et al. 2002), and *Pupa strigosa* (Kurabayashi and Ueshima 2000a). Additionally, enough of the mtDNA sequence of another gastropod, *Euhadra herkotsi* (Yamazaki et al. 1997), and of the bivalve *Mytilus edulis* (F-type, see below) (Hoffmann, Boore, and Brown 1992) have been determined to assess the arrangement of all encoded genes. Like the mitochondrial genomes of nearly all metazoan animals (see Boore 1999), these are small, circular DNA molecules containing almost universally the same 37 genes: 13 for protein subunits of oxidative phosphorylation enzymes (*cox1-3*, *nad1-6*, *nad4L*, *cob*, *atp6*, *atp8*), two for mitochondrial ribosomal RNAs (*rrnS* and *rrnL*), and 22 for the tRNAs necessary to translate these 13 proteins (designated by *trnX*, with *X* being the one-letter code for the corresponding amino acid, and the two for each of serine and for leucine differentiated by numeral). *M. edulis* varies from this by the loss of *atp8* and the gain of a second *trnM*. *M. edulis* is unusual also in maintaining two different mtDNAs, one called F-type and one M-type (Hoeh, Blakley, and Brown 1991), transmitted by an unusual system dubbed "doubly-uniparental inheritance" (Stewart et al. 1995), which now appears to be widespread in bivalves (Passamonti and Scali 2001; Hoeh, Stewart, and Guttman 2002). Partial mtDNA sequences are also available for the bivalve *M. californianus* (also F-type; Beagley, Okimoto, and Wolstenholme 1999) and the gastropods *Albinaria turrita* (Lecanidou, Douris, and Rodakis 1994), *Omalogyra atomus* (Kurabayashi and Ueshima 2000b), *Littorina saxatilis* (Wilding, Mill, and Grahame 1999) and several vermetids (Rawlings, Collins, and Bieler 2001).

For some phyla of animals, mitochondrial gene arrangements seem seldom to have changed. For example, with few notable exceptions, those vertebrates studied have identical gene arrangements, as do most studied arthropods. (In the latter case, exceptions include the highly rearranged mtDNAs found for wallaby louse [Shao, Campbell, and Barker 2001], hermit crab [Hickerson and Cunningham 2000], and metastriate ticks [Black and Roehrdanz 1998; Campbell and Barker 1998].) Mollusks differ, with many gene rearrangements noted for the molluscan taxa listed above. High levels of rearrangement have also been noted for nematodes and brachiopods (see Boore 1999, 2002). Gene rearrangements have been shown to be very powerful characters for reconstructing evolutionary relationships (see Boore and Brown 1998), and the rapidity of rearrangement within a lineage determines the level at which rearrangements are likely to be phylogenetically informative.

The phylogenetic relationships among the different extant molluscan classes are not well established, and anatomical studies have proposed multiple alternatives to this issue. A common proposal is a gradist scenario where chitons (Polyplacophora) and solenogasters and caudofoveates (Aplacophora) are the basal lineages to a grade of valve-bearing taxa (Gastropoda, Bivalvia, Cephalopoda, Monoplacophora, and Scaphopoda), collectively known as

[1] Present address: Medical Scientist Training Program, Department of Genetics, Case Western Reserve University School of Medicine.

Key words: scaphopod, bivalve, mollusk, mitochondria, evolution, genome.

E-mail: jlboore@lbl.gov.

Mol. Biol. Evol. 21(8):1492–1503. 2004
doi:10.1093/molbev/msh090
Advance Access publication March 10, 2004

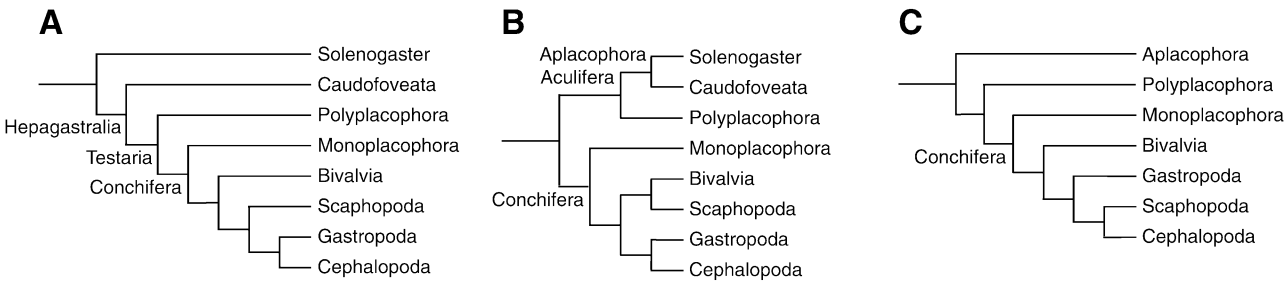

FIG. 1.—A comparison of the most commonly held views on the evolutionary relationships of molluscan classes.

the Conchifera (Salvini-Plawen 1985; Salvini-Plawen and Steiner 1996; Haszprunar 2000) (fig. 1*A*). Some propose that chitons and aplacophorans form a monophyletic clade rather than a grade (Scheltema 1993, 1996), and some view Conchifera as split into a cephalopod/gastropod clade and a scaphopod/bivalve clade with monoplacophorans as the basal conchiferan lineage (Runnegar and Pojeta 1974) (each as in fig. 1*B*). Recent evaluations of morphological and paleontological data (Waller 1998) as well as 18S rRNA sequences (Steiner and Dreyer 2002) alternatively conclude that Scaphopoda is the sister group to Cephalopoda (fig. 1*C*). The many ribosomal RNA sequences have so far only poorly resolved molluscan phylogeny, rendering some taxa paraphyletic (e.g., bivalves; Steiner and Müller 1996), making it difficult to assess whether the proposed anatomical interpretations are identifying true synapomorphies (Steiner and Müller 1996; Steiner and Hammer 2000). Consequently, we are in need of additional characters that can help address phylogenetic relationships among major molluscan lineages. As the mitochondrial genome database continues to grow, we will be able to incorporate both gene order and sequence data into this analysis. Here we present the first complete mitochondrial genome sequence for a member of the Scaphopoda (*Graptacme eborea*), a previously unsampled class, and the completed sequence of the F-type mitochondrial genome of the bivalve *Mytilus edulis*.

Materials and Methods

Determining the mtDNA Sequence of *Graptacme eborea*

A preparation of total DNA prepared from *G. eborea* was the gift of K. Fraser and T. Collins. PCR amplification of a portion of *cox1* with the LCO1490 and HCO2198 oligonucleotides (Folmer et al. 1994) employed an initial heating at 97°C for 5 min, a hot start by adding Taq polymerase (Fisher) at 72°C, then 36 cycles of: (94°C, 1 min; 45°C, 1 min; 72°C, 2 min), followed by incubation at 72°C for 15 min. Reaction volume was 25 µl, which then yielded a single band of approximately 710 nts on a 1% agarose gel stained with ethidium bromide. This amplification product was gel purified using Gene Clean (Bio 101) and ligated into pBluescript (Stratagene), which had been prepared by: digestion with *Eco*RV, tailing on each end with a single T using Taq polymerase and dTTP, then gel purification using Gene Clean (Bio 101). DNA was prepared from recombinant plasmids using alkaline lysis followed by organic extraction and ethanol precipitation (Sambrook, Fritsch, and Maniatis 1989). DNA sequence was determined using AmplitaqFS, dye-terminator cycle sequencing (Perkin-Elmer) according to supplier's instructions and an ABI 377 Automated DNA Sequencer. This sequence was verified as the expected portion of the *cox1* gene by comparison to the homologous regions of *Katharina tunicata* mtDNA (Boore and Brown 1994*a*).

A set of two oligonucleotide primers was designed facing "out" from this fragment, matching to positions that are separated by only 27 nts of the *G. eborea cox1* sequence. These were used to amplify 14,465 nt, nearly the entire mtDNA, in a single reaction. This PCR used rTth-XL polymerase (Perkin-Elmer) with 1.3 mM MgOAc, and was otherwise performed according to supplier's instructions. Reaction volume was 100 µl and conditions were 94°C for 45 sec, followed by 37 cycles of: (94°C, 10 sec; 55°C, 20 sec; 65°C, 12 min, with an additional 15 sec per cycle after the 16th), then an incubation at 72°C for 12 min. An aliquot yielded a single band on a 1% agarose gel stained with ethidium bromide.

Approximately 2 µg of this product was digested separately with the restriction enzymes *Mbo*I and *Taq*I, each of which recognizes 4-nt sites. Several fragments were selected from each digest and gel purified as above, then they were ligated into the compatible *Bam*HI and *Cla*I sites, respectively, of pBluescript plasmid (Stratagene), followed by DNA preparation and sequence determination as above. Additional oligonucleotide primers were designed for determining the sequence "out" from each of these cloned fragments. The 14,465 nt PCR product was passed three times through an Ultrafree Spin Column (30,000 NMWL; Millipore) to eliminate amplification primers and PCR reagents and then used directly as a template for sequencing reactions as above. Using a combination of oligonucleotides matching the ends of the amplified fragments with those matching internal sequences obtained from the cloned *Mbo*I and *Taq*I fragments greatly reduced the time required to "primer walk" through this fragment. All sequence was determined in both directions.

Completing the mtDNA Sequence of *Mytilus edulis*

Most (13.9 kb) of the F-type mtDNA sequence of *M. edulis* has been previously reported (Hoffmann, Boore,

Table 1
Species Used in the Phylogenetic Analysis with Current Taxonomic Classification and Mitochondrial Genome Accession Numbers

Taxon	Classification	Accession Number
Outgroup taxa		
Metridium senile	Cnidaria, Anthozoa	NC_003522
Asterina pectinifera	Ehinodermata, Eleutherozoa	NC_001627
Florometra serratisima	Echinodermata, Crinoidea	NC_001878
Paracentrotus lividus	Echinodermata, Eleutherozoa	NC_001572
Balanoglossus carnosus	Hemichordata, Enteropneusta	NC_001887
Branchiostoma lanceolatum	Chordata, Cephalochordata	NC_001912
Homo sapiens	Chordata, Craniata	NC_001807
Daphnia pulex	Arthropoda, Crustacea	NC_000844
Pagurus longicarpus	Arthropoda, Crustacea	NC_003058
Tetrodontophora bielanensis	Arthropoda, Hexapoda	NC_002735
Drosophila melanogaster	Arthropoda, Hexapoda	NC_001709
Locusta migratoria	Arthropoda, Hexapoda	NC_001712
Limulus polyphemus	Arthropoda, Chelicerata	NC_003057
Mollusks		
Katharina tunicata	Mollusca, Polyplacophora	NC_002010
Loligo bleekeri	Mollusca, Cephalopoda	NC_002507
Crassostrea gigas	Mollusca, Bivalvia, Pteriomorphia	NC_001816
Inversidens japanensis (f)	Mollusca, Bivalvia, Paleoheterodonta	AB055625
Inversidens japanensis (m)	Mollusca, Bivalvia, Paleoheterodonta	AB055624
Venerupis philippinarum	Mollusca, Bivalvia, Heteroconchia	NC_003354
Mytilus edulis	Mollusca, Bivalvia, Pteriomorphia	This study
Albinaria coerulea	Mollusca, Gastropoda, Pulmonata	NC_001761
Pupa strigosa	Mollusca, Gastropoda, Opisthobranchia	NC_002176
Graptacme eborea	Mollusca, Scaphopoda, Dentaliida	This study
Other lophotrochozoans		
Laqueus rubellus	Brachiopoda, Rhynchonelliformea	NC_002322
Terebratalia transversa	Brachiopoda, Rhynchonelliformea	NC_003086
Terebratulina retusa	Brachiopoda, Rhynchonelliformea	NC_000941
Lumbricus terrestris	Annelida, Clitellata	NC_001673
Platynereis dumerilii	Annelida, Polychaeta	NC_000931

and Brown 1992). Although this was sufficient to determine the gene content and arrangement, it omitted the sequences of the central portions of many genes. To complete this, we designed oligonucleotide primers for PCR that match the ends of the previously reported sequences and used these to amplify the undetermined portions using DNA preparations of the appropriate *M. edulis* clones (Hoffmann, Boore, and Brown 1992) as templates. Each PCR reaction yielded a single band on a 1% agarose gel when visualized by ethidium bromide staining and UV irradiation. DNA was purified and the DNA sequence was determined as for *G. eborea*, using the amplifying or internal primers as necessary. All sequence was determined in both directions and was assembled with that previously reported into a complete mtDNA sequence.

Gene Annotation and Gene Order Comparison

Protein-encoding genes of each mtDNA were identified by sequence similarity of open reading frames to mitochondrial gene sequences of *Katharina tunicata* (Boore and Brown 1994*a*). Ribosomal RNA genes were identified by sequence similarity and potential secondary structures. As a class, tRNA genes were identified by their potential to form tRNA-like secondary structures; specific identifications were made according to anticodon sequence.

A search for shared gene arrangements was conducted against all mitochondrial sequence data available in GenBank that included sequence from three or more genes (3,376 entries). This search employed a PERL script that decomposed the query genome into all binary gene arrangements, searched for shared gene orders, and then reassembled any overlapping pairs for each comparison.

Phylogenetic Analysis of Protein Data

We included 27 taxa in the phylogenetic analysis (table 1), 15 of which are mollusks or other lophotrochozoans and 12 of which are metazoan outgroups (five non-lophotrochozoan protostomes, six deuterostomes, and one cnidarian). (One taxon, *Inversidens japanensis*, is represented by two sets of sequences, one F-type and one M-type.) We performed multiple sequence alignments for each protein using the pileup program in the GCG package. Each alignment was then refined by eye and subsequently combined into a concatenated data set. Because *atp8* is missing in several of the taxa, it was excluded from all analyses. Regions of ambiguous alignment were also excluded; table 2 shows the regions corresponding to each gene in the concatenated alignment, the total number of positions per protein, and the number of amino acid sites included in the final analysis. Maximum parsimony (MP) reconstructions were conducted with

Table 2
Number of Amino Acid Positions Used in the Final Analysis

Protein	Alignment position	Total positions	Included positions
Atp6	1-258	258	117
Atp8	259-330	72	0
Cob	331-755	425	358
Cox1	756-1313	558	495
Cox2	1314-1830	517	214
Cox3	1831-2122	292	253
Nad1	2123-2480	358	246
Nad2	2481-2881	401	115
Nad3	2882-3022	141	61
Nad4	3023-3539	517	241
Nad4L	3540-3676	137	28
Nad5	3677-4349	673	272
Nad6	4350-4558	209	20
Total	4558	4558	2420

PAUP*4.0b (Swofford 2001), with branch support estimated from 1,000 bootstrap pseudoreplicates. Quartet-puzzling (QP) was performed with Tree-Puzzle (Strimmer and Haeseler 1997) using both the mtREV24 and Blosum62 models with 100,000 quartet-puzzling steps, with a gamma correction and eight rate categories, and estimating amino acid frequencies from the data set. Bayesian reconstructions (MB) used MrBayes 3.0 (Huelsenbeck and Ronquist 2001). Exploratory Markov Chain Monte Carlo runs were performed starting with different amino acid substitution priors (i.e., mixed models, Poisson). Subsequently, we ran the heated MCMC chain for 1,000,000 generations, which was sampled every 100 updates using the models with higher posterior probabilities (mtREV and Blossum) from the mixed model prior. We discarded 1,000 cycles as burn-in before estimating joint posterior probabilities. We also analyzed each gene individually using MP and QP as above, and MB with mixed and mtRev amino acid models. The Nexus-formatted file of the alignment is available as online Supplementary Material.

Results and Discussion

Gene Content and Organization

The mitochondrial genome of *G. eborea* (GenBank accession number AY484748) contains the 37 genes most commonly found in animal mtDNAs (Boore 1999), including *atp8*, which is missing in *M. edulis* (complete sequence deposited in GenBank as accession number AY484747). (*atp8* is also missing from the mtDNAs of other bivalves [Kim, Je, and Park 1999; Okazaki and Ueshima 2001], secernentean nematodes [Okimoto et al. 1992], platyhelminths [Le et al. 2000], and chaetognaths [unpublished data].) Genes are divided between the two strands about evenly, with one having 12 tRNA- and five protein-encoding genes and the other having 10 tRNA-, eight protein-, and two rRNA-encoding genes (fig. 2). (All genes are on the same strand for *M. edulis* mtDNA.) On the strand reading clockwise in figure 2, there are two regions of co-oriented gene clusters: (1) from *atp6* to *trnH* (eight genes) and (2) from *trnA* to *trnW* (six genes). In addition, this strand contains three tRNA genes (*trnG*, *trnR*, and *trnS2*) that punctuate the gene arrangement on the opposite strand. On the strand reading counter-clockwise in figure 2, there are two regions of gene clusters: (1) from *nad4L* to *trnF* (six genes) and (2) from *trnV* to *cox2* (14 genes), except as interrupted by *trnG*, *trnR*, and *trnS2*.

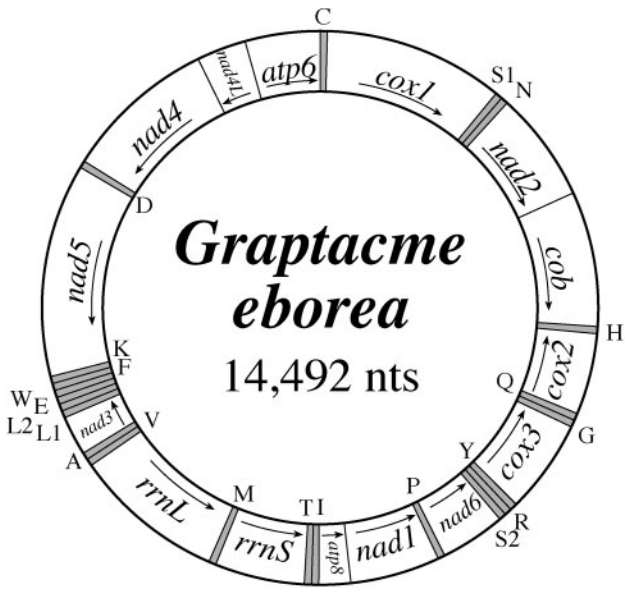

FIG. 2.—Gene map of the mitochondrial genome of the scaphopod mollusk *Graptacme eborea*. Genes have standard abbreviations except for tRNAs, which are designated by the one-letter code for the corresponding amino acid. S1, S2, L1, and L2 designate genes for those tRNAs recognizing the codons AGN, UCN, CUN, and UUR, respectively. Transfer RNA genes shown outside of the circle are transcribed clockwise while those inside are transcribed counter-clockwise. Transcriptional orientation for each of the other genes is indicated with an arrow. Gene scaling is only approximate.

As is typical of animal mtDNAs, genes start with nonstandard initiation codons, but with no consistency of usage among the genes (table 3). Similarly, there is no pattern evident in the use of complete vs. abbreviated stop codons.

In the few cases where it has been studied, mitochondrial genes are expressed as a polycistron, and then the tRNAs are enzymatically removed to liberate the flanking gene specific mRNAs. *G. eborea* mtDNA contains only four cases where non-tRNA genes abut: *nad2-cob*, *atp8-nad1*, *nad4L-nad4*, and *nad4L-atp6*. For this last case, the genes are on opposite strands. We infer that *nad2* and *nad4L* end on abbreviated stop codons (T and TA, respectively; see fig. 3) that would be completed by polyadenylation, but it is not clear how cleavage would occur to end these mRNAs at these nucleotides. The genes for *atp8* and *nad1* are separated by 14 nt, raising the same issue. It has been speculated that other secondary structures can substitute for tRNAs for message cleavage (e.g., Boore and Brown 1994*a*), but no potential secondary structures are obvious at these gene boundaries. Further, it is not clear by what alternative mechanism transcript cleavage could occur at the 3′ end of *cox2*, *cox3*, or *nad6* since these are flanked by tRNA genes on the opposite strand.

Table 3
Comparisons of the Mitochondrial Protein Coding Genes of Four Mollusks, the Scaphopod *Graptacme eborea* (*Geb*), Bivalve *Mytilus edulis* (*Med*), Polyplacophoran *Katharina tunicata* (*Ktu*), and the Gastropod *Albinaria coerulea* (*Aco*)

	Number of Amino Acids				Initiation and Termination Codons			
Protein	*Geb*[a]	*Med*	*Ktu*	*Aco*	*Geb*	*Med*	*Ktu*	*Aco*
Atp6	225	238	230	214	ATG / TAA	ATG / TAG	ATG / TAA	ATG / T*
Atp8	53	[b]—	53	55	ATG / TAA	[b]—	ATG / TAG	ATG / TAG
Cox1	512	551	513	509	GTG / TAA	ATG / TAA	ATG / T*	TTG / TA*
Cox2	223	242	229	224	ATG / T*	ATG / TAG	ATG / TAG	ATG / TAA
Cox3	259	264	259	259	ATG / TAA	ATG / [c]—	ATG / TAA	ATG / TAA
Cob	379	397	379	367	ATG / T*	ATG / TAA	ATG / TAA	ATA / TA*
Nad1	294	305	316	299	ATT / T*	GTG / TA*	ATG / TAA	ATG / TAA
Nad2	331	314	338	307	ATG / T*	ATG / TAG	GTG / TAG	ATG / TAA
Nad3	116	116	120	117	ATG / TAA	ATG / TAA	ATG / TAA	ATA / T*
Nad4	443	435	442	437	ATT / TAA	ATG / TAA	ATA / T*	ATG / TAA
Nad4L	96	93	100	99	ATT / TAA	ATG / TAA	ATG / TAG	ATG / T*
Nad5	559	568	571	545	ATG / T*	ATA / TA*	ATG / T*	ATT / TAG
Nad6	163	158	166	155	ATG / TAG	ATG / TAA	ATG / T*	ATG / TAA

NOTE.—For *Mytilus edulis* see Hoffmann, Boore, and Brown (1992), and also this study; for *Katharina tunicata* see Boore and Brown (1994*a*); and for *Albinaria coerulea* see Hatzoglou, Rodakis, and Lecanidou (1995). *Albinaria* was arbitrarily selected to represent the three gastropods for which complete mtDNA sequence is available. *An asterisk indicates that the predicted stop codon is incomplete, presumably completed to UAA by polyadenylation of the mRNA; in a few cases there is an alternative possibility of a complete stop codon overlapping the adjacent gene by a few nts.

[a] Due to four ambiguities of nucleotide identity, none of which led to an ambiguous amino acid inference, the total number of amino acids here exceeds the number of codons in table 4 by four.

[b] No *atp8* gene is present in *M. edulis* mtDNA.

[c] As interpreted, *cox3* of *M. edulis* ends on an incomplete stop codon; however, the reading frame is open for another 48 codons until a TAA stop codon, but there is no similarity between this extension and other mtDNA sequences.

There are several cases where genes seem to overlap on the same strand. The *cox1* gene is interpreted to overlap the downstream *trnS1* by six nts. The only reasonable alternative is that *cox1* ends on an abbreviated stop codon; the minimal truncation to allow this would be by 24 nucleotides. The seven amino acids that would then not be coded at the carboxyl end are not identical to those of any other animal, so this is possible, but they are chemically similar to the amino acids commonly found at the carboxyl end of animal Cox1 proteins. Secondly, *trnY* overlaps *trnR* by three nts, GGA, which match well in each of these two tRNAs (fig. 4). The third case is for the pair *trnR-trnS2*, which overlaps by a single T; perhaps the tRNA(R) discriminator nucleotide is not encoded in the genomic DNA. Overlapping genes could not be resolved into separate, complete messages from a single polycistron, so this would imply that (1) these genes are transcribed from multiple promoters, (2) there is differential cleavage to generate sometime one or sometimes the other complete RNA, or (3) complete RNAs are restored by some type of post transcriptional mechanism.

The nucleotides CTAA depicted as being part of *trnL1* in figure 4 do not pair well with the 5′ end of the tRNA and overlap with the downstream gene, *trnL2*. It is possible, alternatively, that these genomic nucleotides are not part of the *trnL1* gene, but that the necessary nucleotides to complete the tRNA are added posttranscriptionally after tRNA(L2) is cleaved from the transcript. This could, potentially, be done by an RNA-dependent RNA polymerase, using the 5′ end of the tRNA as a template, as seems likely for centipede mt-tRNAs (Lavrov, Brown, and Boore 2000). However, because the potentially edited nucleotides would all be A's (to match the T's at the 5′ end), this might be completed by a simpler mechanism, tRNA polyadenylation, as is the case for some mt-tRNAs (Yokobori and Pääbo 1995). Otherwise, there is little reason to suppose this as a common process, since the amino-acyl acceptor stems are well-matched for the most 3′ sequences for every other tRNA except tRNA(S2), which has a single mismatched TT terminal pair. This mismatch could be corrected by tRNA polyadenylation, although this gene (as inferred with the mismatch) does not overlap any other.

There have been extensive, unique rearrangements involving nearly every gene (fig. 5) of both of these mtDNAs. *G. eborea* and *M. edulis* mtDNAs have only a few gene boundaries in common with any other animal studied to date. An examination of all 3,376 entries in GenBank of sequences having three or more mitochondrial genes reveals that *G. eborea* shares the arrangement *rrnL*, *trnM*, *rrnS* with the Yesso scallop, *Mizuhopecten yessoensis* (GenBank accession AB052599), and *nad1*, *P*, *nad6* with the squid, *Loligo bleekeri*. The first is an interesting potential synapomorphy that would exclude, among the sampled mollusks, only the polyplacophoran *Katharina tunicata* (Boore and Brown 1994*a*) and the cephalopod *Loligo bleekeri* (Sasuga et al. 1999); others have these genes in autapomorphic arrangements. It is interesting that the sampled gastropods have the arrangement *trnM*, *rrnS*, although *rrnL* is elsewhere. The inferred basal group, Polyplacophora, is represented by *K. tunicata*, which has an arrangement similar to the second case, *nad1*, *-P*, *nad6*; the other studied mollusks have further rearrangements of these genes. The same analysis of *M. edulis* mtDNA reveals that it shares the arrangement *trnL1*, *trnL2*, *nad1* with the

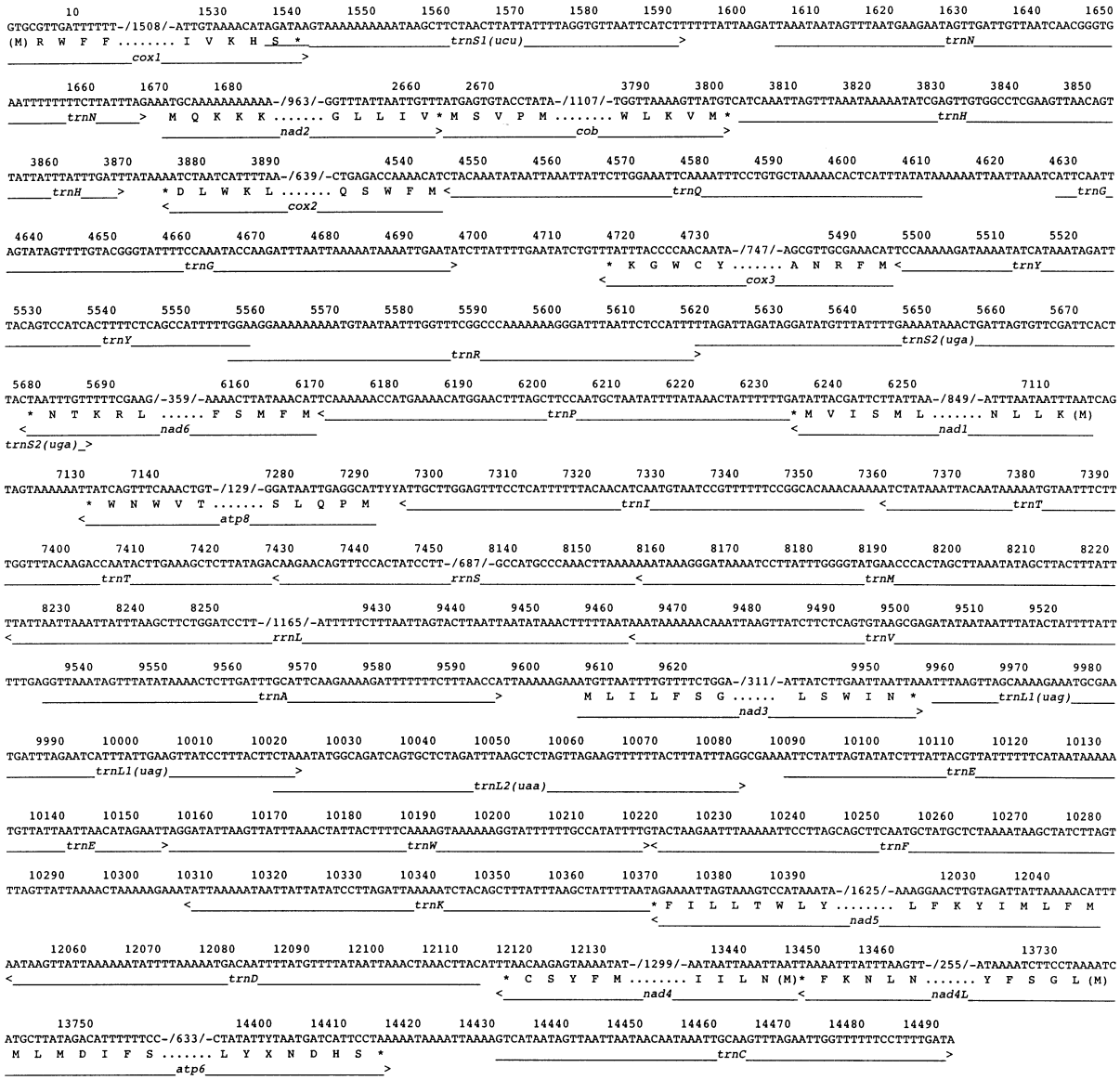

FIG. 3.—A greatly-abbreviated, schematic representation of the mtDNA sequence of *Graptacme eborea*. Numbers within the slash marks indicate omitted nucleotides. Translation for each gene is assumed to initiate with formyl-methionine, so for those genes starting with other than ATR (*cox1*, *nad1*, *nad4*, and *nad4L*), the initiator M is shown in parenthesis to indicate presumed nonconformity with the generally employed genetic code. Stop codons, including those inferred to be "abbreviated," are marked by an asterisk. A dart (>) marks the last nucleotide of each gene and indicates the direction of transcription.

hemichordate *Balanoglossus carnosus* (Castresana et al. 1998) and with *K. tunicata*. Other than these arrangements, neither mtDNA shares more than a single gene boundary with any other animal.

Transfer RNAs

G. eborea mtDNA contains the typical complement of 22 tRNAs. Most have potential to fold into a normal cloverleaf structure, although tRNA(R) and tRNA(S1) do not have paired DHU arms, and a few others have a mismatched nt pair. All have T immediately preceding and R following the anticodon. tRNAs have been described previously for *M. edulis* (Hoffmann, Boore, and Brown 1992).

Mutational Bias Is Reflected in Codon Usage Patterns and Protein Amino Acid Composition

G. eborea mtDNA is 14,492 bp in length and is 74.1% A+T, very high even for a mitochondrial genome. Strand skew measures (Perna and Kocher 1995) for the distribution of GC pairs [(G−C)/(G+C)] and TA pairs [(T−A)/(T+A)] between the two strands are nearly zero (–0.02 and +0.002, respectively).

M. edulis mtDNA is 16,740 bp in length and is 61.8% A+T. GC skew is +0.246 and TA skew is +0.110,

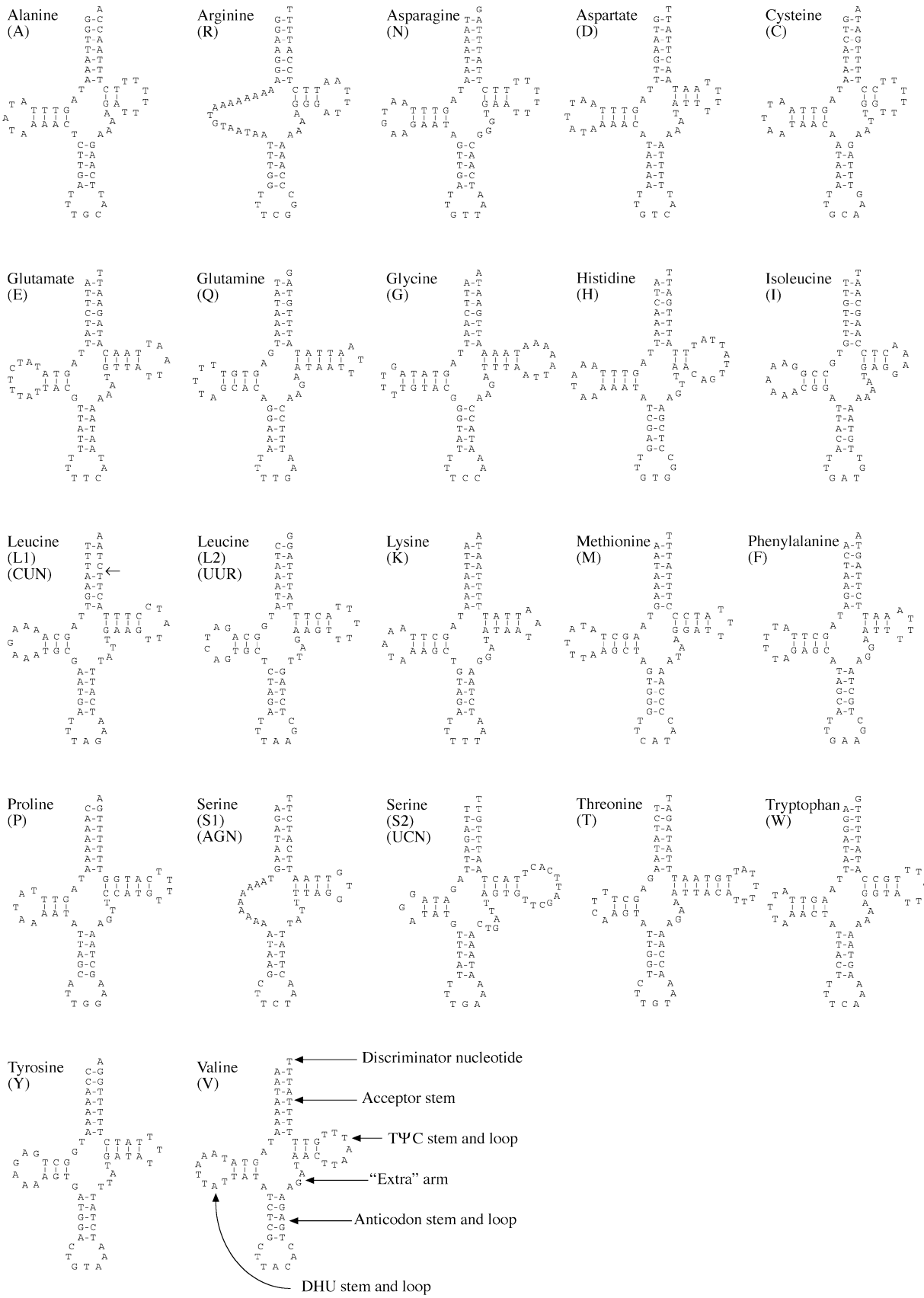

FIG. 4.—The potential secondary structures of the 22 inferred tRNAs of *Graptacme eborea* mtDNA. Nomenclature for portions of tRNA structures is shown for tRNA(V). Codons recognized are shown for the pairs of leucine and serine tRNAs. The arrow on tRNA(L1) marks the beginning of overlap with the downstream gene *trnL2*.

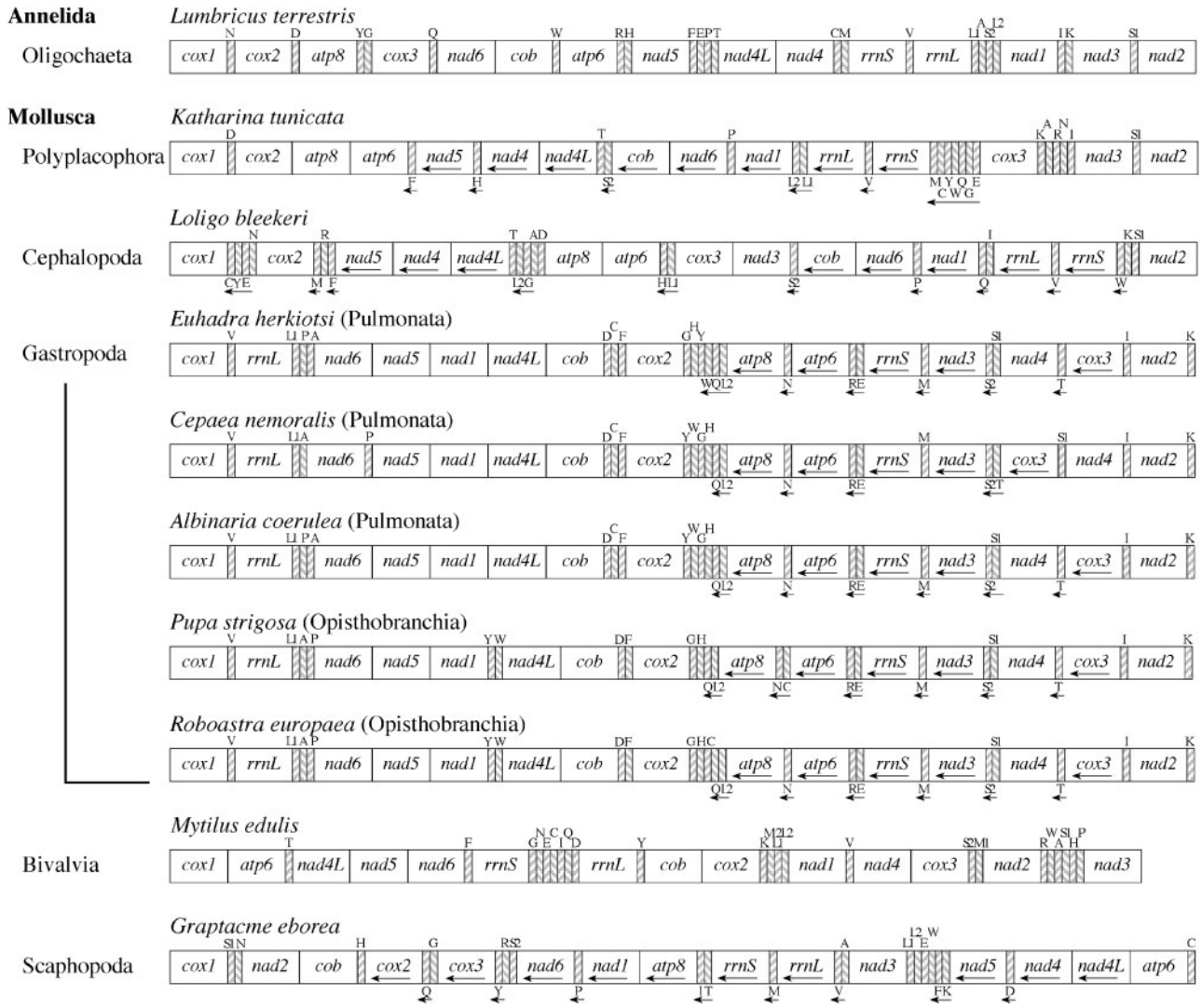

FIG. 5.—The gene map of *Graptacme eborea* mtDNA compared with those of all other mollusks whose complete mt gene arrangement has been published and with that of an outgroup, the annelid *Lumbricus terrestris* (Boore and Brown 1995). Genes are abbreviated as in figure 2. All genes are transcribed from left-to-right except for those with a left-facing arrow.

indicating that the strand containing the genes is quite rich in G and T relative to the other strand. This bias is very evident in comparisons of synonymous codon usage pattern between the two genomes (table 4); for every case where an amino acid can be specified by any NNR codon, *M. edulis* has a much greater proportion of NNG:NNA relative to *G. eborea*.

As has been found for many mtDNAs (Cardon et al. 1994), CpG is the least frequent dinucleotide for these two mtDNAs, both in absolute number and relative to expected frequency calculated from the percentage of C and G in the genomes. For both mtDNAs, all four homodimers are significantly more common than would occur by chance; these values are the four highest except for GC in both mtDNAs and CT for *M. edulis* alone.

The amino acid leucine can be specified by six different codons (TTR and CTN) and the proteins of these two mtDNAs have a very similar number of leucines. As reflects the mutation bias, *M. edulis* has a much greater proportion of leucines specified by TTG and CTG codons, mainly at the expense of TTA codons. The amino acid serine can be specified by eight different codons (TCN and AGN); the proteins of the two mtDNAs also contain similar numbers of this amino acid and, again, the distribution reflects the mutation bias of *M. edulis* toward higher G and T. All AGN codons are used much more frequently in *M. edulis*, especially AGG; TCG usage is also elevated, all at the expense of TCT, TCC, and TCA codons.

This bias is also reflected in patterns of amino acid substitutions between these two mtDNAs. *G. eborea* and *M. edulis* mitochondrial proteins contain nearly identical numbers of nonpolar (A, V, L, I, P, M, F, W) (2,082 and 2,020, respectively) and polar (G, S, T, C, Y, N, Q) (1,170 and 1,219, respectively) amino acids. For nonpolar amino acids, *M. edulis* proteins use many more alanines (GCN) and valines (GTN) at the expense of isoleucine (ATY),

methionine (ATR), and phenylalanine (TTY). For polar amino acids, *M. edulis* proteins contain more glycine (GGN) at the expense of asparagine (AAY). Presumably, the bias toward G and T in the gene-containing strand of *M. edulis* has resulted in amino acid replacements within the tolerance of physio-chemical similarity.

Unassigned DNA

G. eborea mtDNA is very uncommon for lacking any large noncoding regions, as are usually inferred to contain the origin(s) of replication and transcription control signals. The largest noncoding region is only 24 nt between *trnK* and *trnF*. Next in size are the 19 nt gap between *cox3* and *trnG* and the 18 nt between *trnG* and *trnQ*. Noncoding DNA of *M. edulis* mtDNA has been analyzed and described earlier (Hoffmann, Boore, and Brown 1992). There is no obvious conservation of either nucleotide identities or potential secondary structures between the mollusks' noncoding regions. Whatever regulatory elements may be present are apparently short, dispersed, and/or rapidly changing.

Phylogenetic Analysis

Figure 6*A* presents a 70% majority rule consensus tree of MP bootstrap analysis for the taxa outlined in table 1, which has a topology congruent with those from quartet puzzling (QP) and some of the Bayesian analysis. These analyses support the monophyly of the lophophorates, annelids, and brachiopods. Relationships among the major molluscan lineages, however, remain unresolved, as are those among mollusks, brachiopods, and annelids, despite using this relatively large data set of 2,420 confidently aligned amino acid positions. In contrast, relationships among the major groups of deuterostomes and of arthropods are well resolved and conforming to expectation from other analyses, bolstering the view that the relationships among the lophotrochozoan groups are especially difficult to resolve.

Results of MB analyses with different sets of prior probabilities for the amino acid model are depicted in figure 6, panels *B–E*. Concerns over the inability to resolve many of these phylogenetic relationships are exacerbated by the observation that the Bayesian analyses (unlike the MP and QP analyses) return substantially different topologies with high posterior probability values when the prior amino acid substitution models vary.

However, this data set of concatenated protein sequences does give consistent results for support of many metazoan clades regardless of the type of analysis performed (table 5). In contrast, analyses using each individual gene recovered only the deuterostome and protostome nodes with high levels of confidence (>70%) in six cases and not all methods were consistent. The arthropod clade was recovered by only three individual genes and by only one or two methods in the best of cases. The Lophotrochozoa clade was recovered only by Nad2 and only in the case of using the mixed model prior analysis.

Table 4
Codon Usage for *Graptacme eborea* (*Geb*; 3,649 Codons) and *Mytilus edulis* (*Med*; 3,681 Codons) mtDNAs

Amino Acid	Codon	*Geb*	*Med*
Phe (F)	TTT	297	220
(gaa)	TTC	43	69
Leu (L2)	TTA	364	178
(uaa)	TTG	47	104
Leu (L1)	CTT	55	71
(uag)	CTC	6	15
	CTA	71	91
	CTG	6	50
Ile (I)	ATT	299	156
(gau)	ATC	41	46
Met (M)	ATA	241	143
(cau)[a]	ATG	39	76
Val (V)	GTT	79	97
(uac)	GTC	15	32
	GTA	98	132
	GTG	14	125
Ser (S2)	TCT	125	70
(uga)	TCC	31	13
	TCA	69	31
	TCG	8	27
Pro (P)	CCT	71	70
(ugg)	CCC	13	11
	CCA	41	26
	CCG	8	25
Thr (T)	ACT	71	56
(ugu)	ACC	18	14
	ACA	58	38
	ACG	7	15
Ala (A)	GCT	67	74
(ugc)	GCC	15	28
	GCA	53	65
	GCG	6	21
Tyr (Y)	TAT	130	99
(gua)	TAC	28	55
TER[b]	TAA	—	—
	TAG	—	—
His (H)	CAT	68	46
(gug)	CAC	13	27
Gln (Q)	CAA	44	37
(uug)	CAG	13	19
Asn (N)	AAT	144	71
(guu)	AAC	24	42
Lys (K)	AAA	100	69
(uuu)	AAG	16	42
Asp (D)	GAT	60	39
(guc)	GAC	8	28
Glu (E)	GAA	65	45
(uuc)	GAG	13	58
Cys (C)	TGT	44	48
(gca)	TGC	10	30
Trp (W)	TGA	74	42
(uca)	TGG	19	53
Arg (R)	CGT	18	27
(ucg)	CGC	3	12
	CGA	22	24
	CGG	11	19
Ser (S1)	AGT	31	57
(ucu)	AGC	11	28
	AGA	79	88
	AGG	20	75
Gly (G)	GGT	67	66
(ucc)	GGC	11	32
	GGA	87	63
	GGG	40	143

NOTE.—The potential extension of *cox3* of *M. edulis* is not included here. The anticodon of the corresponding tRNA is shown in parentheses.

[a] For *M. edulis* mtDNA, there is a second tRNA for methionine with anticodon UAU.

[b] Stop codons are not included in this analysis.

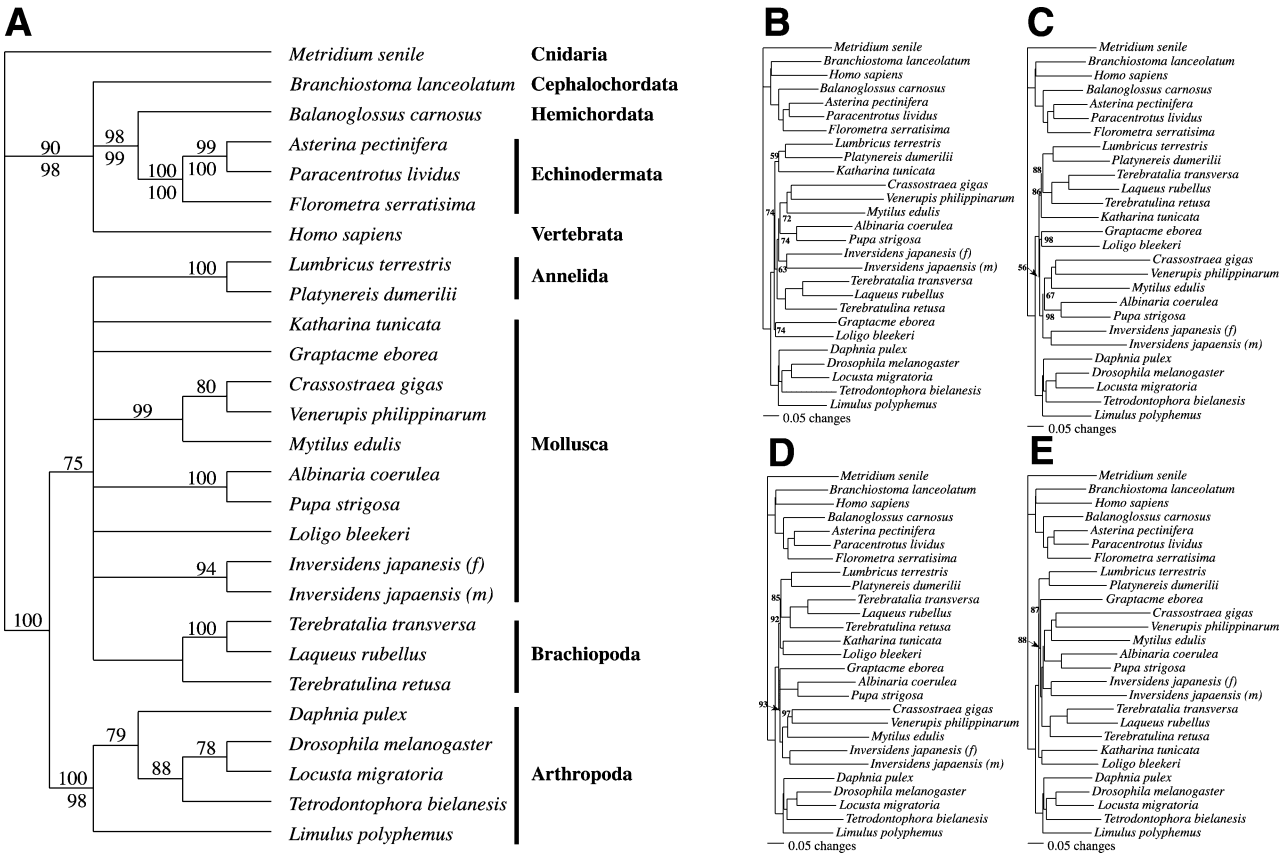

FIG. 6.—Comparisons of various phylogenetic analyses of 2,420 aligned amino acid positions of 12 concatenated mitochondrial proteins. (*A*) 70% majority rule consensus tree of maximum parsimony bootstrap analysis. This tree is congruent with the quartet puzzling (QP) analyses, and the numerals at the nodes are MP bootstrap values followed by QP support from the mtREV24 model analysis (which were very similar to those using Blosum62). Results of Bayesian analyses (50% majority rule consensus) are shown using the Blosum62 (*B*), mtREV24 (*C*), mixed (*D*), and Poisson (*E*) models. Any node unmarked by a numeral has support of 100%.

Although there are key nodes within the Lophotrochozoa that are still unresolved, one hopes that they will yield to further taxon sampling, and these results illustrate the importance of sequencing complete mitochondrial genomes, as opposed to using only single gene sequences, when analyzing deep divergences.

Table 5
Support of Particular Metazoan Clades in Individual Gene (Excluding Atp8) Analyses Using Maximum Parsimony (MP), Quartet Puzzling (QP) with mtREV24, and MrBayes (MB) with a Mixed Model Prior Probability Assignment

Gene	Deuterostomia	Protostomia	Arthropoda	Lophotrochozoa
Atp6	−+−	−−−	+−+	−−−
Cob	−−−	+−−	+−+	−−−
Cox1	−−−	−−−	−−−	−−−
Cox2	+−−	−−−	−−−	−−−
Cox3	−−−	−−−	−−−	−−−
Nad1	+−+	+ + +	+−−	−−−
Nad2	+−−	+−−	−−−	+−−
Nad3	−−−	−−−	−−−	−−−
Nad4	+ + +	+ + +	−−−	−−−
Nad4L	−−−	−−−	−−−	−−−
Nad5	+−−	+ + +	−−−	−−−
Nad6	−−−	+−−	−−−	−−−
All genes	+ + +	+ + +	+ + +	+ + +

NOTE.—Node support of at least 70% is indicated by a plus (+) symbol in analyses, from left to right, using MP, QP, and MB.

Acknowledgments

Thanks to K. Fraser and T. Collins for the gift of *G. eborea* DNA and to B. Dayrat for helpful comments. This work was supported by DEB-9807100 from the National Science Foundation. Part of this work was performed under the auspices of the U.S. Department of Energy, Office of Biological and Environmental Research, by the University of California, Lawrence Berkeley National Laboratory, under contract No. DE-AC03-76SF00098.

Literature Cited

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

Black, W. C., and R. L. Roehrdanz. 1998. Mitochondrial gene order is not conserved in arthropods: Prostriate and metastriate tick mitochondrial genomes. Mol. Biol. Evol. **15**:1772–1785.

Boore, J. L. 1999. Animal mitochondrial genomes. Nucleic Acids Res. **27**:1767–1780.

———. 2002. "Mitochondrial Gene Arrangement Source Guide." (http://www.jgi.doe.gov/programs/comparative/Mito_top_level.html).

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

———. 1994*b*. Mitochondrial genomes and the phylogeny of mollusks. Nautilus **108**(Suppl. 2):61–78.

———. 1995. The complete DNA sequence of the mitochondrial genome of the annelid worm *Lumbricus terrestris*. Genetics **141**:305–319.

———. 1998. Big trees from little genomes: Mitochondrial gene order as a phylogenetic tool. Curr. Opin. Genet. Dev. **8**:668–674.

Campbell, N. J. H., and S. C. Barker. 1998. An unprecedented major rearrangement in an arthropod mitochondrial genome. Mol. Biol. Evol. **15**:1786–1787.

Cardon, L. R., C. Burge, D. A. Clayton, and S. Karlin. 1994. Pervasive CpG suppression in animal mitochondrial genomes. Proc. Natl. Acad. Sci. USA **91**:3799–3803.

Castresana, J., G. Feldmaier-Fuchs, S.-I. Yokobori, N. Satoh, and S. Pääbo. 1998. The mitochodrial genome of the hemichordate *Balanoglossus carnosus* and the evolution of deuterostome mitochondria. Genetics **150**:1115–1123.

Folmer, O., M. Black, W. Hoeh, R. Lutz, and R. Vrijenhoek. 1994. DNA primers for amplification of mitochondrial cytochrome *c* oxidase subunit I from diverse metazoan invertebrates. Mol. Mar. Biol. Biotech. **3**:294–299.

Grande, C., J. Templado, J. L. Cervera, and R. Zardoya. 2002. The complete mitochondrial genome of the nudibranch *Roboastra europaea* (Mollusca: Gastropoda) supports the monophyly of opisthobranchs. Mol. Biol. Evol. **19**:1672–1685.

Haszprunar, G. 2000. Is the Aplacophora monophyletic? A cladistic point of view. American Malacological Bulletin **15**: 115–130.

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

Hickerson, M. J., and C. W. Cunningham. 2000. Dramatic mitochondrial gene rearrangements in the Hermit Crab *Pagurus longicarpus* (Crustacea, Anomura). Mol. Biol. Evol. **17**:639–644.

Hoeh, W. R., K. H. Blakley, and W. M. Brown. 1991. Heteroplasmy suggests limited biparental inheritance of Mytilus mitochondrial DNA. Science **251**:1488–1490.

Hoeh, W. R., D. T. Stewart, and S. I. Guttman. 2002. High fidelity of mitochondrial genome transmission under the doubly uniparental mode of inheritance in freshwater mussels (Bivalvia: Unionoidea). Evolution **56**:2252–2261.

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.

Huelsenbeck, J. P., and F. Ronquist. 2001. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics **17**:754–755.

Kim, S.-H., E.-Y. Je, and D.-W. Park. 1999. *Crassostrea gigas* mitochondrial DNA. GenBank accession number AF177226.

Kurabayashi, A., and R. Ueshima. 2000*a*. Complete sequence of the mitochondrial DNA of the primitive opisthobranch *Pupa strigosa*: Systematic implications of the genome organization. Mol. Biol. Evol. **17**:266–277.

———. 2000*b*. Partial mitochondrial genome organization of the heterostrophan gastropod *Omalogyra atomus* and its systematic significance. Venus (Jap. J. Malac.) **59**:7–18.

Lavrov, D. V., W. M. Brown, and J. L. Boore. 2000. A novel type of RNA editing occurs in the mitochondrial tRNAs of the centipede *Lithobius forticatus*. Proc. Natl. Acad. Sci. USA **97**:13738–13742.

Le, T. H., D. Blair, T. Agatsuma et al. (14 co-authors). 2000. Phylogenies inferred from mitochondrial gene orders-a cautionary tale from the parasitic flatworms. Mol. Biol. Evol. **17**:1123–1125.

Lecanidou, R., V. Douris, and G. Rodakis. 1994. Novel features of metazoan mtDNA revealed from sequence analysis of three mitochondrial DNA segments of the land snail *Albinaria turrita* (Gastropoda: Clausiliidae). J. Mol. Evol. **38**: 369–382.

Okazaki, M., and R. Ueshima. 2001. Evolutionary diversity between the gender-associate mitochondrial DNA genomes of freshwater mussels. Genbank accession numbers AB055624 (male haplotype) and AB055625 (female haplotype).

———. 2002. Gender-associated mtDNA of *Tapes philippinarum*. Genbank accession number NC_003354.

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

Passamonti, M., and V. Scali. 2001. Gender-associated mitochondrial DNA heteroplasmy in the venerid clam *Tapes philippinarum* (Mollusca Bivalvia). Curr. Genet. **39**:117–124.

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

Rawlings, T., T. Collins, and R. Bieler. 2001. A major mitochondrial gene rearrangement among closely related species. Mol. Biol. Evol. **18**:1604–1609.

Runnegar, B., and J. Pojeta. 1974. Molluscan phylogeny: The paleontological viewpoint. Science **186**:311–317.

Salvini-Plawen, L. V. 1985. Early evolution and the primitive groups. Pp. 59–150 *in* E. R. Trueman and M. R. Clark, eds. Evolution. Academic Press, Orlando, Fla.

Salvini-Plawen, L. V., and G. Steiner. 1996. Synapomorphies and plesiomorphies in higher classification of Mollusca. Pp. 29–51 *in* J. Taylor, ed. Origin and evolutionary radiation of the Mollusca. Oxford University Press, Oxford.

Sambrook, J., E. F. Fritsch, and T. Maniatis. 1989. Molecular cloning, a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Sasuga, J., S.-I. Yokobori, M. Kaifu, T. Ueda, K. Nishikawa, and K. Watanabe. 1999. Gene contents and organization of a mitochondrial DNA segment of the squid *Loligo bleekeri*. J. Mol. Evol. **48**:692–702.

Scheltema, A. H. 1993. Aplacophora as progenetic aculiferans and the coelomatic origin of mollusks as the sister taxon of Sipuncula. Biol. Bull. **184**:57–78.

———. 1996. Phylogenetic position of Sipuncula, Mollusca and the progenetic Aplacophora. Pp. 53–58 *in* J. Taylor, ed. Origin and evolutionary radiation of the Mollusca. Oxford University Press, Oxford.

Shao, R., N. J. H. Campbell, and S. C. Barker. 2001. Numerous gene rearrangements in the mitochondrial genome of the wallaby louse, *Heterodoxus macropus* (Phthiraptera). Mol. Biol. Evol. **18**:858–865.

Steiner, G., and H. Dreyer. 2002. Scaphopoda and Cephalopoda are sister taxa—an evolutionary scenario. Abstract for the American Malacological Society Annual Meeting, Charleston, SC.

Steiner, G., and S. Hammer. 2000. Molecular phylogeny of the Bivalvia inferred from 18S rDNA sequences with particular reference to the Pteriomorpha. Pp. 11–29 *in* E. M. Harper, J. D. Taylor, and J. A. Crame, eds. The evolutionary biology of the Bivalvia. The Geological Society of London Special Publications, London.

Steiner, G., and M. Müller. 1996. What can 18S rDNA do for bivalve phylogeny? J. Mol. Evol. **43**:58–70.

Stewart, D. T., C. Saavedra, R. R. Stanwood, A. O. Ball, and E. Zouros. 1995. Male and female mitochondrial DNA lineages in the blue mussel (*Mytilus edulis*) species group. Mol. Biol. Evol. **12**:735–747.

Strimmer, K., and A. von Haeseler. 1997. Puzzle. Maximum likelihood analysis for nucleotide and amino acid alignments. Zoologisches Institut, Munchen, Germany.

Swofford, D. L. 2001. PAUP*: phylogenetic analysis using parsimony (*and other methods). Beta Version 4.0b8. Sinauer Associates, Sunderland, Mass.

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

Tomita, K., S. Yokobori, T. Oshima, T. Ueda and K. Watanabe. 2002. The cephalopod *Loligo bleekeri* mitochondrial genome: Multiplied noncoding regions and transposition of tRNA genes. J. Mol. Evol. **54**:486–500.

Waller, T. R. 1998. Origin of the molluscan class Bivalvia and a phylogeny of major groups. Pp. 1–47 *in* P. A. Johnston and J. W. Haggart, eds. Bivalves: An eon of evolution—plaeobiological studies honoring Norman D. Newell. University of Calgary Press, Calgary, Canada.

Wilding, C. S., P. J. Mill, and J. Grahame. 1999. Partial sequence of the mitochondrial genome of *Littorina saxatilis*: Relevance to gastropod phylogenetics. J. Mol. Evol. **48**: 348–359.

Yamazaki, N., R. Ueshima, J. Terrett et al. (12 co-authors). 1997. Evolution of pulmonate gastropod mitochondrial genomes: comparisons of gene organizations of Euhadra, Cepaea and Albinaria and implications of unusual tRNA secondary structures. Genetics **145**:749–758.

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

Richard Thomas, Associate Editor

Accepted December 30, 2003