Mitochondrial Genomes of *Galathealinum*, *Helobdella,* and *Platynereis*: Sequence and Gene Arrangement Comparisons Indicate that Pogonophora Is Not a Phylum and Annelida and Arthropoda Are Not Sister Taxa

Jeffrey L. Boore and Wesley M. Brown

Department of Biology, University of Michigan

We report a contiguous region of more than half (>7,500 nt) of the mitochondrial genomes for *Platynereis dumerii* (Annelida: Polychaeta), *Helobdella robusta* (Annelida: Hirudinida), and *Galathealinum brachiosum* (Pogonophora: Perviata). The relative arrangements of all 22 genes identified for *Helobdella* and *Galathealinum* are identical to one another and to their arrangements in the mtDNA of the previously studied oligochaete annelid *Lumbricus.* In contrast, *Platynereis* differs from these taxa in the positions of several tRNA genes and in having two additional tRNA genes (*trnC* and *trnM*) and a large noncoding sequence in this region. Comparisons of relative gene arrangements and of the nucleotide and inferred amino acid sequences among these and other published taxa provide strong support for an annelid-mollusk clade that excludes arthropods, and for the inclusion of pogonophorans within Annelida, rather than giving them separate phylum status. Gene arrangement comparisons include the first use of a recently described method on previously unpublished data. Although a variety of alternative initiation codons are typically used by mitochondrial protein-encoding genes, ATG appears to be the initiator for all but one reported here. The large noncoding region (1,091 nt) identified in *Platynereis* has no significant sequence similarity to the noncoding region of *Lumbricus,* although each contains runs of TA dinucleotides and of homopolymers, which could potentially serve as signaling elements. There is strong bias for synonymous codon usage in *Helobdella* and especially in *Galathealinum.* In this latter taxon, 5 codons are completely unused, 13 are used three or fewer times, and G appears at third codon positions in only 26 of the 2,236 codons. Nucleotide composition bias appears to influence amino acid composition of the proteins.

Introduction

The superphylum Protostomia contains animals united by several shared developmental features (e.g., spiral cleavage, strictly determined cell lineages, mesentoblast-derived mesoderm, blastopore becoming the mouth, and schizocoelous coelom formation). The three largest and best studied protostome phyla are Arthropoda (e.g., arachnids, crustaceans, insects), Annelida (e.g., polychaetes, earthworms, leeches), and Mollusca (e.g., chitons, clams, snails). Based primarily on the view that their segmented body plans are a shared-derived feature, arthropods and annelids have commonly been regarded as the most closely related pair of the three (the Articulata; see Brusca and Brusca 1990). This view has been reinforced by long-standing, elegantly described scenarios (e.g., Snodgrass 1938; Raff and Kaufman 1983) that describe the evolutionary transformation of a contiguous lineage from an annelid into an onychophoran and then from an onychophoran into an arthropod by the hypothesized mechanism of progressively increasing body segment specialization. Although this phyletic progression is heuristically appealing, it is probably incorrect.

Recent studies using morphological characters (e.g., Eernisse, Albert, and Anderson 1992), molecular sequences (e.g., Ghiselin 1988; Lake 1990; Garcia-Machado et al. 1999), and fossil evidence (i.e., the halkieriids; see Morris and Peel 1995) have reinvigorated an alternative view, that mollusks and annelids are the most closely related pair. The primary character uniting this group (the Eutrochozoa; Ghiselin 1988) is the shared presence of a trochophore larval form in some species of each of these phyla, whereas such larvae are unknown among arthropods. Deducing the pattern of evolution among these three protostome phyla hinges largely on the subjective interpretation of whether overt body segmentation or a trochophore larval form is the more reliable phylogenetic character (see Eernisse, Albert, and Anderson 1992).

The phylum Annelida is traditionally divided into three classes: Oligochaeta (e.g., earthworms), Hirudinida (leeches), and Polychaeta (marine annelids). There is strong evidence for uniting the first two as the most closely related pair (the Clitellata; see Rouse and Fauchald 1995 and references therein). The third class, Polychaeta, is the most diverse and speciose, and some have concluded that polychaetes are a paraphyletic assemblage (e.g., McHugh 1997; Kojima 1998).

Another phylum, the Pogonophora (‘‘beard worms’’), are also vermiform animals with a trochophore larva. Pogonophorans live in thin-walled tubes anchored in the ocean’s sediment, often at great depths. Opinions about both the phylogenetic placement and the taxonomic level of pogonophorans have differed widely, but most now regard them as a protostome phylum related to the Annelida (see Rouse and Fauchald 1995 and references therein). However, some molecular sequence comparisons (McHugh 1997) and more recent morphological comparisons (Rouse and Fauchald 1997) have found support for their inclusion as a family within the Annelida. The work presented here supports this latter view, that the Pogonophora are not an independent phylum and should most properly revert to the name Si-

Key words: *Galathealinum, Helobdella, Platynereis, Lumbricus,* mitochondria, evolution, genome.

Address for correspondence and reprints: Jeffrey L. Boore, Department of Biology, University of Michigan, 830 North University Avenue, Ann Arbor, Michigan 48109-1048. E-mail: jboore@umich.edu.

Mol. Biol. Evol. 17(1):87–106. 2000

boglinidae Caullery, 1914, as a family within Annelida in accordance with the proposal by Rouse and Fauchald (1997).

Our proposal of the relationships among annelids, mollusks, arthropods, and pogonophorans is based on the comparisons of the sequences and gene arrangements for homologous segments constituting about 50% of the mitochondrial genomes of the polychaete *Platynereis dumerii*, the hirudinid *Helobdella robusta*, the pogonophoran *Galathealinum brachiosum*, the oligochaete *Lumbricus terrestris* (Boore and Brown 1995), and homologous portions of several published mollusk, arthropod, and chordate species (for a list and summary descriptions of studied mitochondrial genomes, see Boore 1999 and links at http://biology.lsa.umich.edu/∼jboore).

Metazoan mitochondrial genomes are usually unicircular DNA molecules of about 16 kb that encode the same set of 37 genes (for 2 rRNAs [*rns*, *rnl*], 22 tRNAs [*trnX*, with anticodon shown when more than one tRNA specifies the same amino acid], and 13 proteins [*cox1–3, cob, atp6, atp8, nad1–6, nad4L*]). (For historical reasons, these genes are sometimes named differently in animal mitochondrial genomes; see Boore 1999 for a table of synonymous gene names.) This set of 37 genes can potentially be rearranged in an enormous number of combinations, and the large number of different arrangements found among (and occasionally within) metazoan phyla suggest that this character is relatively unconstrained. Major rearrangements of genes, here defined as translocations and/or inversions of one or more multigene tracts, appear to be infrequent on a geological timescale, although minor rearrangements, such as exchanges of position or polarity between neighboring tRNA genes, are encountered with greater frequency. Therefore, with the possible exclusion of some minor rearrangements, identical, convergent rearrangements in independent lineages are highly unlikely, and arrangements promise to be a reliable character to use for determining very ancient relationships, such as those that exist between major taxonomic categories (e.g., phyla, classes) (see Boore and Brown 1998). Indeed, comparisons of mitochondrial gene arrangements have proven especially informative in several recent phylogenetic studies (Smith et al. 1993; Boore et al. 1995; Boore, Lavrov, and Brown 1998).

Materials and Methods

Molecular Analysis

A DNA preparation of the leech *H. robusta* (Annelida: Hirudinida) was a gift of Monica Dixon and David Weisblat, and a DNA preparation of the polychaete *P. dumerii* (Annelida: Polychaeta) was a gift of Daniel Sellos. A DNA preparation of the deep-sea pogonophoran *G. brachiosum* (Pogonophora: Perviata) was a gift of Robert Vrijenhoek and Mike Black; the specimen had been collected on Alvin dive #2798 at the Oregon subduction zone at 45°61.1′N 125°17.4′W at a depth of 2,628 m.

Initially, three small fragments were amplified by PCR from each of *Platynereis*, *Helobdella*, and *Galathealinum* mtDNA using the following oligonucleotide pairs: (1) for a 710-nt fragment of *cox1*, LCO1490 (GGT CAA CAA ATC ATA AAG ATA TTG G) and HCO2198 (TAA ACT TCA GGG TGA CCA AAA AAT CA) (Folmer et al. 1994); (2) for a 540-nt fragment of *cox3*, COIIIF (TGG TGG CGA GAT GTK KTN CGN GA) and COIIIR (ACW ACG TCK ACG AAG TGT CAR TAT CA); and (3) for a 450-nt fragment of *cob*, CytbF (GGW TAY GTW YTW CCW TGR GGW CAR AT) and CytbR (GCR TAW GCR AAW ARR AAR TAY CAY TCW GG) (see fig. 1 for primer placement). Reactions used *Taq* polymerase with supplied buffer (Fisher Scientific or Qiagen); Mg^{++} ion concentration and cycling conditions were optimized as necessary. Each reaction produced a single band when visualized under UV light following ethidium bromide staining on a 1% agarose gel. PCR reaction products were purified by three serial passages through an Ultrafree (30,000 NMWL) spin column (Millipore). This purified DNA (50–300 ng) was used in a dye-terminator cycle sequencing reaction according to manufacturer's (Perkin-Elmer) recommendations. Unincorporated nucleotides were removed by ethanol precipitation, and the purified product was analyzed on an ABI 377 automated DNA sequencer.

Species-specific oligonucleotides were designed on the basis of these sequences in order to amplify much larger fragments using ''long-PCR'' (Barnes 1994). For spanning the region *rnl-cox1*, one conserved primer matching the *rnl* sequence (16S-BL [ACG TGA TCT GAG TTC AGA CCG G]; Palumbi et al. 1991) was used along with species-specific oligonucleotides matching within the determined fragment of *cox1* (PlatCOIR [TCT CCC GAG TAG CGA TCC GGG TTG ACC TAG TTC], HeloCOIR [AAA AAG GAC CCT GGT TGG GCT AGT TCA ATT CGA], or GalaCOIR [AGA AAA GAT CCT GGT TGT CCT AGT TCG AGA CGG]). Pairs of species-specific oligonucleotides were used for amplifying the *cox3-cob* region (PlatCOIIIF [GCT ACG GGC TTT CAT GGG TTA CAT G], HeloCOIIIF [CTG GAT TCC ACG GAG CAC ATG], or GalaCOIIIF]CCA TGG ACT TCA TGT TCT GGT AG], along with PlatCytbR [CTA CTA GCA TTG GCC CGA TAT AGG G], HeloCytbR [ACC CGC CTC AAA TTC ACT CTA C], or GalaCytbR [AGC TCC AAT ATA GGG AAT AGC TG], respectively). For *Helobdella* and *Galathealinum*, the *cox1–cox3* amplification used HeloCOIF (TTT GAT CCT GTT GGA GGT GGA GAC CCA GTA CT) or GalaCOIF (CTT TGA TCC TAG AGG AGG TGG TGA TCC TGT TCT) with HeloCOIIIR (CCG AAG AAG AAG CAA ATT TCA G) or GalaCOIIIR (CCT ATA GGA GGT CAA GTA CAT CC), respectively.

The corresponding fragment was very difficult to amplify in *Platynereis*, with the reactions generally yielding multibanded products, perhaps because of the presence of signaling elements within the large noncoding region (see Shadel and Clayton 1997). This region was amplified in shorter, overlapping fragments by using oligonucleotides designed to sequences conserved among the other annelid species (WormCOIF-3′ [TAC

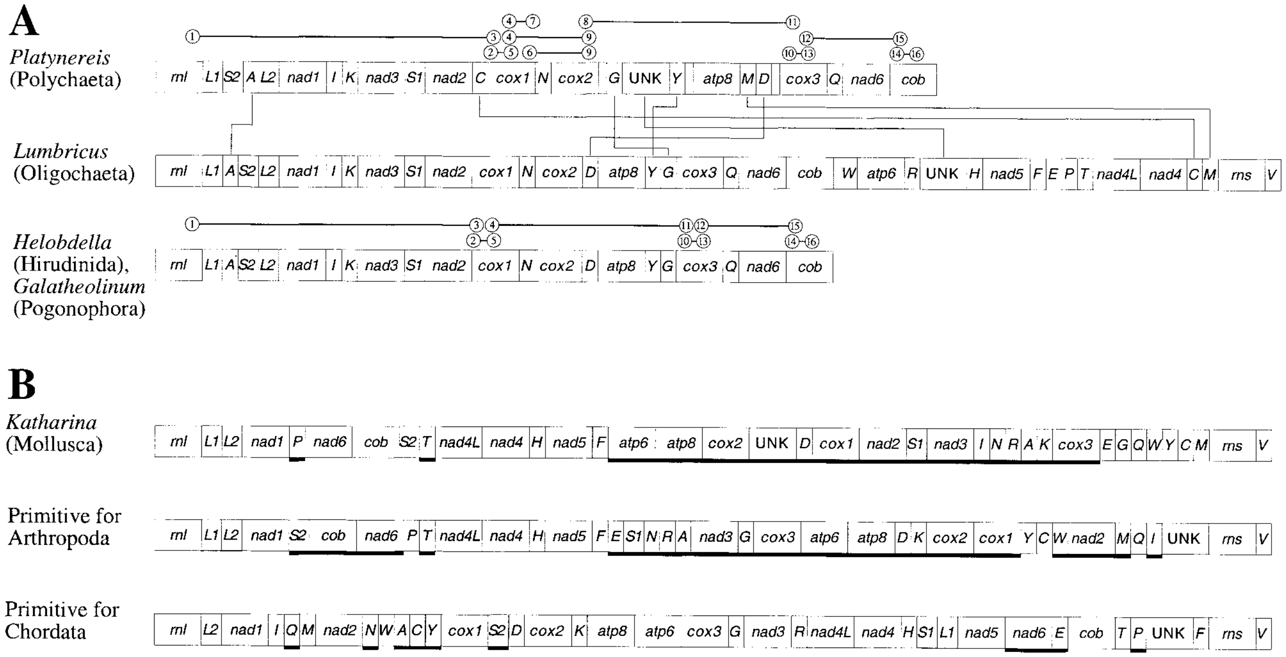

A
Platynereis (Polychaeta)
rnl L1 S2 A L2 nad1 I K nad3 S1 nad2 C cox1 N cox2 G UNK Y atp8 M D cox3 Q nad6 cob
1 2-5 3 4 4-7 6 8 9 10-13 11 12 14-16 15
Lumbricus (Oligochaeta)
rnl L1 A S2 L2 nad1 I K nad3 S1 nad2 cox1 N cox2 D atp8 Y G cox3 Q nad6 cob W atp6 R UNK H nad5 F E P T nad4L nad4 C M rns V
Helobdella (Hirudinida), Galatheolinum (Pogonophora)
rnl L1 A S2 L2 nad1 I K nad3 S1 nad2 cox1 N cox2 D atp8 Y G cox3 Q nad6 cob
1 2-5 3 4 10-13 11 12 14-16 15
B
Katharina (Mollusca)
rnl L1 L2 nad1 P nad6 cob S2 T nad4L nad4 H nad5 F atp6 atp8 cox2 UNK D cox1 nad2 S1 nad3 I N R A K cox3 E G Q W Y C M rns V
Primitive for Arthropoda
rnl L1 L2 nad1 S2 cob nad6 P T nad4L nad4 H nad5 F E S1 N R A nad3 G cox3 atp6 atp8 D K cox2 cox1 Y C W nad2 M Q I UNK rns V
Primitive for Chordata
rnl L2 nad1 I Q M nad2 N W A C Y cox1 S2 D cox2 K atp8 atp6 cox3 G nad3 R nad4L nad4 H S1 L1 nad5 nad6 E cob T P UNK F rns V

FIG. 1.—*A,* Schematic representation of the partial mitochondrial genomes of the polychaete *Platynereis dumerii,* the leech *Helobdella robusta,* and the pogonophoran *Galathealinum brachiosum* compared with the complete mitochondrial genome of the oligochaete *Lumbricus terrestris* (Boore and Brown 1995). Genes are not to scale, are all transcribed in the same orientation (i.e., left to right as depicted), and are abbreviated as in the text, except that tRNA genes are designated only by the one-letter code for the corresponding amino acid, with L1/L2 and S1/S2 designating the tRNA pairs recognizing the leucine codons CUN/UUR (anticodons tag/taa) and the serine codons AGN/UCN (anticodons tct/tga), respectively. "UNK" designates the largest noncoding (i.e., "unknown") region, and each other noncoding region larger than 30 nt is shown by a small shaded box. Lines connecting homologous genes show differences between the arrangements of *Platynereis* and *Lumbricus.* Lines above the genome depictions show the extents of various PCR amplification products. These are flanked by circled numerals representing the amplifying primers according to the following key: (1) 16S-BL; (2) LCO1490; (3) PlatCOIR, HeloCOIR, or GalaCOIR; (4) PlatCOIF, HeloCOIF, or GalaCOIF; (5) HCO2198; (6) WormCOIF-3′; (7) WormCOIR-3′; (8) PlatCOIF; (9) WormCOIIR; (10) COIIIF; (11) PlatCOIIIR, HeloCOIIIR, or GalaCOIIIR; (12) PlatCOIIIF, HeloCOIIIF, or GalaCOIIIF; (13) COIIIR; (14) CytbF; (15) PlatCytbR, HeloCytbR, or GalaCytbR; (16) CytbR. See *Materials and Methods* for sequences of these primers. *B,* The complete mitochondrial gene arrangements of the other animals included in this phylogenetic analysis, that of the mollusk *Katharina tunicata* (Boore and Brown 1994*a,* 1994*b*) and those inferred to be primitive for Arthropoda (identical to the arrangement of a chelicerate mtDNA; Staton, Daehler, and Brown 1997; Boore, Lavrov, and Brown 1998; Boore 1999) and for Chordata (identical to the arrangement of fish mtDNA (Chang, Huang, and Lo 1994), with the exception of one tRNA translocation (Boore, Daehler, and Brown 1999; Boore 1999). Underlining designates opposite (i.e., right to left as depicted) transcriptional orientation.

TAC GTA GTA GCA CAC TTT CAC TA], WormCOIR-3′ [TAR TCT GAG TAT CGT CGD GGT ATT CC], and WormCOIIR [GCT CCG CAA ATT TCT GAA CAT TGT CC]) or specific to sequence obtained (PlatCOIF [CCG AAA CCT AAA CAC TGC GTT CTT TGA TCC TGC], PlatCOIIF [GTG TAC TAG TAT CGG CTG CTG ACG], and PlatCOIIIR [GCA CTC TAA ATG GGT TGA TAG GGG TC]). Each of these amplifications that did not include the large noncoding region gave single-banded products in good quantity with minimum optimization efforts; however, the primer pair flanking this region (PlatCOIIF and PlatCOIIIR) continued to produce a multibanded product. One of these bands was of significantly greater quantity than the others and corresponded in size to the length of sequence obtained by using sequencing primers internal to the amplifying primers. This is evidence that this band corresponds to the actual mtDNA sequence, along with the presence within it of the expected genes. Although it is possible that the multiple bands were legitimate products representing multiple states of the mtDNA genome, the supernumerary bands were neither consistent in appearance among various attempts nor of regular size variation as expected of tandem repeat sequences that have been identified in the large noncoding regions of some mtDNAs.

All long-PCR reactions used rTth XL polymerase (Perkin-Elmer) with supplied buffer. Reactions were optimized for Mg^{++} concentration and cycling conditions as required. Each of these amplification products was purified and sequenced as above with additional oligonucleotides obtained as necessary (Gibco-BRL) for primer walking through each fragment. All nucleotides were determined on both strands except for a few short regions that were less than 200 nt in length, within 400 nt of the sequencing primer, and without any hint of ambiguity on the sequenced strand.

Sequence Analysis

Sequences were produced and assembled using the ABI suite of programs (e.g., Sequencing Analysis, Sequence Navigator, Autoassembler). Subsequent manipulations used MacVector 6.5 and GCG (Oxford Molecular Group).

Amino acid sequences were inferred for all protein-encoding genes determined for *Galathealinum, Helob-*

della, and *Platynereis,* along with the homologous genes of *L. terrestris* (Annelida: Oligochaeta) (Boore and Brown 1995), *Katharina tunicata* (Mollusca: Polyplacophora) (Boore and Brown 1994*a*, 1994*b*), *Artemia fransiscana* (Arthropoda: Crustacea) (Valverde et al. 1994), *Drosophila yakuba* (Arthropoda: Hexapoda) (Clary and Wolstenholme 1985), *Cyprinus carpio* (Chordata) (Chang, Huang, and Lo 1994), and *Squalus acanthias* (Chordata) (Rasmussen and Arnason 1999) using the genetic code for *Drosophila* or vertebrate mtDNA, as appropriate. All proteins were inferred to initiate with formyl-methionine regardless of the DNA sequence of the designated start codon (Smith and Marcker 1968). Each protein-encoding gene and ribosomal RNA gene was easily identified by comparison with homologs in *Lumbricus* mtDNA. Transfer RNA genes were identified generically by their potential secondary structures and specifically by anticodon sequence.

Phylogenetic Analysis of Inferred Protein Sequences

Sequences were aligned using CLUSTAL W, as implemented in MacVector 6.5 (Oxford Molecular Group); the BLOSSUM matrix was used to weight shared amino acids, with gap and extension penalties of 5 and 1. Nad6 could not be aligned with confidence and so was not used in the analyses. Gap placement is sometimes ambiguous near the ends of each gene alignment due to occasional variation in the lengths of some genes, poor conservation of the sequences in these regions, or uncertainty in the actual initiation and/or termination codons. To deal with this, some positions were eliminated from phylogenetic analysis according to the following criterion: In any case where gaps were introduced in the alignment of the gene ends, positions were eliminated progressively until the first occurrence of a residue conserved in at least eight of the nine taxa. This resulted in elimination of up to 3 positions at the carboxyl end of Atp8, the first 1–4 and the last 6–12 of Cox1, the last 3–8 of Cox2, the first 4–7 of Cox3, the first 7–12 of Cob, the first 14–19 and the last 16–25 of Nad1, the first 10–23 and the last 38–55 of Nad2, and the first 21–24 and the last 2–5 of Nad3. This left 1,948 aligned positions, which constituted the ‘‘whole’’ data set. Because lesser, but still significant, ambiguities remained in the alignments of Nad2 and Nad3, those genes were omitted from some analyses. We refer to the 1,579 aligned positions from this more conserved set of genes as the ‘‘limited’’ data set.

PAUP* 4.0 (Swofford 1998) was used for phylogenetic analyses. For maximum parsimony, all characters were unordered and of equal weight, and all searches employed the exhaustive search algorithm. The accelerated transformation option (ACCTRANS) was used to determine branch lengths. In separate analyses of both the whole and the limited data sets, gaps were considered ‘‘missing data’’ or ‘‘21st amino acids,’’ and a subset of taxa that omitted the chordates and arthropods was analyzed. Unrooted trees were also produced by the neighbor-joining method (Saitou and Nei 1987). Confidence estimates included consistency, retention, and rescaled consistency indices and bootstrap analysis with 1,000 replicates of a heuristic search with random order of taxon entry. Trees were rooted by designating as outgroup taxa the vertebrates *C. carpio* (Chang, Huang, and Lo 1994) and *S. acanthias* (Rasmussen and Arnason 1999), or, for the limited taxon analysis, the mollusk *K. tunicata* (Boore and Brown 1994*a*).

The sequences of tRNA genes were analyzed as a separate data set. Each of the sequences for the 12 tRNAs determined for all of *Katharina, Platynereis, Galathealinum, Helobdella,* and *Lumbricus* was aligned by eye, using potential secondary structure as a guide. The resulting 810 aligned nucleotide positions were analyzed by parsimony and neighbor-joining, as above. Maximum-likelihood analysis used quartet puzzling with empirically derived nucleotide frequencies, a 2:1 assumed ratio of the rate of transitions to transversions, and the HKY85 model (Hasegawa, Kishino, and Yano 1985). Gaps were separately treated as ‘‘missing data’’ or ‘‘fifth nucleotides.’’ The mollusk *Katharina* was used as the outgroup to root these trees.

Phylogenetic Analysis of Gene Arrangements

The mitochondrial gene arrangements of *Platynereis, Lumbricus,* and *Katharina* were also compared with those previously inferred to be primitive for Arthropoda (Staton, Daehler, and Brown 1997; Boore, Lavrov, and Brown 1998) and Chordata (Boore 1999; Boore, Daehler, and Brown 1999) (see below). We assumed that *Galathealinum, Helobdella,* and *Lumbricus* share all of the same synapomorphies, since their gene arrangements are identical for the regions determined.

The gene arrangements were analyzed using the minimum-breakpoint method (Sankoff and Blanchette 1998; Blanchette, Kunisawa, and Sankoff 1999). Briefly, this method compares each pair of arrangements and determines the number of breakpoints required to change one arrangement into the other. To simplify calculations, an early analysis had then applied distance methods to a matrix of these differences (Sankoff et al. 1992), but this approach was unsatisfactory due to problems in handling unequal rates of rearrangement and due to information lost by not identifying the specific genes involved in translocations. An improved method is now available which bases phylogeny reconstructions on parsimony and specifies genes involved in translocations (Sankoff and Blanchette 1998; Blanchette, Kunisawa, and Sankoff 1999). We employed the latter method.

Results

Gene Content and Organization

The sequenced portions of the mtDNAs of the leech *H. robusta* (GenBank accession number AF178680) and the pogonophoran *G. brachiosum* (AF178679) each contain 1 partial and 8 complete protein-encoding genes, part of *rnl,* and 12 tRNA genes (Fig. 1 and the appendices). These regions are of similar lengths in *Helobdella* and *Galathealinum* (7,553 and 7,576 nt, respectively). The gene compositions and arrangements are identical in both, and also in the ho-

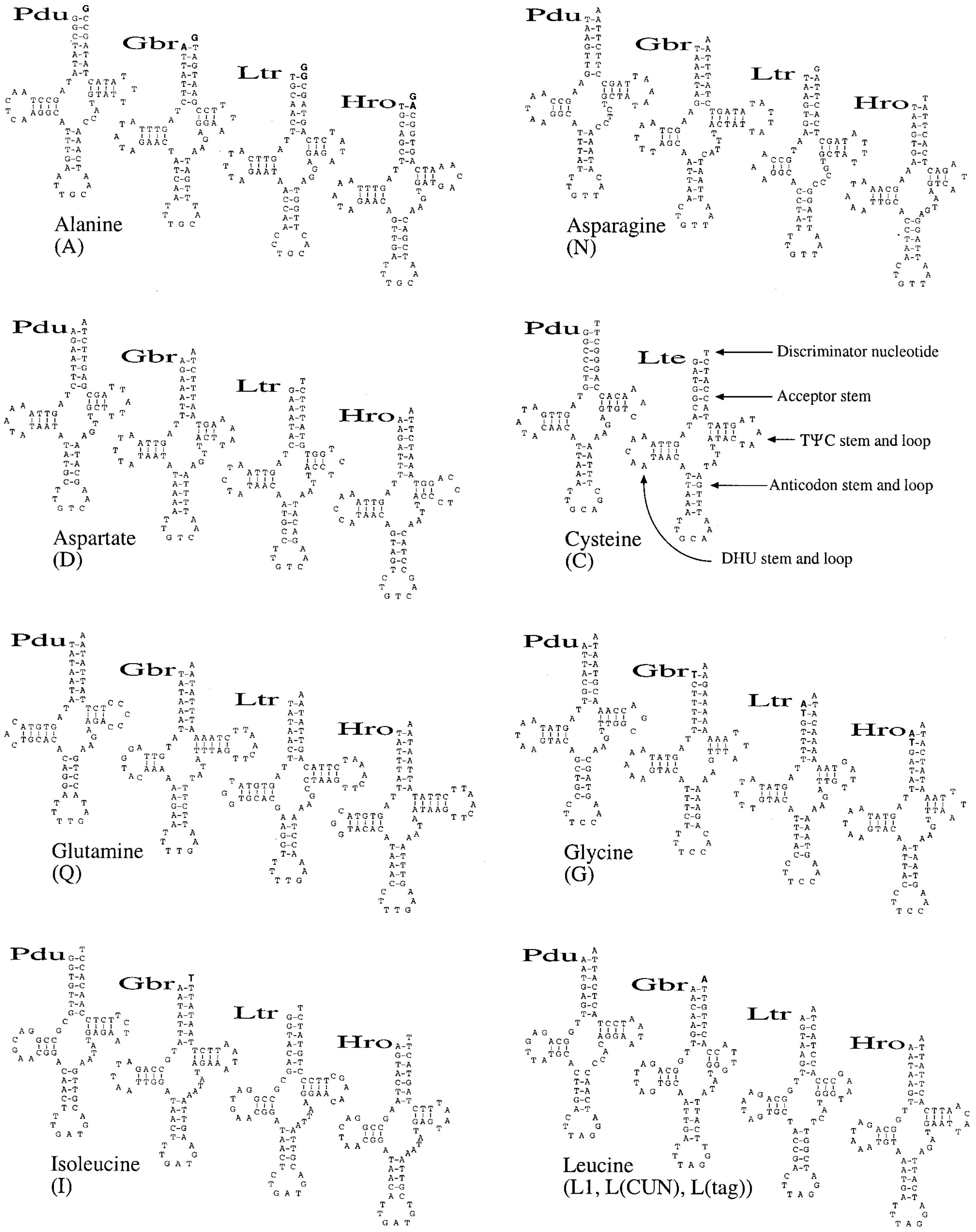


FIG. 2.—The potential secondary structures of the inferred tRNAs of *Platynereis dumerii* (Pdu), *Galathealinum brachiosum* (Gbr), *Lumbricus terrestris* (Ltr), and *Helobdella robusta* (Hro). Boldface nucleotides indicate overlap with adjacent tRNA genes. Designations for structural elements are shown for *Lumbricus* tRNA(C). For a few tRNAs, lines connect nucleotides having potential for additional base-pairing.

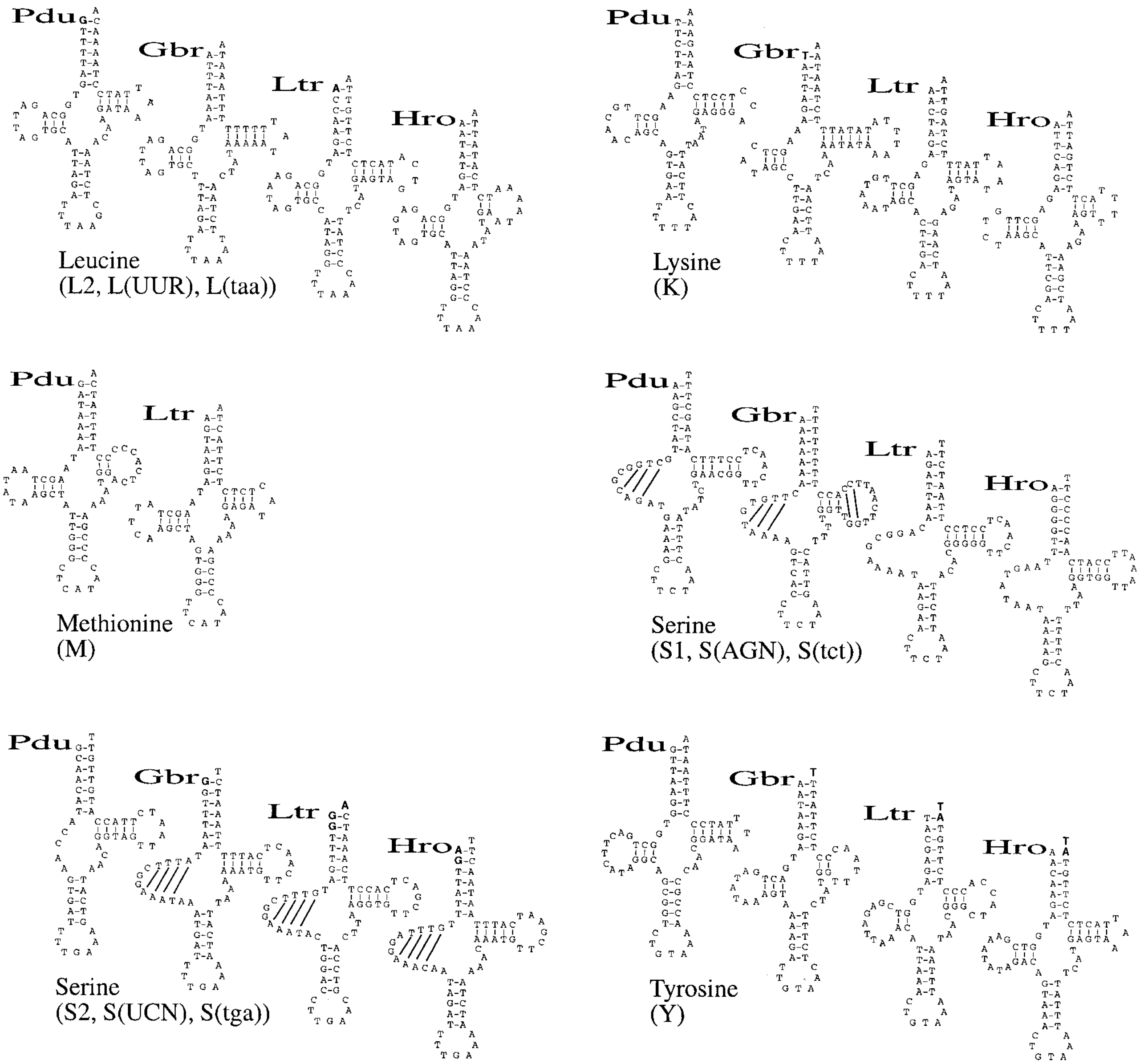


FIG. 2 (*Continued*)

mologous portion of the mtDNA of the oligochaete *L. terrestris* (Boore and Brown 1995). The gene content and arrangement of the corresponding region in the polychaete *P. dumerii* (AF178678) is similar, but there are differences in the positions of tRNA genes and these of noncoding regions. The portion of *Platynereis* mtDNA containing these same genes is significantly longer, at 8,925 nt, due mainly to the presence of two additional tRNA genes and several noncoding regions.

Initiation/Termination Codons

Alternatives to ATG start codons are very common among metazoan mtDNAs, so it is unusual to find an ATG codon at the beginning of all but one of the protein-encoding genes among these species (appendices 1–3). The *cox3* gene of *Helobdella* is the only exception and is most probably initiated by TTG. Although the ATA preceding it could be the start codon, this would cause a 3-nt overlap with the preceding *trnG.*

Many of the protein gene sequences in this study appear to end with a single T that is directly adjacent to the downstream gene. It is common for termination codons to be truncated (to T- or TA-) in metazoan mtDNAs; such codons are converted to complete (UAA) stop codons by polyadenylation after transcript processing (Ojala, Montoya, and Attardi 1981). Several of the genes we sequenced have complete, in-frame stop codons that would require short overlaps with the downstream genes, sometimes of only 1 or 2 nt. We speculate that these are normally unused; that normal cleavage of the polycistronic transcript yields incomplete stop codons, subsequently completed by polyadenylation; and that the encoded TAA codons function (if at all) only as ‘‘backups’’ to prevent translational readthrough if the transcripts are not properly cleaved.

Transfer RNAs

There are 12 tRNA genes in the sequenced portions of *Galathealinum* and *Helobdella* mtDNAs, each identical in relative position and polarity to its homolog in *Lumbricus.* The same set of 12 tRNA genes is also found in the sequenced portion of *Platynereis* mtDNA, along with two additional tRNA genes, *trnC* and *trnM.* Figure 1 shows the relative location of each of these genes, and figure 2 shows their potential secondary structures and compares them with their homologs in *Lumbricus.*

All tRNA gene sequences have the potential to form a 7-nt-pair acceptor stem and a 5-nt-pair anticodon stem, despite a single mismatch in a few and two mismatches in one (the acceptor stem of tRNA(N) in *Platynereis*). Except for the serine tRNAs, all can form standard cloverleaf structures, with the DHU and TΨC stems nearly always 3–5 bp, with loops of 3–7 nt. With only a few exceptions, the nucleotides preceding and following anticodons are T and A, respectively, and the most common dinucleotide separating the acceptor stem from the DHU stem is TA. A single A separates the DHU stem from the anticodon stem in all but 6 of the 52 tRNA genes shown in figure 2. Forty-one of these 52 tRNAs have 4 nt in the ‘‘extra’’ arm, 10 have 5 nt (tRNA(N) and tRNA(I) in all four species, and tRNA(K) and tRNA(S1) in *Platynereis*), and 1 has 3 nt (tRNA(G) in *Galathealinum*). (This tRNA could be folded alternatively with a four-member extra arm if the TΨC stem were only 2 bp and the acceptor stem were 1 nt shorter at the 3′ end with one mismatch; however, this seems less likely.)

There are 10 cases in which the sequences of adjacent tRNA genes overlap (marked with boldface nucleotides in fig. 2 and shown in the appendices). No other gene overlaps occur, assuming that all stop codons have been correctly inferred (see below). For the structures in figure 2 to form, complete individual tRNA gene transcripts are required. Given the overlaps, however, these are not possible unless (1) transcription of each originates from a different promoter; (2) polycistronic transcript processing alternates, yielding sometimes one or the other complete tRNA; or (3) transcript editing restores the nucleotides that are lost in processing. For 6 of these 10 overlapping pairs, the overlap involves only the discriminator nucleotide. It is possible that this nucleotide is not encoded by these genes but is added posttranscriptionally either by polyadenylation (demonstrated for some mt tRNAs; Yokobori and Pääbo 1997) or by a mechanism similar to the one that adds CCA to the 3′ ends of tRNAs. There are four cases of 2-nt overlap, involving the adjacent gene pairs *trnA*-*trnS2*(*tga*) and *trnY*-*trnG* in both *Lumbricus* and *Helobdella*; since these two are the most closely related pair of taxa (see below), it is possible that they share some mechanism for resolving these larger overlaps of these identical gene pairs.

As is found in many other mitochondrial systems, the DHU arms of the serine tRNAs cannot be folded into standard stem-loop structures (fig. 2). In some cases, alternative folding yields a structure with a DHU stem having only 1 nt between the acceptor and DHU stems and 2 nt between the DHU and anticodon stems. (Similar alternative folding is possible for the serine tRNAs of *Katharina*; see Boore and Brown 1994*a.*)

Nucleotide and Amino Acid Composition

These mtDNA sequences like those of most metazoans, are AT-rich, ranging from 61% to 76% (table 1). *Platynereis* and *Lumbricus* have similar biases in codon usage (table 2), and their values for these are not greatly different from those found in the mollusk *Katharina* (Boore and Brown 1994*a*). Codon usage in *Helobdella* is somewhat different, generally reflecting its greater A+T richness, and is most extreme for the codons ACG and CGG, neither of which is used within the portion of the mtDNA sequenced.

Codon usage is most biased in *Galathealinum* mtDNA, in which 5 codons are never used and 13 others are used three or fewer times. The occurrence of C at third codon positions is much less frequent than it is in the other three species (consistent with *Galathealinum*’s higher A+T richness), but the most extreme bias is against G, which occurs at the third position in only 26 of the 2,236 codons in the *Galathealinum* sequence. Furthermore, 9 of these are start (ATG) codons that are evidently maintained by strict selection, judging by their near universality of use within this group, leaving only 17 codons ending in G that are truly synonymous with another codon.

Although selection for A+T richness could account, in part, for the infrequency of G and C at third codon positions, it cannot account specifically for the bias against G relative to C. It is possible that an anti-G bias results from selection for translational efficiency or, alternatively, that it results from a bias in mutational tendency.

The higher A+T bias in *Galathealinum* mtDNA may account for the differences in the amino acid compositions of some of its proteins from those of the other species (table 3). Alanine, valine, and threonine, each encoded by a G- or C-containing codon, are underrepresented in *Galathealinum,* and isoleucine and phenylalanine, each encoded by AT-rich codons, are overrepresented. Interestingly, the ratios of nonpolar to polar amino acids are very similar among these four mtDNAs, indicating the importance of physicochemical characteristics on substitution patterns. Also reflecting *Galathealinum*’s higher A+T bias is its greater usage of TTR than CTN to encode leucine (TTR : CTN ratio = 1.67); CTN is more commonly used in the other three species (TTR : CTN ratios of 0.75, 0.49, and 0.79). Finally, further emphasizing *Galathealinum*’s anti-G bias, all but three of the 206 TTR codons are TTA.

The bias against G in *Galathealinum* mtDNA, so evident at third codon positions, may also account for the low frequency of occurrence of certain codons. All of the 18 codons occurring three or fewer times contain

Table 1
Frequency of Occurrence of Each Nucleotide Within the Partial Mitochondrial Genome of *Platynereis dumerii* (Pdu), *Galathealinum brachiosum* (Gbr), *Helobdella robusta* (Hro), and *Lumbricus terrestris* (Ltr; Boore and Brown 1995)

	tRNA Genes				Protein-Encoding Genes				Third Codon Positions				NC Region[a]	
	Pdu 763[b]	Gbr 766	Ltr[c] 765	Hro 761	Pdu 6,663	Gbr 6,708	Ltr[c] 6,714	Hro 6,699	Pdu 2,221	Gbr 2,236	Ltr[c] 2,238	Hro 2,233	Pdu 1,091	Ltr 384
C	0.187	0.095	0.182	0.130	0.214	0.144	0.237	0.196	0.211	0.064	0.242	0.170	0.156	0.151
G	0.177	0.111	0.187	0.151	0.160	0.099	0.160	0.131	0.093	0.012[d]	0.101	0.054	0.126	0.206
A	0.325	0.385	0.319	0.378	0.296	0.305	0.276	0.325	0.391	0.407	0.355	0.457	0.335	0.326
T	0.311	0.409	0.312	0.340	0.330	0.452	0.326	0.348	0.305	0.518	0.301	0.319	0.384	0.318
A+T	0.636	0.794	0.631	0.718	0.626	0.757	0.602	0.673	0.696	0.925	0.656	0.776	0.719	0.644
GC-skew[e]	−0.027	0.078	0.014	−0.075	−0.144	−0.185	−0.194	−0.199	−0.390	−0.690	−0.409	−0.517	−0.106	0.154
TA-skew[f]	−0.022	0.030	−0.011	−0.053	0.054	0.194	0.083	0.034	−0.123	0.120	−0.082	−0.175	0.068	−0.012

[a] Scores for the nucleotides in the single largest noncoding region.

[b] Total number of nucleotides scored is given in each case beside the taxon designation.

[c] Scores for only the homologs (or homologous portion of *cob*) to those reported for the other animals.

[d] In total, of the 2,236 codons considered for *Galathealinum*, only 26 end in G. Nine of these 26 are ATG start codons. If these are assumed to be maintained by strict selection and omitted from the analysis, GC-skew for *Galathealinum* is −0.786.

[e] GC-skew is a measure of the bias for one strand in having G rather than C. This value ranges from 1 (the G of all GC pairs is on the considered strand) to −1 (the C of all GC pairs is on the considered strand). GC-skew is zero if there is no bias between the strands for G rather than C. GC-skew is computed as (G − C)/(G + C).

[f] TA-skew is analogous to GC-skew, but considering the bias of T over A for one strand. TA-skew is computed as (T − A)/(T + A).

at least one G (mean 1.5). Despite this, six of the eight amino acids that can be specified only by G-containing codons are not underrepresented in *Galathealinum* mtDNA (table 4). It is possible that these persist against a strong anti-G bias because they constitute the minimal set required for function. This possibility is supported by the higher than average degree of conservation of these amino acids at these positions (table 4). Of the 1,579 aligned amino acid positions in the limited data set (see *Materials and Methods*), 843 are completely conserved in the three annelids and the pogonophoran, and 650 are completely conserved in all nine taxa. The degree of conservation of five of the six amino acids specified by G-containing codons (all except cysteine) is significantly higher than the average in both cases, supporting the view that these positions are under strong functional constraints. Why cysteines are less well conserved than others is not obvious and may be an artifact of their low frequency of occurrence.

In addition to showing nucleotide composition and A+T richness, table 1 presents values for the skewness of GC and AT pairs (Perna and Kocher 1995), which reflect the amount of interstrand bias in A versus T and in G versus C. There is no appreciable skew in the tRNA genes, presumably due to the structural requirement for intrastrand base-pairing. There is a small amount of GC-skew (ca. −0.2) in the protein genes and significantly more at third codon positions, both reflecting the anti-G bias already noted. TA-skew is highest among the four taxa in *Galathealinum* mtDNA for all categories in table 1 and is especially high in the total protein-encoding sequences.

Unassigned Nucleotides

Metazoan mitochondrial genomes typically contain at least one relatively large region devoid of structural genes. In vertebrates, this region contains elements that control replication and transcription (see Shadel and Clayton 1997), and analogous regions may function similarly in other taxa. No such region is found in the sequenced portions of *Galathealinum* or *Helobdella* mtDNAs, but there is a 1,091-nt noncoding region in *Platynereis* mtDNA. This region is somewhat higher in A+T (72%) than an analogous 384-nt region in *Lumbricus* (64%) and is located between *trnG* and *trnY* in *Platynereis* and between *trnR* and *trnH* in *Lumbricus.* The primitive position for this region in Annelida is uncertain, but the presence in *Platynereis* of apparently translocated tRNA genes (see below) flanking it, along with the frequent movement of such regions in conjunction with other translocations (Boore 1999), suggests that its position in *Platynereis* is derived.

Proportional to their individual frequencies, all four homodinucleotides and all four homotrinucleotides, except for CCC in *Lumbricus,* are overrepresented for the noncoding regions of both *Platynereis* and *Lumbricus.* The dinucleotide CG, normally one of the least common in metazoan mtDNAs (Cardon et al. 1994), is also present at higher frequencies than expected in both *Platynereis* and *Lumbricus* noncoding regions. The dinucleotide TA occurs 151 times in the *Platynereis* and 34

Table 2
Codon Usage of Nine[a] Mitochondrially Encoded Genes of Four Annelids, Including the Pogonophoran

		PDU[b]		GBR[c]		LTR[d]		HRO[e]				PDU		GBR		LTR		HRO	
AMINO ACID	CODON	*N*	%	*N*	%	*N*	%	*N*	%	AMINO ACID	CODON	*N*	%	*N*	%	*N*	%	*N*	%
Phe (F) (GAA)[f] . . .	TTT	107	4.8	219	9.8	102	4.6	118	5.3	Tyr (Y) (GUA) . . .	TAT	40	1.8	78	3.5	39	1.7	54	2.4
	TTC	60	2.7	33	1.5	67	3.0	50	2.2		TAC	29	1.3	8	0.4	35	1.6	40	1.8
Leu (L2) (UAA) . . .	TTA	114	5.1	203	9.1	90	4.0	128	5.7	TER[g]	TAA	1	—	2	—	0	—	1	—
	TTG	17	0.8	3	0.1	15	0.7	12	0.5		TAG	0	—	0	—	0	—	0	—
Leu (L1) (UAG) . . .	CTT	64	2.9	77	3.4	56	2.5	48	2.1	His (H) (GUG) . . .	CAT	26	1.2	52	2.3	25	1.1	33	1.5
	CTC	16	0.7	6	0.3	40	1.8	22	1.0		CAC	32	1.4	5	0.2	37	1.7	24	1.1
	CTA	80	3.6	39	1.7	104	4.6	99	4.4	Gln (Q) (UUG) . . .	CAA	35	1.6	37	1.7	38	1.7	42	1.9
	CTG	13	0.6	1	0	14	0.6	7	0.3		CAG	9	0.4	0	0	3	0.1	1	0
	ATT	139	6.2	227	10.2	126	5.6	149	6.7	Asn (N) (GUU) . . .	AAT	44	2.0	89	4.0	38	1.7	66	3.0
Ile (I) (GAU)	ATC	47	2.1	33	1.5	58	2.6	64	2.9		AAC	44	2.0	12	0.5	42	1.9	33	1.5
	ATA	126	5.7	137	6.1	101	4.5	129	5.8	Lys (K) (UUU) . . .	AAA	40	1.8	56	2.5	38	1.7	48	2.1
Met (M) (CAU) . . .	ATG	29	1.3	11	0.5	32	1.4	31	1.4		AAG	4	0.2	3	0.1	10	0.4	8	0.4
Val (V) (UAC)	GTT	43	1.9	27	1.2	29	1.3	29	1.3	Asp (D) (GUC) . . .	GAT	14	0.6	33	1.5	28	1.3	23	1.0
	GTC	13	0.6	3	0.1	21	0.9	11	0.5		GAC	28	1.3	5	0.2	21	0.9	28	1.3
	GTA	72	3.2	39	1.7	59	2.6	81	3.6	Glu (E) (UUC) . . .	GAA	44	2.0	48	2.1	25	1.1	39	1.7
	GTG	13	0.6	1	0	25	1.1	6	0.3		GAG	12	0.5	0	0	23	1.0	10	0.4
Ser (S2) (UGA) . . .	TCT	34	1.5	115	5.1	41	1.8	38	1.7	Cys (C) (GCA) . . .	TGT	16	0.7	18	0.8	7	0.3	14	0.6
	TCC	29	1.3	8	0.4	44	2.0	19	0.9		TGC	6	0.3	4	0.2	12	0.5	6	0.3
	TCA	43	1.9	57	2.5	55	2.5	83	3.7	Trp (W) (UCA) . . .	TGA	52	2.3	63	2.8	54	2.4	59	2.6
	TCG	8	0.4	1	0	4	0.2	5	0.2		TGG	16	0.7	0	0	13	0.6	6	0.3
Pro (P) (UGG)	CCT	28	1.3	60	2.7	37	1.7	20	0.9	Arg (R) (UCG) . . .	CGT	8	0.4	21	0.9	9	0.4	6	0.3
	CCC	28	1.3	5	0.2	34	1.5	4	0.2		CGC	10	0.4	3	0.1	4	0.2	5	0.2
	CCA	46	2.1	37	1.7	32	1.4	75	3.4		CGA	17	0.8	14	0.6	24	1.1	29	1.3
	CCG	6	0.3	1	0	11	0.5	2	0.1		CGG	6	0.3	1	0	7	0.3	0	0
Thr (T) (UGU)	ACT	39	1.8	55	2.5	46	2.1	37	1.7	Ser (S1) (UCU) . . .	AGT	11	0.5	6	0.3	9	0.4	11	0.5
	ACC	42	1.9	6	0.3	43	1.9	25	1.1		AGC	17	0.8	1	0	6	0.3	5	0.2
	ACA	63	2.8	34	1.5	51	2.3	77	3.4		AGA	32	1.4	39	1.7	37	1.7	25	1.1
	ACG	17	0.8	0	0	8	0.4	0	0		AGG	8	0.4	3	0.1	10	0.4	5	0.2
Ala (A) (UGC)	GCT	49	2.2	43	1.9	59	2.6	37	1.7	Gly (G) (UCC) . . .	GGT	16	0.7	39	1.7	23	1.0	30	1.3
	GCC	31	1.4	7	0.3	51	2.3	31	1.4		GGC	37	1.7	3	0.1	26	1.2	12	0.5
	GCA	52	2.3	36	1.6	48	2.1	58	2.6		GGA	51	2.3	69	3.1	39	1.7	47	2.1
	GCG	18	0.8	0	0	12	0.5	5	0.2		GGG	30	1.3	1	0	40	1.8	23	1.0

[a] The inferred protein sequences analyzed are those of *nad1–3*, *nad6*, *cox1–3*, *atp8*, and the amino-terminal 269 residues of *cob*.

[b] The 2,221 codons of *Platynereis dumerii*.

[c] The 2,236 codons of *Galathealinum brachiosum*.

[d] The 2,238 codons of the homologous regions of *Lumbricus terrestris* (Boore and Brown 1995).

[e] The 2,233 codons of *Helobdella robusta*.

[f] The anticodon of the corresponding tRNA is shown in parentheses for each codon designation. The 10 that are unknown for *Galathealinum* and *Helobdella* are underlined.

[g] Termination codons were not included in this analysis; this is the count of predicted complete stop codons that do not overlap the adjacent genes.

Table 3
Amino Acid Compositions of Nine Protein-Encoding Genes

	PLATYNEREIS		GALATHEOLINUM		LUMBRICUS		HELOBDELLA	
	No.	%	No.	%	No.	%	No.	%
Nonpolar								
A (GCN)	150	6.74	**86**	3.85	170	7.59	131	5.86
V (GTN)	141	6.34	**70**	3.13	134	5.98	128	5.73
L (total)	304	13.69	329	14.71	319	14.25	316	14.15
(TTR)	(131)	(5.90)	(206)	(9.21)	(105)	(4.69)	(140)	(6.27)
(CTN)	(173)	(7.79)	(123)	(5.50)	(214)	(9.56)	(176)	(7.88)
I (ATY)	186	8.36	**261**	11.67	184	8.22	213	9.53
P (CCN)	108	4.86	103	4.61	114	5.09	101	4.52
M (ATR)	155	6.97	148	6.62	133	5.94	160	7.16
F (TTY)	167	7.51	**252**	11.27	169	7.55	168	7.52
W (TGR)	68	3.06	63	2.82	67	2.99	65	2.91
Total	1,279	57.59	1,312	58.68	1,290	57.64	1,282	57.41
Polar								
G (GGN)	134	6.03	112	5.01	128	5.72	112	5.01
S (total)	182	8.19	230	10.29	206	9.20	191	8.55
(TCN)	(114)	(5.13)	(181)	(8.09)	(144)	(6.43)	(145)	(6.49)
(AGN)	(68)	(3.06)	(49)	(2.19)	(62)	(2.77)	(46)	(2.06)
T (ACN)	161	7.24	**95**	4.25	148	6.61	139	6.22
C (TGY)	22	0.99	22	0.98	19	0.85	20	0.90
Y (TAY)	69	3.10	86	3.85	74	3.31	94	4.21
N (AAY)	88	3.96	101	4.52	80	3.57	99	4.43
Q (CAR)	44	1.98	37	1.65	41	1.83	43	1.92
Total	700	31.52	683	30.54	696	31.10	698	31.26
Acidic								
D (GAY)	42	1.89	38	1.70	49	2.19	51	2.28
E (GAR)	56	2.52	48	2.15	48	2.14	49	2.19
Total	98	4.41	86	3.85	97	4.33	100	4.48
Basic								
K (AAR)	44	1.98	59	2.64	49	2.19	56	2.51
R (GCN)	41	1.84	39	1.74	44	1.97	40	1.79
H (CAY)	59	2.65	57	2.55	62	2.77	57	2.55
Total	144	6.48	155	6.93	155	6.93	153	6.85

NOTE.—Included are the complete inferred amino acid sequences of *nad1–3, nad6, cox1–3, atp8,* and the 269 amino-terminal residues of *cob.* The codon for each amino acid is given in parentheses. Shown in bold are the numbers for those amino acids whose compositions in *Galathealinum* mtDNA differ by >30% from the mean of their occurrence in the other three mtDNAs. For the amino acids leucine and serine, separate subtotals reflecting the use of each of the two codon families are given in parentheses after the summary total.

times in the *Lumbricus* noncoding region. This is not greatly different from expectation, given the high A+T composition, but is noteworthy because many of these dinucleotides are found in runs of alternating TA pairs. In the noncoding regions of *Platynereis* and *Lumbricus* mtDNAs, respectively, 61 and 14 of the TA pairs occur adjacent to at least one other TA pair, with the longest runs being 14 and 6 consecutive TA pairs for each of the two mtDNAs. This has also been observed in other mtDNA noncoding regions (e.g., *Katharina* mtDNA contains a run of 36 TA pairs; Boore and Brown 1994*a*). Finally, no blocks of significant sequence similarity were identified between the noncoding regions of *Lumbricus* and *Platynereis*, although short T, A, and G homopolymer runs occur in the noncoding regions in both species.

There are totals of only six noncoding nucleotides in the sequenced portion of *Galathealinum* mtDNA and only two for *Helobdella.* In *Platynereis* mtDNA, in addition to the large noncoding region described above, there are numerous intergenic nucleotides: 5 (CACAT) between *trnL1*(*tag*) and *trnS2*(*tga*), 1 (T) between *trnC* and *cox1*, 61 between *cox2* and *trnG*, 60 between *trnY* and *atp8*, 3 (AAT) between *trnM* and *trnD*, 32 between *trnD* and *cox3*, and 9 (GGATATCCT) between *trnQ* and *nad6* (appendix 1). All except the 9 nt separating *trnQ* and *nad6* are adjacent to tRNAs whose positions appear to be derived. The presence of a block of intergenic nucleotides is a condition often associated with a recent

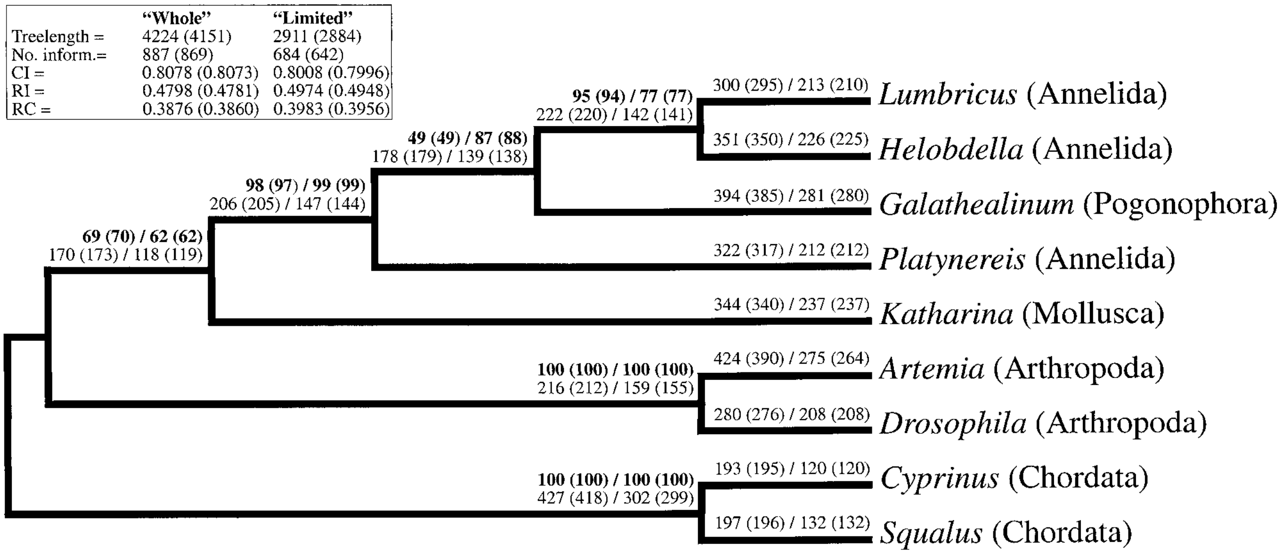
Table 4
Those Amino Acids that Are Specified by G-Containing Codons but Are Not Underrepresented in *Galathealinum* Mitochondrial Proteins Correspond to Highly Conserved Residues

Amino Acid	Codon	No. of Codons[a]	No. Unvarying in All Taxa[b] (%)		No. Unvarying in Annelida[c] (%)	
W	TGR	52	47	90.4	50	96.2
G	GGN	93	71	76.3	77	82.8
C	TGY	17	3	17.6	4	23.5
D	GAY	36	25	69.4	29	80.6
E	GAR	40	24	60.0	32	80.0
R	CGN	34	24	70.6	30	88.2
All . . .	—	1,579	650	41.2	843	53.4

[a] The number of occurrences of the codon the in aligned 1,579-amino-acid alignment referred to as the limited data set.

[b] *Squalus, Cyprinus, Drosophila, Artemia, Katharina, Platynereis, Galathealinum, Helobdella,* and *Lumbricus.*

[c] *Galathealinum, Platynereis, Helobdella,* and *Lumbricus.*



	"Whole"	"Limited"
Treelength =	4224 (4151)	2911 (2884)
No. inform.=	887 (869)	684 (642)
CI =	0.8078 (0.8073)	0.8008 (0.7996)
RI =	0.4798 (0.4781)	0.4974 (0.4948)
RC =	0.3876 (0.3860)	0.3983 (0.3956)

FIG. 3.—Single most-parsimonious tree rooted by assuming the vertebrates as an outgroup. Bootstrap values (1,000 replicates of heuristic search with random order of taxon entry) are shown in boldface, and branch lengths (accelerated transformation option) are shown in lightface. Values shown before the slash are for the whole data set (1,948 aligned amino acid positions; see *Materials and Methods*); those after the slash are for the limited data set (1,579 positions), which omits Nad2 and Nad3, the two least conserved genes. In each case, gaps were first scored as additional characters, followed in parentheses by the values obtained with gaps scored as missing data. Tree length, number of parsimony-informative characters, consistency index (CI), retention index (RI), and rescaled consistency index (RC) are shown. The neighbor-joining tree has an identical topology.

gene translocation (Boore 1999). The tRNA(Q) of *Platynereis* is quite dissimilar to those of the other animals and is shorter; it is tempting to speculate that these nine adjacent noncoding nucleotides are the vestige of a recent shift in the sequences coding for this tRNA. None of these noncoding sequences of *Platynereis* mtDNA are similar to any portion of *Lumbricus* mtDNA.

Phylogenetic Reconstruction

The whole data set of 1,948 amino acid positions for nine taxa was analyzed by parsimony, using PAUP* 4.0 (Swofford 1998). The limited data set, which omits the two most variable inferred proteins (Nad2 and Nad3), was subjected to the same analysis. Gaps introduced to maximize alignment similarity were alternatively scored as additional characters or as missing data. Trees were rooted using the two vertebrate species as an outgroup. All four of these analyses (using whole or limited data sets with two modes of gap scoring) yielded the same most-parsimonious tree, with the next-shortest tree being at least three steps longer (fig. 3). An alternative method, neighbor-joining analysis (Saitou and Nei 1987), yielded trees with the same topology.

Considering the possibility that using a distant outgroup could bias phylogenetic reconstruction, analyses were also performed using only the five most closely related taxa in figure 3 (*Katharina, Platynereis*, *Galathealinum, Helobdella,* and *Lumbricus*). This subset of taxa was analyzed by parsimony, neighbor joining, and maximum likelihood, using the aligned amino acid sequences for the protein genes or the 810 aligned nucleotides for the shared tRNA genes. The mollusk *Katharina* was used as the outgroup to root the trees. All analyses yielded the same single most-parsimonious tree (fig. 4). In parsimony analyses of the protein-encoding genes, regardless of gap scoring, the next-most-parsimonious trees are only one or two steps longer than the shortest tree. However, analysis of the tRNA gene sequences with either method of gap scoring yielded no alternative trees less than 10 steps longer than the minimum-length tree, and maximum-likelihood analysis provided 100% puzzling support for each of its nodes.

A potentially confounding factor in any sequence-based phylogeny is the artifactual attraction of branches with higher rates of change (Felsenstein 1978). This does not appear to have been a significant bias for the trees reported here. As can be seen in figures 3 and 4, branch lengths assigned by parsimony analyses vary only within a fairly small range, and the controversial relationships advanced (i.e., Annelida + Mollusca excluding Arthropoda, and Pognophora + Clitellata excluding the polychaete) do not unite branches that are longer than their neighbors. Pairwise distance measures vary over only a small range between each of the four annelid/pogonophoran ingroup taxa and their outgroup *Katharina* (0.374–0.419), between each of these five taxa and their arthropod outgroup (0.370–0.446), and between these seven taxa and their chordate outgroup (0.387–0.449).

As an independent source of phylogenetic information, mitochondrial gene arrangements were also compared. The mitochondrial gene arrangements analyzed include those of the mollusk *K. tunicata* (Boore and Brown 1994*a*), the oligochaete *L. terrestris* (Boore and Brown 1995), the polychaete *P. dumerii* (with missing genes in positions to match their arrangement in *Lumbricus*), and those inferred to be primitive for Arthropoda and Chordata (see fig. 1*B*).

	Protein genes	tRNA genes
Treelength =	1597 (1587)	984 (864)
No. inform.=	226 (226)	251 (229)
CI =	0.9130 (0.9124)	0.8537 (0.8484)
RI =	0.3850 (0.3850)	0.4263 (0.4279)
RC =	0.3514 (0.3512)	0.3639 (0.3631)

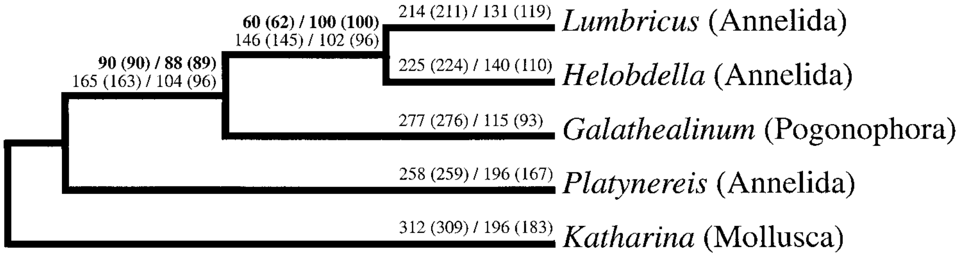


FIG. 4.—Single most-parsimonious tree considering only a limited set of the studied taxa and rooted by assuming the mollusk *Katharina* as an outgroup. Bootstrap values (1,000 replicates of heuristic search with random order of taxon entry) are shown in boldface, and branch lengths (accelerated transformation option) are shown in lightface. Values shown before the slash are for the limited data set (1,579 aligned amino acid positions; see *Materials and Methods*), and those after the slash are from the alignment of 12 tRNA gene sequences (810 aligned nucleotide positions). In each case, gaps were first scored as additional characters, and the values obtained are followed in parentheses by the values obtained with gaps scored as missing data. Tree length, number of parsimony-informative characters, consistency index (CI), retention index (RI), and rescaled consistency index (RC) are shown. The neighbor-joining trees and, for the tRNA gene data set, the maximum-likelihood tree have identical topologies.

In order to infer the primitive arrangement for Arthropoda, we consider the two most distantly related arthropod groups so far studied, Cheliceriformes, represented by the horseshoe crab *Limulus polyphemus* (Staton, Daehler, and Brown 1997) and Insecta, represented by *Drosophila* (Clary and Wolstenholme 1985). The mitochondrial gene arrangements of these two animals differ by only a single tRNA gene position, that of *trnL2*(*taa*). It is parsimonious to infer that their common ancestor had one or the other of these two arrangements, which can easily be determined by considering those of less related taxa. The position of *trnL2*(*taa*) in *Limulus* mtDNA (i.e., *rnl-trnL1-trnL2-nad1*) is shared by many outgroup taxa (Boore et al. 1995; Boore, Lavrov, and Brown 1998), including the mollusk *Katharina,* as shown in figure 1*B,* and so it is the gene arrangement found in *Limulus* that must be primitive for Arthropoda or, more exactly, for whatever arthropods diverged after the split of Cheliceriformes and Insecta.

The same type of logic allows inference of the primitive arrangement for Chordata. The early-branching cephalochordate *Branchiostoma* (Spruyt et al. 1998; Boore, Daehler, and Brown 1999) has a gene arrangement differing from that of a fish (Chang, Huang, and Lo 1994) (which is, itself, identical to the arrangements found in dozens of other diverse vertebrates; see Boore 1999) by only four tRNA positions. For one of these four differences, the *Branchiostoma* arrangement (*trnN-trnW-trnA-trnC-trnY*) is also found in the even more distantly related hemichordate *Balanoglossus* (Castresana et al. 1998*b*), so this must be the primitive arrangement for the common ancestor of the Chordata. For the other three cases (positions of *trnF, trnM,* and *trnG*), similar or identical positions are shared between the mtDNAs of the fish and outgroups such as arthropods and/or echinoderms, so it is most parsimoniously inferred that the fish arrangement is primitive, with separate, later translocations in the lineage leading to *Branchiostoma* (see further discussion in Boore, Daehler, and Brown 1999).

We applied the minimum-breakpoint method for reconstructing patterns of gene rearrangement (Sankoff and Blanchette 1998; Blanchette, Kunisawa, and Sankoff 1999). Given a tree topology with gene arrangements associated with each branch, this method reconstructs ancestral gene arrangements for each node in such a way that the total number of breakpoints between neighboring nodes is minimized. A particular ancestral reconstruction at a node may not be unique and can range from being identical to one descendent arrangement (with the branch to the other descendent having all differences) to the opposite condition. To overcome this, the search is run iteratively with random trials of ancestral reconstructions, each searching for the minimum number of breakpoints. All possible unrooted trees are scored this way for the total number of breakpoints; the shortest corresponds to a parsimony analysis of gene arrangements and forms a phylogenetic hypothesis.

This analysis yields a single shortest tree requiring a total of 76 breakpoints (fig. 5); the next-shortest tree requires a total of 80. The tree produced was rooted by designating Chordata as the outgroup.

Discussion

Mitochondrial genome comparisons serve as models of genome evolution. In this system, much smaller and simpler than that of the nucleus, are all of the same factors of genome evolution, where one may find tractable the changes in tRNA structure, base composition, genetic code, gene arrangement, etc. Several observations are noteworthy in this study: (1) There are strong biases in codon usage for *Helobdella* and, especially, for *Galathealinum,* where it seems to influence the amino acid compositions of the encoded proteins. Complete loss of use of a particular codon is the precondition for genetic code change in the most commonly invoked model (Osawa and Jukes 1989; Castresana et al. 1998*a*), and *Galathealinum* mtDNA is the most extreme exam-

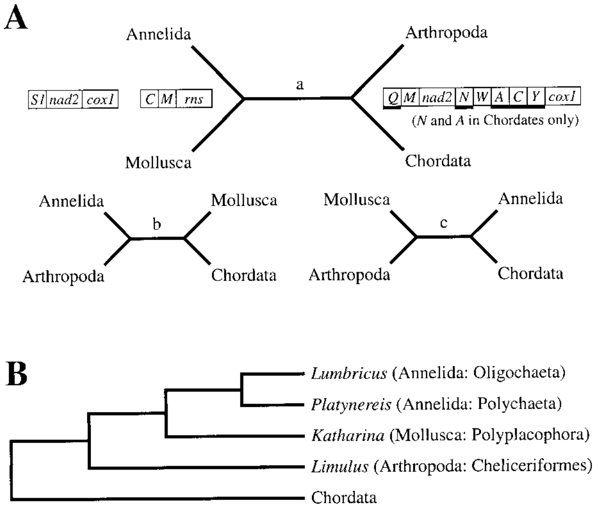

A
Annelida
Arthropoda
S1 nad2 cox1
C M rns
a
Q M nad2 N W A C Y cox1
(N and A in Chordates only)
Mollusca
Chordata
Annelida
b
Mollusca
Arthropoda
Chordata
Mollusca
c
Annelida
Arthropoda
Chordata
B
Lumbricus (Annelida: Oligochaeta)
Platynereis (Annelida: Polychaeta)
Katharina (Mollusca: Polyplacophora)
Limulus (Arthropoda: Cheliceriformes)
Chordata

FIG. 5.—*A,* The three possible unrooted trees for the four phyla Annelida, Mollusca, Arthropoda, and Chordata are shown (a, b, and c). Annelida is represented by the gene arrangement of *Lumbricus* (Boore and Brown 1995), Mollusca by that of *Katharina* (Boore and Brown 1994*a,* 1994*b*), Arthropoda by that of *Limulus* (Staton, Daehler, and Brown 1997), and Chordata by that inferred to be primitive for this phylum (Boore, Daehler, and Brown 1999), which differs from fish mtDNA (Chang, Huang, and Lo 1994; Rasmussen and Arnason 1999) in the location of only one tRNA gene (see above). Genes are abbreviated as in figure 1, with underlining signifying genes oriented right to left as drawn. Only branching pattern a has support, with several gene boundaries shared in a phylogenetically informative pattern, as shown by their schematic representation on each side of this unrooted tree. Note the positions of *nad2, trnC,* and *trnM,* all of which would have had to experience convergent translocations if branching patterns b or c were correct. *B,* The results of comparing mitochondrial gene arrangements using the minimum-breakpoint method. Taxa analyzed were *Lumbricus terrestris, Platynereis dumerii, Katharina tunicata,* and the gene arrangements inferred to be primitive for each of Arthropoda and Chordata (see text). This is the single shortest tree from the analysis (76 total breakpoints) rooted by designating Chordata as the outgroup.

ple of this condition so far observed. It is possible that any of the unique features of *Galathealinum* mtDNA may be related to its living in an extreme environment, in hydrothermal vent communities at abyssal depths. (2) ATG start codons are found for all but one of the protein-encoding genes in all of the annelid (including pogonophoran) mtDNAs, whereas it is very common for metazoan mtDNAs to use a variety of alternate start codons. (3) Although it is not demonstrated that the large noncoding regions of *Platynereis* and *Lumbricus* mtDNAs contain regulatory signaling elements, this is the common function for analogous regions of other mtDNAs (Shadel and Clayton 1997). There is little similarity in these sequences between the two mtDNAs, suggesting that the signals necessary for regulating transcription and/or replication, if present, are short, difficult to recognize, and/or rapidly evolving.

All genes are identically arranged in the studied portions of the mtDNAs of the oligochaete *Lumbricus,* the hirudinid *Helobdella,* and the pogonophoran *Galathealinum.* Several tRNA genes differ in their locations in the mtDNA of the polychaete *Platynereis.* It is unclear which of these two gene arrangements is primitive for the Annelida. For three of the differently located tRNA genes, *trnG, trnY,* and *trnD,* there is a noncoding region flanking their positions in *Platynereis* mtDNA, as is often found for recently translocated genes (Boore 1999). One portion of the gene arrangement for the clitellates and pogonophoran (*trnL1-trnA-trnS2-trnL2*) is also found in the echiuran *Urechis caupo* (Boore, Lavrov, and Brown 1998). Some view echiurans as an outgroup to the Annelida (Brusca and Brusca 1990); if this is correct, then the parsimonious reconstruction would be that this is the primitive annelid condition, with the change being derived for *Platynereis.* However, others view echiurans as the sister taxon to clitellates to the exclusion of both pogonophorans and the polychaete family Nereidae (McHugh 1997), so this arrangement could, alternatively, be derived from the primitive one found in *Platynereis* (*trnL1-trnS2-trnA-trnL2*) for a pogonophoran-echiuran-clitellate clade. The positions of *trnC* and *trnM* are the same in the *Katharina* and the *Lumbricus* mtDNAs (*trnC-trnM-rns*), so this can be inferred as the primitive annelid condition, with translocations of these two genes in the lineage leading to *Platynereis.*

Comparisons of mitochondrial genomes using both gene arrangements and inferred amino acid sequences provide strong support for an Annelida-Mollusca clade that excludes Arthropoda, as has been found in other studies (e.g., Ghiselin 1988; Lake 1990; Eernisse, Albert, and Anderson 1992; Morris and Peel 1995; Garcia-Machado et al. 1999). This revised view of the relationship of these phyla (i.e., accepting the Eutrochozoa rather than the Articulata as the correct superphylum group) compels a reinterpretation of the patterns of morphological evolution. For example, the parsimonious interpretation is that body segmentation is primitive for all three taxa. Traditionally, mollusks have been viewed as unsegmented, but some have serially arranged structures, notably Polyplacophora, Monoplacophora, and Nautiloidea (Vagvolgyi 1967; Lemeche 1959; Wingstrand 1985). These have often been interpreted as not being ‘‘true’’ segmentation, due partly to the lack of metamerism in groups thought to be primitive mollusks, the solenogasters and caudofoveates, and in groups thought to be closely related to mollusks, such as sipunculids and echiurans. It may also be that the long-standing description of mollusks arising from a nonsegmented ancestor, either the hypothetical ancestral mollusk (see Ghiselin 1988 for discussion) or a flatworm (see Haszprunar 1996), has stifled the more straightforward interpretation of molluscan iterative structures as being segmental.

Finally, these data indicate that Pogonophora is more closely related to the Clitellata than is *Platynereis* (order Phyllodocida, family Nereidae). This is consistent with the results of comparing partial EF1-α sequences (McHugh 1997). Thus, the Pogonophora should no longer be considered an independent phylum, but rather a group within Annelida, and should revert to the name Siboglinidae Caullery, 1914 in accordance with the suggestion of Rouse and Fauchald (1997). The current status of the Polychaeta is uncertain, with recent studies pointing out the possibility that they are a paraphyletic group (McHugh 1997, Kojima 1998). With the inclusion of only a single representative, this study cannot address this specifically. Whether or not other polychaete groups would cluster with *Platynereis* awaits further study.

Acknowledgments

We are grateful to Monica Dixon and David Weisblat for the gift of *Helobdella* DNA, to Daniel Sellos for *Platynereis* DNA, and to Robert Vrijenhoek and Mike Black for *Galathealinum* DNA. Thanks to Mathieu Blanchette for performing the minimum-breakpoint analysis of gene arrangements using his computer program and to Susan Fuerstenberg, Kevin Helfenbein, Dennis Lavrov, Markus von Nickisch-Rosenegk, and Lewis Rosenberg for comments on an earlier draft of this manuscript. This work was supported by DEB-9807100 from the NSF to J.L.B. and W.M.B.

APPENDIX 1

Abbreviated sequence of a portion (8,925 nt) of the mtDNA of the polychaete *Platynereis dumerii.* Transfer RNAs and noncoding nucleotides are shown in their entirety, but protein-encoding genes are greatly abbreviated in order to be concise. The numerals between slashes are the numbers of intragenic nucleotides omitted. The plus signs indicate that more sequence of the *rnl* and *cob* genes remains undetermined beyond the limits of this fragment. Transfer RNA genes are designated by the amino acid with which they are charged and, for the two serine and two leucine tRNAs, differentiated by their anticodons. Asterisks mark stop codons, either abbreviated or complete. Carets mark nucleotides that could form a complete stop codon if there were 1- or 2-nt overlaps with the downstream gene.

```
        10        20        30        40        50        60        70        80        90       100
TTAGATTCTATCCTTATTATCTAAATAAAACATAGTACGAAAGGACCTGGCTTTATAGATTAGATCTTTATAAATTAATATTATTAATGAGTTGGCAGAG
+_______________________________________rnl__________________________________________
                                                                                     ___trnL1(tag)__

       110       120       130       140       150       160       170       180       190       200
TTATGCACTTGACTTAGGATCAACCCACAGGAATAATCCTACTCATTACACATGCAACATACCAAGTGATTTTGAAAGTCATAACAGGTAGTTAATCTTA
___________________trnL1(tag)___________________     ___________________trnS2(tga)__________________

       210       220       230       240       250       260       270       280       290       300
CCATGTTGTTAAGGCTATATAGCCTAACTCAAGGCATTTGAATTGCAATCAAACCATGTATTTTATACTATAGCCGTTTTAGTGGCAGATTAGTGCATTA
__________                                                                 _________trnL2(taa)______
            ______________________________trnA______________________________

       310       320       330       340       350       360                1230      1240      1250
GATTTAAGCTCTAAACAAGATAAATTATCCTAAAACAATGCTACTTAAGTCCCCAATTACCAT-/870/-TAACTCTAGTATTAATAACTCTTATAATTA
_____________trnL2(taa)______________ M  L  L  K  S  P  I  T  I       M  T  L  V  L  M  T  L  M  I
                                     ____________________________nad1_______________________________

      1260      1270      1280      1290      1300      1310      1320      1330      1340      1350
CGACAGCCCTATGGTGTTGCGCCGGACGAACGGACAACTTTGATGACGTTGAATATGAGAATCTTCTCCAACACCTTTCTTAGAAGCTTGCAACAGCAGT
T  T  A  L *_____________________________trnI_______________________________
____nad1____                                                                ___________trnK_________

      1360      1370      1380      1390      1400      1410      1420               1710      1720
GAATTTTTACTTCATTAATAGAGGGACCTCCTCCTAAGAAAATGAATATAACTATTATTAATACGCTAATTG-/278/-CATGAATGAAATGAAGGCTCT
_________________trnK____________________ M  N  M  T  I  I  N  T  L  I          H  E  W  N  E  G  S
                                         _________________________________nad3______________________

     1730      1740      1750      1760      1770      1780      1790      1800      1810      1820
TTAGAATGAAAATAAGCTATGCTGGCGCAGATGAAAGCTTCTAACTTTATATCTGAACGGTTCAACTCCTTTCATAGCTTTATGACACCTTCATCTCTTT
 L  E  W  K *^^___________________________trnS1(tct)_____________________________ M  T  P  S  S  L
____nad3_____                                                                    _______nad2________

     1830             2750      2760      2770      2780      2790      2800      2810      2820
TATTTTTTT-/910/-AATTCTTACATCATTATGTACACTAACGCCCTTAATACTATACCTATAGGCCCTGTAGTTGATATACAACATTAAATTGCAGCT
L  F  F           I  L  T  S  L  C  T  L  T  P  L  M  L  Y  L **^_________________trnC______________
___________nad2_________________________________________________

  2830      2840      2850      2860      2870           4370      4380      4390      4400
TTAAAAGAGTGTCAAACACCAGGGCTTTATGCGCTGATTTTTTTCCAC-/1493/-CAGTCATTATTACATCATCTTTAAGTTGAAGCCAATAAGGCATT
___________trnC____________  M  R  W  F  F  S  T        T  V  I  I  T  S  S *___________trnN________
                            ______________________cox1_______________________

       4420      4430      4440      4450      4460      4470      4480      4490      4500
TATTTGTTACATAAACCTCTGCTATAAATTAGCCTTCTTAAATGTCTCACTGAAGACAAATCGCATTCCAAGACGCCGCCACACCTATTATAA-/600/-
______________trnN_______________________ M  S  H  W  S  Q  I  A  F  Q  D  A  A  T  P  I  M
                                         ______________________________________________________cox2_

      5110      5120      5130      5140      5150      5160      5170      5180      5190      5200
CAGTTTTATTAGATGAATCAACTCATTTAAAGAAGAAGAATAACCAACAAACCCACATAATACAGAATAAATAAAAGGAAAGCCTAGCTATAGAGCTACA
  S  F  I  S  W  I  N  S  F  K  E  E  E ***
___________________cox2____________________

      5210      5220      5230      5240      5250      5260      5270      5280      5290      5300
TCATATTACGTTAGTATAATAAGTACAGCTGCCTTCCAAGTAGCAAGTTTGGCGACCAAACGTAATAGTTTTTTATAATGTTTTATTTTTCATTTTATAA
    ____________________________trnG_______________________________

      5310      5320      5330      5340      5350      5360      5370      5380      5390      5400
ATATCACTTATATTTAGTTATATATAATTACATATATATATATATATATATATATATATATAATATAACAATTAATGTATATTAAATTTTTAATATGGACTA

      5410      5420      5430      5440      5450      5460      5470      5480      5490      5500
TTAATATTAATTTGTAGGGTATTATATAATATTTAATCCAATTATTCTTATAAAATACATCGTCTTATTAAACTTCTATTGTAAAGCAGGGCGCAACGAC

      5510      5520      5530      5540      5550      5560      5570      5580      5590      5600
TGCCCGGCTGAGGGCTTCCGCTCATAAACTGGCGGCCCTTCGCCGTTGTCGGCGCCGCTAATGGGAAAATAACGAATTTATAAGAAAAATATGATAGAGC

      5610      5620      5630      5640      5650      5660      5670      5680      5690      5700
CATATATTTTGAGTATTATATATATAACGCTGATTTTTTGATAGATGCTATTTCAGAAATGGGGGTTCGTCCCAAATTTGCTCAATCCACGTTTTTGTGC

      5710      5720      5730      5740      5750      5760      5770      5780      5790      5800
ATTTTTGGTTATAAATTCTTTACAGTAAAACACCTTATCAACACCAAATTTTTCATCAAGTTTTTCAATTACATAAAGGGTCCTTGATAAAATAAGTCAT

      5810      5820      5830      5840      5850      5860      5870      5880      5890      5900
TAGTTAACATCAATCCGTCATTTTGCGCCTATTTTTCAAAATCACAAATAATTTAAATTTTACCTGAAAACCTAGCTGGTTTTTTAGTATTAATTTATAT

      5910      5920      5930      5940      5950      5960      5970      5980      5990      6000
AATCAATTTATTAAAGGGTATAGTGTTACACCAAATATCATAATTTCATAAATGTAGTAAAATCCCGTTTAAAATTACTATGACCCTTATTTCACCACC
```

```
      6010      6020      6030      6040      6050      6060      6070      6080      6090      6100
TTTCACCTAATAGAGGGGGGTTTTCTGATGTTTCGTTAAATGGTGACGGATTGACCCTTCTATTTTTCCGGACAGACCCCTTAATTGGCATGTATTTTAT

      6110      6120      6130      6140      6150      6160      6170      6180      6190      6200
CAGAAACCCCCATTGCCTGTAACAGCACTAACCAATATTCTCTATTCGATAGGGGGTTTAAGAGAAAACCACTTTTTCCACTTAAGCAAAAAAAAAACGC

      6210      6220      6230      6240      6250      6260      6270      6280      6290      6300
TGATTTTTATAAATTTTTATTAAAATATACAGCCCAGCTTTAAAGTTTATCTAAAATTCCTTTTTTTAGTTAATATTTAGTATTAACAGAAATATTATAA

      6310      6320      6330      6340      6350      6360      6370      6380      6390      6400
ATTTACATGGCGCCTATTATTCATTATCATATTGCTTTACTATATAACCATATAATTTGTTAAGGTGGCTGACTTCATAGGCAGCGGTCTGTAAAACCGC
                                                         ____________________trnY___________________

      6410      6420      6430      6440      6450      6460      6470      6480      6490      6500
CAACGGATAATTTATCCCTTTAATATTATATGATTACATATTATCATGTTATTAATCTATTTTATAAATATTAAATTACACATGAATGCCTCACCTATCC
__________trnY___________
                                                                                      M  P  H  L  S
                                                                                    ______atp8______

      6610      6620      6630      6640      6650      6660      6670      6680      6690      6700
-/107/-TACCCTTAACGCGTCAATAACCAAGCCGTGAAAATGATGATAAAATAAGCTAATATAAGCTATTGGGCTCATACCCCGAAAATGGACTCACCC
         T  L  N  A  S  M  T  K  P  W  K  W *__________________________trnM_________________________
___________________atp8______________________

      6710      6720      6730      6740      6750      6760      6770      6780      6790      6800
CCTTTTATCAAATAGAATTCTAGTTAAAATATAATATATGCTTGTCAAGCATAAGTTGCTTTATTAGCGAGTTCTAGCTGTAATATTATTAAATTACTCA
__________   _____________________________trnD______________________________

      6810      6820      6830      7560      7570      7580      7590      7600      7610      7620
ATATAGCCATGGTACGCCAACCATTTCATT-/727/-TTATCTATGTATTTACTGATGAGGAGCTTTATAATATAGTGTACACTGCACACAGGATTTTGA
         M  V  R  Q  P  F  H           Y  L  C  I  Y  W  W  G  A *__________________trnQ____________
        ________________________cox3______________________________

      7630      7640      7650      7660      7670      7680             8110      8120      8130
TACCTGAAGAAGACCCCCTCTTATTATAAGGATATCCTATGACATTATTAACTTTAAACA-/420/-TGCGCCCCTTCAACTATGTTCAGCCCTGTTCGT
___________trnQ______________
                                       M  T  L  L  T  L  N         L  R  P  F  N *
                                      ________________________nad6________________ M  F  S  P  V  R
                                                                                  ________cob_______

   8140      8150      8160      8170      8180      8190      8200              8910      8920
AAAACTCACCCAGGCATTAAAATTGCTAATGGCGCACTAGTAGACCTGCCAGCACCTGGCAATTTAT-/700/- CTTAGTTACCCCTATTCATATTAAA
 K  T  H  P  G  I  K  I  A  N  G  A  L  V  D  L  P  A  P  G  N  L            L  V  T  P  I  H  I  K
_____________________________________________________________________cob___________________________+
```

APPENDIX 2

Abbreviated sequence of a portion (7,576 nt) of the mtDNA of the pogonophoran *Galathealinum brachiosum.* Annotations are as in appendix 1.

```
         10        20        30        40        50        60        70        80        90       100
AAATAAAACTAATAGTTATTAGTACGAAAGGACCATAATTATTAAAAATTTTAAATCTATAAGAAATCATATAAATTTATTTAATTTAAACAAGTTGGCA
                                                                                        ____________
+_________________________________________rn1___________________________________________

        110       120       130       140       150       160       170       180       190       200
GATTAGTGCAATTGATTTAGGTTCATTATATGGGTTACCACTTGTTATTATTCTAGTTTATATAGAACTATCAATTTGCATTTGATAAGAGGAATTTCCG
__________________trnL1(tag)___________________
                                              _________________________trnA_________________________

        210       220       230       240       250       260       270       280       290       300
AATGATGGTTTAATATTTCGGAAATAAATGATTTTGAAAATCATTAAAAAATGTTCAACTCATTTTTAAATCTATTTAAATGGCAGATTAGTGCTTTAGA
_______                                                                  _________trnL2(taa)________
      ____________________________trnS2(tga)_____________________________

        310       320       330       340       350                 1250      1260      1270
TTTAAATTCTAACTATAAAAATATTTTTTTTTAAATAATGAGGTCTATTATTATT-/885/-TCTATTCTTTTCTTTTTATAATATTATGCCAGATTAAT
_____________trnL2(taa)______________ M  S  S  I  I  I         S  I  L  F  F  L  *^^_____trnI______
                                     ____________________nad1____________________

        1290      1300      1310      1320      1330      1340      1350      1360      1370
TGGATAACTTTGATGAAGTTAAATATAGAAATAATTCTTAATATTTATTAGAAAGCTCATAGCCTTGAACTTTTAATTCAATCAAAATATAAATTTATAT
____________________trnI______________________
                                             ______________________________trnK_____________________

80      1390      1400      1410             1720      1730      1740      1750      1760      1770
ATTTCTAATAAATGTTATTAAGTTTTTATTTC-/300/-TGAAATGAAGGATCTTTAAAATGATTAAAATAAAAAAACTTGTGTAAAAGTCACCTTCTAA
__trnK_____ M  L  L  S  F  Y  F         W  N  E  G  S  L  K  W  L  K  *^^_________trnS1(tct)________
           _______________________nad3________________________________

       1780      1790      1800      1810      1820            2800      2810      2820
GTTACTTTTGGTTGGTTCAATTCCACCTTTTTTTTATGAAAAATTTTATT-/969/-ATTTACTTTTCTTTTTATGCGTTGACTTTA-/1514/-CTCCT
__________trnS1(tct)_______________ M  K  N  F  I         I  Y  F  S  F *
                                   _________________nad2_________________ M  R  W  L  Y        S  P
                                                                         ________________cox1_______

       4350      4360      4370      4380      4390      4400      4410      4420      4430
AATTTTCTAAATAATTTAAATAGAAGCTAATTTAGCAATTAACTGTTAATTAATCATTTACTATTATTATATAGTCTATTTAAAATGGCACAATGAAATC
 N  F  L  N  N *_______________________________trnN________________________________  M  A  Q  W  N
________________                                                                    _______cox2_____

                5100      5110      5120      5130      5140      5150      5160      5170      5180
AACTT-/648/-TCTTTATTTAATAATTAGAAATTTAGTTAATTTATAATATAAAATTGTCAATTTTAAGTTACTTAAAAGTAATTTCTAATGCCTCATC
Q  L         S  L  F  N  N *^^___________________________trnD_____________________________ M  P  H
_______cox2________________                                                               ___atp8___

             5320      5330      5340      5350      5360      5370      5380      5390      5400
TT-/130/-AATGAAAATGATAATAAAATAAGATGACTGATATAAAGTAAAAGATTGTAACTCTTTCTATGGTTTAACCCTCTTATTTCTTTTTTAGTAT
L        K  W  K  W ***   _____________________________trnY_____________________________
____atp8______________                                                                 ______trnG___

5410      5420      5430      5440      5450      5460      5470      5480
AAAAAGTACATTTCTTTTCCACAGAAAAAGTTTATTAAAAAAAAGAAATGATCCGTCAACCATATCATATTGTTGAATTTAGT-/721/-TAATTTATTG
____________________trnG_______________________ M  I  R  Q  P  Y  H  I  V  E  F  S        L  I  Y  W
                                               _____________________________________cox3____________

         6230      6240      6250      6260      6270      6280      6290      6300      6310
ATGAGGATCTTAATTTATAAATATGTTAGGTACAAAAAAGAATTTTGAATTCTTTATATTTAGTTCAATTCTAAAATTTATAAATGTTAATAATATTAAT
  W  G  S ***  ________________________________trnQ________________________________ M  L  M  M  L  M
_____________                                                                      ________nad6_____

         6330            6750      6760      6770      6780      6790           7560      7570
ATCCTTATTAATA-/411/-AAAAGTCCGTTACGTCCATTTATATATGTTTAAACCTATTCATAAATCT-/762/-TTAGTAACACCAATTCATATTAAA
  S  L  L  M         K  S  P  L  R  P  F  M *
_______________nad6__________________________ M  F  K  P  I  H  K  S         L  V  T  P  I  H  I  K
                                             __________________________cob_________________________+
```

APPENDIX 3

Abbreviated sequence of a portion (7,553 nt) of the mtDNA of the leech *Helobdella robusta.* Annotations are as in appendix 1.

```
        10        20        30        40        50        60        70        80        90       100
TCTTCAATAAAATAAATTTATTTGTAGTACGAAAGGACCCAAATAACCCTATAATAGGATTAAATAAAATTAACTATATACTATAATAAGTTGGCAGATA
+______________________________________rn1__________________________________________
                                                                                    ___trnL1(tag)___

       110       120       130       140       150       160       170       180       190       200
AATGTAAATGATTTAGGATCATTAGATGAATTACAATTCACTTATTATGCAGCTTAGTTTAAATAGAACAGTTGATTTGCAATCGACAAGTGATGACAAA
__________________trnL1(tag)___________________
                                               _________________________trnA________________________

       210       220       230       240       250       260       270       280       290       300
TCAGTGGCAGTTATTTGTTTAGGAAACAATAGATTTTGAAAATCTAAAACAAATGTTCGAATCATTTAATAACTTAATATGATGGCAGAGTAGTGCATTA
__________                                                                 _________trnL2(taa)______
        ______________________________trnS2(tga)___________________________

       310       320       330       340       350                1260      1270      1280      1290
GGTTTAAACCCTAAATATGAATAAAATCTCATATTAATGAACCTATCAGCTTCAATA-/897/-ATCTTAAATACTATAATTTAATAGTACTAAGCCGGA
____________trnL2(taa)______________ M  N  L  S  A  S  M         I  L  N  T  M  I *** ______trnI____
                                    _______________________nad1______________________

      1300      1310      1320      1330      1340      1350      1360      1370      1380      1390
CTAACGGATAACATTGATGTCGTTAAATATGAGTAATTTCTAGTACTTATTCAGAGAGCTTGTCTAAGCATTCGACTTTTAATCGAAAGAAAGTTTTTAC
_____________________trnI_______________________
                                                __________________________trnK______________________

      1400      1410      1420             1730      1740      1750      1760      1770      1780
TTCTGATTAATGCTTATTATAACAATTATT-/300/-TGAAATGAAGGCTCTATTAATTGACTTTCTTAGGGGTTTAAGTATAATAAAAGCTTCTAACTT
__trnK___ M  L  I  M  T  I  I         W  N  E  G  S  I  N  W  L  S *^^____________trnS1(tct)________
         _______________________nad3________________________________

   1790      1800      1810      1820      1830                 2810      2820
TTTTAAGGTGGTTAAATTCCATCAACCCCTTATGATTATTTCCTTCATC-/966/-TTACTCTACATTATTTATGCGATGACTCTAC-/1501/-CAGGA
_________trnS1(tct)____________ M  I  I  S  F  I         L  L  Y  I  I  *
                               ____________________nad2_________________ M  R  W  L  Y         T  G
                                                                        _________________cox1_______

 4340      4350      4360      4370      4380      4390      4400      4410      4420      4430
ATTGTCACTACCCCATTAAGTGAAAGCAAATAAATTGCACCTAACTGTTAATTAGGAGTAAGTCATAGACTCACTTATATGCCTTATTGAGGACAACTAC
 I  V  T  T  P *____________________________trnN______________________________ M  P  Y  W  G  Q  L
____cox1________                                                              _________cox2_________

 4440      5080      5090      5100      5110      5120      5130      5140      5150      5160
TGCTT-/633/-TGATTAAAAACATAAGATTTTAGTTAAACCATAACAGTAGTCTGTCAGCCTACAATTACCCTCCAGGTAAATCTTAATGCCACACCTT
L  L         W  L  K  T *^^____________________________trnD_____________________________ M  P  H  L
_______cox2______________                                                               _____atp8___

          5300      5310      5320      5330      5340      5350      5360      5370      5380
GCC-/129/-AGTTGAAAATGATAACAAGATGGTCGAAATATAGACAGTAAACTGTAAATTTATTCATGAGTAAATTACTCTCTTGTATGATATTAGTAT
 A         S  W  K  W *^^____________________________trnY________________________________
_____atp8______________                                                                _____trnG____

5390      5400      5410      5420      5430      5440      5450      5460                      6200
AATAAGTACAATTACCTTCCAAGTAATAAGTTTAATTTAAATATCATATTGCTCCGACAACCCTATCATTTAGTAGAACCAAG-/721/-TGCATGTATT
______________________trnG______________________(M) L  R  Q  P  Y  H  L  V  E  P  S        C  M  Y
                                                _____________________________________cox3___________

      6210      6220      6230      6240      6250      6260      6270      6280      6290      6300
GATGAGGCTCATTATTTATTAGTGTACGGTACACATAAACCTTTGAAGTTTAAATAATAAGTTCAATTCTTATATTAATAAATGTCCTTAATCGTAATAA
W  W  G  S *________________________________trnQ_________________________________ M  S  L  I  V  M
___cox3_____                                                                     _________nad6______

                   6730      6740      6750      6760      6770           7530      7540      7550
ACATA-/415/- AGGACCTCTACGACCATTCATATATGTTTAAACCATTCCGAAGCCATCAT-/753/-AGGTCATCTGTAACCCCTATTCACATTAAA
N  M         K  D  L  Y  D  H  S **
_______nad6________________________  M  F  K  P  F  R  S  H  H         S  S  S  V  T  P  I  H  I  K
                                    _____________________________cob_______________________________+
```

LITERATURE CITED

Barnes, W. M. 1994. PCR amplification of up to 35-kb DNA with high fidelity and high yield from bacteriophage templates. Proc. Natl. Acad. Sci. USA **91**:2216–2220.

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

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

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. 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.

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

Boore, J. L., L. L. Daehler, and W. M. Brown. 1999. Complete sequence, gene arrangement and genetic code of mitochondrial DNA of the cephalochordate *Branchiostoma floridae* (''amphioxus''). Mol. Biol. Evol. **16**:410–418.

Boore, J. L., D. Lavrov, and W. M. Brown. 1998. Gene translocation links insects and crustaceans. Nature **392**:667–668.

Brusca, R. C., and G. J. Brusca. 1990. Invertebrates. Sinauer, Sunderland, Mass.

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, N. Satoh, and S. Pääbo. 1998*a.* Codon reassignment and amino acid composition in hemichordate mitochondria. Proc. Natl. Acad. Sci. USA **95**:3703–3707.

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

Chang, Y.-S., F.-L. Huang, and T.-B. Lo. 1994. The complete nucleotide sequence and gene organization of carp (*Cyprinus carpio*) mitochondrial genome. J. Mol. Evol. **38**:138–155.

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

Eernisse, D. J., J. S. Albert, and F. E. Anderson. 1992. Annelida and Arthropoda are not sister taxa: a phylogenetic analysis of spiralian metazoan morphology. Syst. Biol. **41**: 305–330.

Felsenstein, J. 1978. Cases in which parsimony and compatability methods may be positively misleading. Syst. Zool. **27**:401–410.

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. Biotechnol. **3**:294–299.

Garcia-Machado, E., M. Pempera, N. Dennebouy, M. Oliva-Suarez, J. C. Mounolou, and M. Monnerot. 1999. Mitochondrial genes collectively suggest the paraphyly of Crustacea with respect to Insecta. J. Mol. Evol. **49**:142–149.

Ghiselin, M. T. 1988. The origin of molluscs in the light of molecular evidence. Oxf. Surv. Evol. Biol. **5**:66–95.

Hasegawa, M., H. Kishino, and T. Yano. 1985. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. J. Mol. Evol. **22**:160–174.

Haszprunar, G. 1996. The mollusca: coelomate turbellarians or mesenchymate annelids? Pp. 1–28 *in* J. Taylor, ed. Origin and evolutionary radiation of the Mollusca. Oxford University Press, Oxford, England.

Kojima, S. 1998. Paraphyletic status of Polychaeta suggested by phylogenetic analysis based on the amino acid sequences of elongation factor-1α. Mol. Phylogenet. Evol. **9**:255–261.

Lake, J. A. 1990. Origin of the Metazoa. Proc. Natl. Acad. Sci. USA **87**:763–766.

Lemche, H. 1959. Protostomian interrelationships in the light of *Neopilina.* Pp. 381–389 *in* Proceedings of the XVth International Congress of Zoology, London.

McHugh, D. 1997. Molecular evidence that echiurans and pogonophorans are derived annelids. Proc. Natl. Acad. Sci. USA **94**:8006–8009.

Morris, S. C., and J. S. Peel. 1995. Articulated halkieriids from the Lower Cambrian of North Greenland and their role in early protostome evolution. Philos. Trans. R. Soc. Lond. Biol. Sci. **347**:305–358.

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

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

Palumbi, S., A. Martin, S. Romano, W. O. McMillan, L. Stice, and G. Grabowski. 1991. The simple fool's guide to PCR. Version 2.0.

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.

Raff, R. A., and T. C. Kaufman. 1983. Embryos, genes and evolution. Macmillan, New York.

Rasmussen, A.-S., and U. Arnason. 1999. Phylogenetic studies of complete mitochondrial DNA molecules place cartilaginous fishes within the tree of bony fishes. J. Mol. Evol. **48**:118–123.

Rouse, G. W., and K. Fauchald. 1995. The articulation of the annelids. Zool. Scripta **24**:269–301.

———. 1997. Cladistics and polychaetes. Zool. Scripta **26**: 139–204.

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

Sankoff, D., and M. Blanchette. 1998. Multiple genome rearrangement and breakpoint phylogeny. J. Comput. Biol. **5**:555–570.

Sankoff, D., G. Leduc, N. Antoine, B. Paquin, B. F. Lang, and R. J. Cedergren. 1992. Gene order comparisons for phylogenetic inference: evolution of the mitochondrial genome. Proc. Natl. Acad. Sci. USA **89**:6575–6579.

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

Smith, A. E., and K. A. Marcker. 1968. *N*-formylmethionyl transfer RNA in mitochondria from yeast and rat liver. J. Mol. Biol. **38**:241–243.

Smith, M. J., A. Arndt, S. Gorski, and E. Fajber. 1993. The phylogeny of echinoderm classes based on mitochondrial gene arrangements. J. Mol. Evol. **36**:545–554.

Snodgrass, R. E. 1938. Evolution of Annelida, Onychophora and Arthropoda. Smithson. Misc. Collect. **97**:1–77.

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

Staton, J. L., L. L. Daehler, and W. M. Brown. 1997. Mitochondrial gene arrangement of the horseshoe crab *Limulus polyphemus* L.: conservation of major features among arthropod classes. Mol. Biol. Evol. **14**:867–874.

Swofford, D. L. 1998. PAUP*, phylogenetic analysis using parsimony (*and other methods). Version 4. Sinauer, Sunderland, Mass.

Vagvolgyi, J. 1967. On the origin of molluscs, the coelom, and coelomic segmentation. Syst. Zool. **16**:153–168.

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

Wingstrand, K. G. 1985. On the anatomy and relationships of recent Monoplacophora. Galathea Rep. **16**:7–94.

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

B. Franz Lang, reviewing editor

Accepted September 21, 1999