The mitochondrial genome of the sipunculid *Phascolopsis gouldii* supports its association with Annelida rather than Mollusca

Jeffrey L. Boore[1,2] and Joseph L. Staton[1,3]

[1] Department of Biology, University of Michigan

830 N. University Avenue

Ann Arbor MI 48109-1048 USA

[2] DOE Joint Genome Institute and Lawrence Berkeley National Laboratory

2800 Mitchell Drive

Walnut Creek, CA 94598 USA

[3] Belle W. Baruch Institute for Marine Biology and Coastal Research

University of South Carolina

Columbia, SC 29208, USA

Corresponding author: J.L. Boore, (925) 296-5691, FAX (925) 296-5666,

JLBoore@LBL.gov

List of keywords: Phascolopsis, mitochondria, evolution, phylogeny, genome, Annelida

Running title: Phascolopsis mtDNA

Abstract

We have determined the sequence of about half (7470 nts) of the mitochondrial genome of the sipunculid *Phascolopsis gouldii*, the first representative of this phylum to be so studied. All of the 19 identified genes are transcribed from the same DNA strand. The arrangement of these genes is remarkably similar to that of the oligochaete annelid *Lumbricus terrestris*. Comparison of both the inferred amino acid sequences and the gene arrangements of a variety of diverse metazoan taxa reveals that the phylum Sipuncula is more closely related to Annelida than to Mollusca. This requires reinterpretation of the homology of several embryological features and of patterns of animal body plan evolution.

Introduction

Complete mitochondrial genome sequences are available for 56 invertebrate animals (Boore 1999 and Mitochondrial Genomics link at http://www.jgi.doe.gov). With a few notable exceptions, animal mitochondrial DNAs (mtDNA) are circular molecules, about 16 kb in size, containing 37 genes: 13 for proteins of electron transport (cytochrome oxidase subunits I-III [*cox1-3*], cytochrome *b* [*cob*], NADH dehydrogenase subunit 1-6, 4L [*nad1-6, nad4L*], ATP synthase subunits 6, 8 [*atp6, 8*], two for rRNAs (large and small subunit rRNA, *rrnL* and *rrnS*, respectively), and 22 for tRNAs (abbreviated *trnX*, where "X" is the one-letter code for the corresponding amino acid, with anticodon where necessary for uniquely specifying identity). In a few lineages, the arrangement of these genes varies greatly, such as in bivalve (e.g. Hoffmann, Boore, and Brown 1992) and gastropod (e.g. Yamazaki et al. 1997) mollusks. However, in other lineages, it has remained unchanged for very long periods of time. For example, the arrangements of the 37 mitochondrial genes of the horseshoe crab (Staton, Daehler, and Brown 1997; Lavrov, Boore, and Brown 2000) and of the fruit fly (Clary and Wolstenholme 1985) differ by the location of only one tRNA gene.

Comparisons of mitochondrial gene arrangements have been very effective for reconstructing evolutionary relationships (Smith et al. 1993; Boore et al. 1995; Boore and Brown 1998, 2000; Boore, Lavrov and Brown 1998; Dowton 1999; Stechmann and Schlegel 1999; Kurabayashi and Ueshima 2000). Despite the occasional finding of mtDNAs with "scrambled" gene arrangements, the data set remains nearly free of homoplasy (Boore 1999, 2000). Since the rate of rearrangement varies among lineages, these gene order characters may provide resolution at differing taxonomic levels. Within

rapidly changing lineages the signal may be best at lower levels, with anciently shared arrangements being eroded. Conversely, in more slowly evolving lineages the more ancient signal may be retained, but without rearrangements to resolve more recent branching. The merits of using gene rearrangements for recovering phylogenetic patterns lie not in their infrequency (as has been suggested, e.g. Saccone et al. 1999; Le et al. 2000), but in their complexity (i.e. there are a very large number of potential arrangements) and their near-irreversibility (as judged by the infrequency of identified homoplasy). Thus, the ability to resolve any particular phylogenetic relationship may depend on the rate of gene rearrangement, but the confidence in a clade that is defined by a rearrangement, as analyzed by phylogenetic methods, does not.

In addition to providing gene order characters for inferring phylogeny, this dataset also allows comparison of sequences as an independent estimator of evolutionary relationships. Further, mitochondrial genome comparisons serve as a model system for genome evolution, where we can examine such factors as: What determines whether genes are all encoded by the same DNA strand vs. being distributed between the two strands? How does mutation bias influence A+T-richness, dinucleotide frequency, codon usage, strand-skew between purines and pyrimidines, and amino acid composition of proteins? How do the structures of tRNAs evolve?

We describe here the partial mitochondrial genome of *Phascolopsis gouldii*, the first representative of the phylum Sipuncula to be so examined. The placement of this phylum within metazoan evolution has varied since its recognition as a taxon. Lamarck (1816) confused sipunculans with sea cucumbers (Holothuroidea), and they were later considered to be a derived group of annelids (Delle Chiaje 1823). In his reorganization

of metazoans, Quatrefages (1847) created the Gephyrea, or "bridge group", that contained sipunculans, echiurans, sternapsid annelids and priapulans. With their lack of segmentation and simple internal anatomies, Quatrefages envisioned these groups as the transitional forms from "lower" to "higher" metazoan archetypes. Sipunculans were later elevated to the phylum level (Sedgwick 1898) and associated with other spiralians either as a sister taxon to Annelida on the basis of biochemical properties (Florkin 1976, Henry 1987), or as sister to mollusks based on similarities in development (Scheltema 1993, 1996).

The developmental lynchpin that supports a sipunculan/molluscan affiliation is a shared "molluscan cross" cleavage pattern in the blastomeres (Baba 1951; Heath 1899; Gerould 1907; van Dongen and Geilenkirchen 1974; Rice 1975, 1985; Verdonk and van den Biggelar 1983), which has been interpreted as a synapomorphy uniting these two groups (Scheltema 1993, 1996). Annelids and echiurans possess a different arrangement of the blastomeres termed an "annelid cross" (Gerould 1907). As an extension of this proposed relationship, Scheltema (1996) hypothesized further homology between sipunculan larval characteristics to those of larval and adult mollusks including: the ventral buccal organ of sipunculan larvae with the odontophore of mollusks, the ciliated lip below the mouth of sipunculan pelagosphera stage larvae with the foot of a molluscan pediveliger larva, and the lip glands of sipunculan larvae with pedal glands of larval chitons and adult neomenioid aplacophorans.

We report the sequence of about half (7470 nts) of the mtDNA of the first representative of the phylum Sipuncula, *Phascolopsis gouldii*. We analyze genomic features in comparison to those of other animal mtDNAs and compare both gene

arrangement and inferred amino acid sequences to resolve the phylogenetic placement of this phylum. These comparisons strongly support the closer relationship of Sipuncula with Annelida to the exclusion of Mollusca and other taxa.

Materials and Methods

Molecular Analysis

Live specimens of *Phascolopsis gouldii* (Fisher 1950) were purchased from the Marine Biological Laboratory at Woods Hole, Massachusetts. We isolated total DNA using a CTAB method described in Collins et al. (1996) from excised retractor muscles. We also isolated total DNA from the upper "nuclear" fraction of the CsCl purification method for mitochondrial DNA (Dowling et al. 1996). The upper layer of that preparation contains both nuclear and non-supercoiled (i.e. nicked) mitochondrial DNA.

Initially, primers designed to match generally conserved regions of the mtDNA were used in the PCR to amplify four short (450-710 nts) fragments from *cox1* (primers LCO1490 and HCO2198; Folmer et al. 1994), *cox3* (primers COIIIF and COIIIB; Boore and Brown 2000), *cob* (primers CytbF and CytbR; Boore and Brown 2000), and *rrnL* (primers 16SARL and 16SBRH; Palumbi 1996). DNA sequences obtained from these fragments were used to design oligonucleotides that were then employed in "long-PCR" (Barnes 1994) to amplify the portions of the mtDNA spanning *cox1-cox3*, *cox3-cob*, and *cob-rrnL*. Reaction conditions and results, fragment purification, DNA sequence determination and assembly, and gene identifications, were as in Boore and Brown (2000). All sequence was determined for both strands using synthetic oligonucleotides to

"primer-walk" through the long-PCR amplified fragments. This 7470 nt sequence was deposited in GenBank under accession number AF374337.

Phylogenetic analysis of sequences and gene arrangements

The invertebrate mitochondrial genetic code was used to infer the amino acid sequences of the six protein-encoding genes identified in the studied portion of *P. gouldii* mtDNA. Those of Cob, Cox1, Cox2, and Cox3 were aligned to the homologues of 16 other, phylogenetically diverse animals (listed, along with citations, in table 1) using ClustalW as incorporated in MacVector™ 6.5 (Oxford Molecular Group). The other two inferred protein sequences, Atp8 and Nad6, were judged to be too divergent to align with confidence, and so were not used in this phylogenetic analysis. The BLOSUM matrix was used to weight shared amino acids, with gap and extension penalties of 10 and 1, respectively. Amino acid alignments were then adjusted manually; these alignments can be viewed at EMBL accessions ALIGN_000119 and ALIGN_000121-123.

Protein alignments were analyzed using three methods: 1. maximum likelihood (ML) quartet sampling with Tree-Puzzle 5.0 (Strimmer and von Haeseler 1996) using the mtREV24 model of substitution (Adachi and Hasegawa 1996), with gamma distributed rates (eight categories) and parameters estimated from the data set, with no clock assumed; 2. equal-weighted parsimony (MP) (heuristic search of 1000 random sequence additions); and 3. neighbor-joining (NJ) analysis of total pairwise distance of amino acids (Saitou and Nei 1987). The two latter methods were each done using PAUP* ver. 4.0b4a (Swofford 2000). Gaps were treated as missing data. Data were also bootstrapped using MP and NJ methods (1000 replicates each). The single shortest tree

for the heuristic parsimony analysis was used to calculate branch support values (Bremer 1994) using the program AutoDecay (Eriksson 1998) in combination with PAUP*.

Additional analyses tested specifically the effect of grouping *P. gouldii* with the two annelids vs. with the mollusk/brachiopod group using first a Kishino-Hasegawa test (Kishino and Hasegawa) as implemented in Tree-Puzzle (http://www.tree-puzzle.de/) and secondly using a non-parametric Wilcoxon signed-ranks test (Templeton 1983) as implemented in PAUP*. For the partially determined mtDNA sequences of two other annelids, *Galathealinum brachiosum* and *Helobdella robusta* (Boore and Brown 2000), only the partial sequences of the *cob* genes are known, along with the complete sequences of *cox1*, *cox2*, and *cox3*. These were included in a separate analysis to verify that this does not affect the phylogenetic placement of *P. gouldii*.

We also compared the gene arrangement of this portion of *P. gouldii* mtDNA to those completely determined for 13 other animals (table 1). These taxa were the same as those included in the sequence analysis except for three: *Daphnia pulex* and *Ixodes hexagonus* were omitted since they have gene arrangements identical to those of *Drosophila yakuba* and *Limulus polyphemus*, respectively, and *Locusta migratoria* was omitted, since it differs from the gene arrangement of *D. yakuba* by only a single tRNA position. A matrix was constructed that scored 36 characters, each as "upstream of X" or "downstream of X", where "X" refers to each of the 19 genes identified in the studied portion of *P. gouldii* mtDNA. Coded character states were the 5-prime or 3-prime end of the adjacent gene. This matrix of scored gene adjacencies was analyzed using parsimony criteria with the programs PAUP* ver. 4.0b4a (Swofford 2000) and MacClade (Maddison and Maddison 1992).

Results and Discussion

Gene Content and Organization

This 7470 nucleotide portion of *Phascolopsis gouldii* mtDNA contains six protein-encoding genes (*cox1* is incomplete by only a few nucleotides at the 5' end), 11 tRNA genes, and two rRNA genes (*rrnL* is incomplete at the 3' end). Remarkably, the arrangement of these genes is very similar to that of the oligochaete annelid *Lumbricus terrestris* (Boore and Brown 1995; fig. 1). Among animal mtDNAs studied so far (see Boore 1999 and Mitochondrial Genomics link at http://www.jgi.doe.gov), the entire block of 11 genes spanning *cox1* through *cob* is uniquely shared between *P. gouldii* and *L. terrestris* (table 1). Further, the five genes in the block from *trnC* through *rrnL* are arranged identically between these two mtDNAs, and also in the mtDNAs of the chiton *Katharina tunicata* (Boore and Brown 1994) and the brachiopod *Terebratulina retusa* (Stechmann and Schlegel 1999). Intervening between these two blocks of genes are *trnP*, *trnE,* and *trnS2(uga)*. This is not similar to the arrangement of these genes in any other animal so far studied (Boore 1999). All of the genes identified in this portion of *P. gouldii* mtDNA are found in the same transcriptional orientation, as is true of the mtDNAs of *L. terrestris* (Boore and Brown 1995), the polychaete annelid *Platynereis dumerilii* (Boore and Brown 2000 and GenBank AF178678), an ascidian (Yokobori et al 1999), a mussel (Hoffmann, Boore and Brown 1992), the two sampled brachiopods (Stechmann and Schlegel 1999; Noguchi et al. 2000), the four sampled nematodes (Okimoto et al. 1991, 1992; Keddie, Higazi and Unnasch 1998), the seven sampled

flatworms (Le et al. 2000; GenBank accession AB018440), and a sea anemone (Beagley, Okimoto and Wolstenholme 1998).

In many mtDNAs, *atp6* follows directly after *atp8*. This is the case even for organisms that are unambiguously outgroup taxa to those considered here, including a sponge (Watkins and Beckenbach 1999) and yeast (e.g. Sekito et al. 1995). The explanation may be that the transcript of these two genes is not cleaved, so that these genes are co-translated from a single bicistron, as has been shown for mammalian mitochondrial systems (Fearnley and Walker 1986). The gene for *atp6* is not contained in the studied portion of *P. gouldii* mtDNA but, clearly, it does not follow *atp8*. So far, these genes have been found separated (see Boore 1999) only for animals that are part of the proposed superphylum Eutrochozoa (Ghiselin 1988; Mollusca, Annelida, and others, but not including Arthropoda), leading to the speculation that the loss of co-translation of these mRNAs could be a derived feature uniting this group.

Base Composition and Codon Usage

Overall, the 7470 nts determined for *P. gouldii* mtDNA are 63.1% A+T, similar to that found for the whole mtDNA sequence of *L. terrestris* (61.6%; Boore and Brown 1995). As is common among mtDNAs, CG is the least frequent dinucleotide, occurring at only 0.53% of expectation given the proportion of G and of C observed in this sequence.

Leucine is inferred to be the most frequent amino acid, present 243 times in these six inferred amino acid sequences, followed by four amino acids in nearly equal numbers, each about half as common as leucine: alanine (124), isoleucine (123), phenylalanine

(135), and serine (125). Cysteine is the least frequently used amino acid, occurring only 14 times. These values are all similar to the amino acid usage of *L. terrestris* homologues (see table 2).

The DNA strand shown in figure 2 is rich in pyrimidines (C and T). One way to measure this is called skewness, an assessment of whether the G of GC pairs (GC-skew, calculated as [G–C] / [G+C]) and the A of AT pairs (AT-skew, calculated as [A–T] / [A+T]) are more commonly on the considered strand (Perna and Kocher 1995). GC-skew for the entire 7470 nts is -0.239, for the protein-encoding portion only is -0.265, and for third codon positions (see below), which, presumably, are more free to change and therefore better reflecting mutational bias, is a remarkable -0.754. AT-skew values are much less biased: for the entire region, -0.080, for the protein-encoding portion, -0.138, and for the third codon positions, -0.028.

Gene Initiation and Termination

Most of the studied animal mitochondrial genomes contain one or more genes that initiate with some alternative to the standard ATG. Annelids appear as a notable exception; based on the few animals sampled, ATG is used as the initiation codon for all, or nearly all, genes (Boore and Brown 2000). Consistent with this, and with the phylogenetic placement of Sipuncula advanced here (see below), each of the *P. gouldii* protein-encoding genes in this portion of the mtDNA appears to start with ATG (the initiator of *cox1* is not present in this sequence).

Mitochondrial genes often terminate with abbreviated stop codons, where a single T or TA is presumably completed by polyadenylation to a TAA stop codon (Ojala,

Montoya, and Attardi 1981). This appears to be the case for three genes in *P. gouldii* mtDNA. The least certain of this is for *atp8*, since the next nucleotide (A), otherwise part of *trnY*, would complete the stop codon (fig. 2). If *cob* were to extend to the first complete stop codon, it would overlap *trnP* by 41 nts and, if *nad6* were to do so, it would overlap *cob* by 20 nts. In each case, homologous genes of related animals are similar in predicted amino acid sequences to the abbreviated forms in *P. gouldii* and have no similarity to the overlapping extensions.

All genes are in a compact arrangement, with a total of only 13 non-coding nucleotides. These are distributed in six regions of one to four nts each. There is a CC between *trnY* and *trnG* but, otherwise, all are A or T. No genes are inferred to overlap except for the possibility of a single nucleotide shared between *atp8* and *trnY* (see above).

Transfer RNAs

There are 11 sequences identified with the potential for folding into tRNA-like structures (fig. 3). Each has an anticodon matching exactly to one tRNA in *L. terrestris* mtDNA. All have a seven-member amino-acyl acceptor stem, four with a single mismatch each, and a five-member anticodon stem, again, four with a single mismatch each. Two have five nts in the extra arm and all others have four. All but four have A immediately preceding the anticodon arm. For all except tRNA(N) and tRNA(S2), there are three to five nucleotide pairs in both DHU and TΨC arms. All but four tRNAs have TA immediately preceding the DHU arm. The nucleotides preceding the anticodon are YT for all tRNAs and the nucleotide following the anticodon is A for all but tRNA(P).

It is common for tRNA(S) in many mtDNAs to lack a paired DHU arm and, in some cases, to have a six-member anticodon stem (Yokogawa et al. 1991). There is potential for *P. gouldii* tRNA(S2) to have such an extended anticodon stem, and also to have a paired DHU arm with an unusual structure, having only one nucleotide separating the amino-acyl acceptor arm from the DHU arm. Several other serine-specifying tRNAs have been identified in animal mtDNAs that have this identical pairing potential (Boore and Brown 2000).

Phylogenetic Reconstruction

Analyses based on maximum likelihood (ML), maximum parsimony (MP) and neighbor-joining of pairwise distances (NJ) of inferred amino acids are in almost complete agreement (fig. 4). In all cases, *P. gouldii* is sister to the two annelids, and the sipunculan, annelids, mollusks and brachiopods are placed within a well-supported Eutrochozoa (Ghiselin 1988). Further, there is strong support for the monophyly of several traditionally recognized groups (at least as comprised by this small sampling), such as Insecta, Chelicerata, Arthropoda, and Echinodermata. Of the 1396 aligned inferred amino acids, 673 were parsimony informative. A heuristic search with 1000 random step-wise additions of taxa yielded a single tree (CI=0.618, RI=0.366, and RC= 0.226), 4552 steps in length, identical to that in figure 4 except for the resolution of two trichotomies.

Discrepancies among the ML, MP, and NJ analyses are in the placement of the brachiopod *T. retusa* and the relationship among the crustaceans *D. pulex* and *A. franciscana*. In MP and NJ trees, *T. retusa* is placed as sister to *L. bleekeri*, although

with less than 52% bootstrap support in either case, whereas in the ML tree, *T. retusa* is placed as sister to a clade of the two mollusks (69% of quartets sampled). The two crustaceans are paraphyletic in the MP analysis, although with poor (less than 50%) bootstrap support, but group together in both the ML and NJ analyses. This association is found in 95% of the quartets in ML and with 82% of the bootstrap resamplings in NJ.

The tree shown in figure 4 was specifically compared to the alternative which repositions *P. gouldii* to be more closely related to mollusks than to annelids. Using ML, a Kishino-Hasegawa test (Kishino and Hasegawa 1989) demonstrates that the grouping of *P. gouldii* with the two annelids is significantly better (at the 5% significance level) than the grouping of *P. gouldii* with the brachiopod/mollusk clade. Using MP, non-parametric tests (Templeton 1983) of the single tree with *P. gouldii* as sister to annelids (4552 steps) and the single most parsimonious tree with *P. gouldii* sister to mollusks (4582 steps) demonstrate that the (*P. gouldii*, (*P. dumerilii*, *L. t*errestris) tree is significantly shorter ($P<0.0001$). In case of bias from the inclusion of the ambiguously placed *T. retusa* among the mollusks, a second test omitting *T. retusa* was performed. The resultant sipunculan/annelid tree in this analysis is also significantly shorter ($P<0.0001$) than (*P. gouldii*, (*L. bleekeri*, *K. tunicata*).

Two other annelids, *Galathealinum brachiosum* and *Helobdella robusta*, were included in a separate analysis, since each lacks a complete sequence for *cob* (Boore and Brown 2000). Their inclusion has no affect on the phylogenetic placement of *P. gouldii*, and these two taxa join *P. dumerilii* and *L. terrestris* in a topology consistent with the results of the earlier study (not shown).

Gene arrangements were also compared for phylogenetic reconstruction by scoring gene adjacencies as characters for cladistic analysis. The strict consensus of 30 equally parsimonious trees is shown in figure 5. The unusual placement of *Alligator mississippiensis* as basal to the other included deuterostomes is not well supported, since it is based on the sharing of a single gene boundary (*trnI, trnM*) between *Branchiostoma floridae* and *Balanoglossus carnosus* in contrast to an alternative arrangement (*trnI, –trnQ, trnM*) in *A. mississippiensis* and two outgroup taxa (*D. yakuba* and *L. polyphemus*). This shared gene boundary has evidently been created by two independent translocations of *trnQ* from its primitive position (between *trnI* and *trnM*), in one case leading to the arrangement *nad1, –trnQ, trnI, trnM, nad2* in *B. carnosus* and in the other to *nad1, trnI, trnM, –trnQ, nad2* in *B. floridae*.

Moving *P. gouldii* to make it sister taxon to a (*L. bleekeri, K. tunicata*) clade or to a ((*L. bleekeri, K. tunicata*), *T. retusa*) clade increases tree lengths by seven steps in each case for this gene arrangement analysis.

Our findings are in general agreement with several published studies (Winnepenninckx, Backeljau, and De Wachter 1995 [parsimony-based analysis]; Giribet et al. 2000 [18S-only tree]; Regier and Shultz 2000) in that sipunculans are closely associated with annelids to the exclusion of mollusks. Other studies, or portions thereof, give results to the contrary. Comparisons of 18S rDNA sequences have placed Sipuncula as basal and sister to an assemblage of worms and mollusks (Field et al. 1988); Winnepenninckx, Backeljau, and De Wachter 1995 [neighbor-joining analysis]; Giribet et al. 2000 [when combined with morphological characters] or as sister to only one of two ectoprocts analyzed (Mackey et al. 1996). The conflicts among these studies indicate that

comparisons of short DNA sequences lack the necessary resolving power at this level of relationship. Further, some of the associations between sipunculans and annelids found in these past studies were probably discounted based on the problematica associated with these data (Maley and Marshall 1998).

Conclusions

The results of this study constitute the first genomic-level sequence from the phylum Sipuncula, and the first molecular evolutionary treatment of the group by comparing mitochondrial genomes. The shared mitochondrial gene orders suggests a close phylogenetic affinity between sipunculan and annelid taxa, and phylogenetic analysis of their combined amino acid sequences also reflects this relationship. In short, sipunculans and annelids form a natural group within the Eutrochozoa separate from that of mollusks. While it might seem controversial from a modern evolutionary perspective, a vermiform baüplan may not represent the simplest ancestral holdover or a failure to develop elaboration beyond the simplest hydrostatic skeleton, but rather it may ultimately prove to be a unifying characteristic among Eutrochozoan worms.

Likewise, many characteristics that have been hypothesized to link sipunculans with mollusks, e.g., developmental pattern, lack of segmentation, etc., must be reevaluated. The presence of a “molluscan cross” may be ancestral in all eutrochozoans and subsequently lost in the annelids, or possibly is a convergent pattern that has independent origins in the two phyla. Recent reports of segmental development in chitons (Jacobs et al. 2000) suggests that segmentation in some form may be present in all spiralians, so its loss in other mollusks and worm groups may not be surprising. Certainly, the detailed

examination of hypothesized synapomorphic larval and adult morphology in the sipunculans and molluscan groups (Scheltema 1993, 1996) will need careful reconsideration to exclude convergence or oversight of annelid larval and adult homologues.

Acknowledgments

Thanks to Inaki Ruiz-Trillo for assistance with the program Tree-Puzzle and to Kevin Helfenbein for helpful comments on the manuscript. This work was supported by NSF grants DEB-9807100 and DEB-9303301. 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

ADACHI, J. and M. HASEGAWA. 1996. Model of amino acid substitution in proteins encoded by mitochondrial DNA. J. Mol. Evol. **42:**459-468.

ASAKAWA, S., H. HIMENO, K. MIURA and K. WATANABE. 1995. Nucleotide sequence and gene organization of the starfish *Asterina pectinifera* mitochondrial genome. Genetics **140:**1047-1060.

BABA, K. 1951. General sketch of the development of a solenogastre, *Epimenia verrucosa* (Nierstrasz). J. Dept. Agric. Kyusyu Imp. Univ. **6:**21-40.

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.

BEAGLEY, C. T., R. OKIMOTO and D. R. WOLSTENHOLME. 1998. The mitochondrial genome of the sea anemone *Metridium senile* (Cnidaria): Introns, a paucity of tRNA genes, and a near-standard genetic code. Genetics **148:**1091-1108.

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. Nucl. Acids Res **27:**1767-1780.

BOORE, J. L. 2000. The duplication/random loss model for gene rearrangement exemplified by mitochondrial genomes of deuterostome animals, Pp. 133-147, in D. SANKOFF and J. NADEAU, eds. Comparative Genomics, Computational Biology Series vol 1, Kluwer Academic Publishers, Dordrecht, Netherlands.

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

BOORE, J. L. and W. M. BROWN. 1995. Complete DNA sequence of the mitochondrial genome of the annelid worm, *Lumbricus terrestris*. Genetics **141:**305-319.

BOORE, J. L., and W. M. BROWN. 1998. Big trees from little genomes: Mitochondrial gene order as a phylogenetic tool. Curr. Opin. Genet. Dev. **8:**668-674.

BOORE, J. L. and W. M. BROWN. 2000. 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. Mol. Biol. Evol. **17(1):**87-106.

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 from the cephalochordate *Branchiostoma floridae* ("amphioxus"). Mol. Biol. Evol. **16(3):**410-418.

BOORE, J. L., D. LAVROV, and W. M. BROWN. 1998. Gene translocation links insects and crustaceans. Nature **393:**667-668.

BREMER, K. 1994. Branch support and tree stability. Cladistics **10:**295-304.

CANTATORE, P., M. ROBERTI, G. RAINALDI, M. N. GADALETA and C. SACCONE. 1989. The complete nucleotide sequence, gene order and genetic code of the mitochondrial genome of *Paracentrotus lividus*. J. Biol. Chem. **264:**10965-10975.

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.

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.

COLLINS, T. M., K. FRAZER, A. R. PALMER, G. J. VERMEIJ and W. M. BROWN. 1996. Evolutionary history of northern hemisphere *Nucella* (Gastropoda Muricidae): molecular, morphological, ecological, and paleontological evidence. Evolution. **50(6)**:2287-2304.

CREASE, T. J. 1999. The complete sequence of the mitochondrial genome of *Daphnia pulex*. Gene **233(1-2):**89-99.

DELLE CHIAJE, S. 1823. Memorie sulla storia e notomina degli animali senza vertebre del Regno di Napoli. Vol. I. Fratelli Fernandes, Naples. 184 pp.

DOWLING, T. E., C. MORITZ, J. D. PALMER and L. H. RIESENBERG. 1996. Nucleic acids III: Analysis of fragments and restriction sites. Pp. 249-320, in D. M. HILLIS, C. MORITZ and B. K. MABLE, eds. Molecular Systematics. Sinauer Associates, Sunderland, Massachusetts, USA.

DOWTON, M. 1999. Relationships among the cyclostome braconid (Hymenoptera: Braconidae) subfamilies inferred from a mitochondrial tRNA gene rearrangement. Mol. Phylogenet. Evol. **11:**283-287.

ERIKSSON, T. 1998. AutoDecay ver. 4.0 (program distributed by the author). Bergius Foundation, Royal Swedish Academy of Sciences, Stockholm.

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

FIELD, K. G., G. J. OLSEN, D. J. LANE, S. J. GIOVANNONI, M. T. GHISELIN, E. C. RAFF, N. R. PACE and R. A. RAFF. 1988. Molecular phylogeny of the animal kingdom. Science **239:**748-753.

FLOOK, P., C. H. F. ROWELL and G. GELLISSEN. 1995. The sequence, organization, and evolution of the *Locusta migratoria* mitochondrial genome. J. Mol. Evol. **41:**928-941.

FLORKIN, M. 1976. Biochemical evidence for the phylogenetic relationships of the Sipuncula. Pp. 95-108, in M. E. RICE and M. TODOROVIC, eds. Proceedings of the International Symposium on the Biology of the Sipuncula and Echiura. Vol. II. Naucno Delo Press, Belgrad.

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.

GEROULD, J. H. 1907. The development of *Phascolosoma*. Studies on the embryology of the Sipunculidae. I. Zool. Jahrb., Anat. **23:**77-162.

GHISELIN, M. T. 1988. The origin of molluscs in the light of molecular evidence. Pp. 66-95, in P. HARVEY and L. PARTRIDGE, eds. Oxford Surveys in Evolutionary Biology. Oxford University Press, Oxford.

GIRIBET, G., D. L. DISTEL, M. POLZ, W. STERRER and W. C. WHEELER. 2000. Triploblastic relationships with emphasis on the acoelomates and the position of Gnathostomulida, Cycliophora, Plathelminthes, and Chaetognatha: A combined approach of 18S rDNA sequences and morphology. Syst. Biol. **49:**539-562.

HEATH, H. 1899. The development of *Ischnochiton*. Zool. Jahrb., Anat. **12:**567-656.

HENRY, R. P. 1987. Invertebrate red blood cell carbonic anhydrase. J. Exp. Zool. **242:**113-116.

HOFFMANN, R. J., J. L. BOORE and W. M. BROWN. 1992. A novel mitochondrial genome organization for the blue mussel, *Mytilus edulis*. Genetics **131:**397-412.

Jacobs, D. K., C. G.Wray, C. J. Wedeen, R. Kostriken, R. DeSalle, J. L. Staton, R. D. Gates and D. R. Lindberg. 2000. Molluscan engrailed expression, serial organization, and shell evolution. Evol. Dev. **2:**340-347.

JANKE, A. and U. ARNASON. 1997. The complete mitochondrial genome of *Alligator mississippiensis* and the separation between recent Archosauria (birds and crocodiles). Mol. Biol. Evol. **14(12):**1266-1272.

KEDDIE, E. M., T. HIGAZI, and T. R. UNNASCH. 1998. The mitochondrial genome of *Onchocerca volvulus*: sequence, structure and phylogenetic analysis. Mol. Biochem. Parasitol. **95(1):**111-127.

KISHINO, H. AND M. HASEGAWA. 1989. Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data and the branching order in Hominoidea. J. Mol. Evol. **29:**170-179.

KURABAYASHI, A., and R. UESHIMA, 2000 Complete sequence of the mitochondrial DNA of the primitive opisthobranch *Pupa strigosa*: Systematic implications of the genome organization. Mol. Biol. Evol. **17(2):** 266-277.

LAMARCK, J. A. 1816. Histoire naturelle des animaux sans vertèbres. Vol. 3: Radiaires, vers, crustacés, insectes. Verdière, Paris. 586 pp.

LAVROV, D., J. L. BOORE, and W. M. BROWN. 2000. The complete mitochondrial DNA sequence of the horseshoe crab *Limulus polyphemus*. Mol. Biol. Evol. **17(5):**813-824.

LE, T. H., D. BLAIR, T. AGATSUMA, P. F. HUMAIR, N. J. CAMPBELL, M. IWAGAMI, D. T. LITTLEWOOD, B. PEACOCK, D. A. JOHNSTON, J. BARTLEY, D. ROLLINSON, E. A. HERNIOU, D. S. ZARLENGA and D. P. MCMANUS. 2000. Phylogenies inferred from mitochondrial gene orders-a cautionary tale from the parasitic flatworms. Mol. Biol. Evol. **17(7):**1123-1125.

MACKEY, L. Y., B. WINNEPENNINCKX, R. DE WACHTER, T. BACKELJAU, P. EMSCHERMANN and J. R. GAREY. 1996. 18S rRNA suggests that Entoprocta are protostomes, unrelated to Ectoprocta. J. Mol. Evol. **42:**552-559.

MADDISON, W. P. and D. R. MADDISON. 1992. Analysis of Phylogeny and Character Evolution, Sinauer, Sunderland, Massachusetts.

MALEY, L. E., and C. R. MARSHALL. 1998. Evolution - The coming of age of molecular systematics. Science **279:**505-506.

NOGUCHI, Y., K. ENDO, F. TAJIMA and R. Ueshima. 2000. The mitochondrial genome of the brachiopod *Laqueus rubellus*. Genetics **155(1):**245-259.

OJALA, D., J. MONTOYA and G. ATTARDI. 1981. tRNA punctuation model of RNA processing in human mitochondria. Nature **290:**470-474.

OKIMOTO, R., H. M. CHAMBERLIN, J. L. MACFARLANE and D. R. WOLSTENHOLME. 1991. Repeated sequence sets in mitochondrial DNA molecules of root knot nematodes (*Meloidogyne*): nucleotide sequences, genome location and potential for host race identification. Nucl. Acids Res. **19:**1619-1626.

OKIMOTO, R., J. L. MACFARLANE, D. O. CLARY and D. R. WOLSTENHOLME. 1992. The mitochondrial genomes of two nematodes, *Caenorhabditis elegans* and *Ascaris suum*. Genetics **130(3):**471-498.

Palumbi, S. R. 1996. Nucleic acids II: The polymerase chain reaction. Pp. 205-247, in D. M. HILLIS, C. MORITZ and B. K. MABLE, eds. Molecular Systematics. Sinauer Associates, Sunderland, Massachusetts, USA.

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.

QUATREFAGES, M. A. de 1847. Études sur les types inférieurs de l'embranchement des Annelés. Mémoire sur l'échiure de Gaertner (Echiurus Gaertnerii NOB.). Annales des Sciences Naturelles, ser. 3, Zoologie (Paris) **7:**307-343.

REGIER, J. C., and J. W. SHULTZ. 1998. Molecular phylogeny of arthropods and the significance of the Cambrian 'explosion' for molecular systematics. Am. Zool. **38:**918-928.

RICE, M. E. 1975. Sipuncula. Pp. 67-127, in A. GIESE and J. PEARSE, eds. Reproduction of Marine Invertebrates. Vol. 2. Entoprocts and lesser coelomates. Academic Press, New York.

RICE, M. E. 1985. Sipuncula: Developmental evidence for phylogenetic inference. Pp. 274-296, in S. CONWAY MORRIS, J. D. GEORGE, R. GIBSON and H. M. PLATT, eds. The Origins and Relationships of Lower Invertebrates. Oxford University Press, Oxford.

SACCONE, C., C. DEGIORGI, C. GISSI, G. PESOLE and A. REYES. 1999. Evolutionary genomics in Metazoa: the mitochondrial DNA as a model system. Gene **238:**195-209.

SAITOU, N. and M. NEI. 1987. The neighbor-joining method: A new method for reconstructing phylogenetic trees. Mol. Biol. Evol. **4:**1406-425.

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(6):**692-702.

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

SCHELTEMA, A. H. 1996. Phylogenetic position of the Sipuncula, Mollusca and the progenetic Aplacophora. Pp. 53-58, in J. D. TAYLOR, ed. Origin and Evolutionary Radiation of the Mollusca. Oxford University Press, Oxford.

SEDGWICK, A. 1898. Sipunculoidea (Gephrea, Achaeta). Pp. 534-539, in: A Student's Textbook of Zoology. Swan Sonnenschein, London.

SEKITO, T., K. OKAMOTO, H. KITANO and K. YOSHIDA. 1995. The complete mitochondrial DNA sequence of *Hansenula wingei* reveals new characteristics of yeast mitochondria. Curr. Genet. **28:**39-53.

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.

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(8)**:867-874.

STRIMMER, K. and A. VON HAESELER. 1996. Quartet puzzling: A quartet maximum likelihood method for reconstructing tree topologies. Mol. Biol. Evol. **13:**964-969.

STECHMANN, A. and M. SCHLEGEL. 1999. Analysis of the complete mitochondrial DNA sequence of the brachiopod *Terebratulina retusa* places Brachiopoda within the protostomes. Proc. R. Soc. Lond., B, Biol. Sci. **260:**2043-2053.

SWOFFORD, D. L. 2000. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. Sinauer Associates, Sunderland, Massachusetts.

TEMPLETON, A. R. 1983. Phylogenetic inference from restriction endonuclease cleavage site maps with particular reference to humans and apes. Evolution **37**:221-244.

VAN DONGEN, C. A. M. and W. L. M. GEILENKIRCHEN. 1974. The development of *Dentalium* with special reference to the significance of the polar lobe. I, II and III. Division chronology and development of the cell pattern in *Dentalium dentale* (Scaphopoda). Proc. K. Nederl. Akad. Wetens., ser. C, **77:**57-100.

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.

VERDONK, N. H. and J. A. M. VAN DEN BIGGELAR. 1983. Early development and the formation of the germ layers. Pp. 91-122, in N. H. VERDONK, J. A. M. VAN DEN BIGGELAR, and A. S. TOMPA, eds. The Mollusca. Vol. 3 Development. Academic Press, New York.

WATKINS, R. F. and A. T. BECKENBACH. 1999. Partial sequence of a sponge mitochondrial genome reveals sequence similarity to Cnidaria in cytochrome oxidase subunit II and the large ribosomal RNA subunit. J. Mol. Evol. **48:**542-554.

WINNEPENNINCKX, B., T. BACKELJAU and R. DE WACHTER. 1995. Phylogeny of protostome worms derived from 18S rRNA sequences. Mol. Biol. Evol. **12:**641-649.

YAMAZAKI, N., R. UESHIMA, J. TERRETT, S-I. YOKOBORI, M. KAIFU, R. SEGAWA, T. KOBAYASHI, K.-I. NUMACHI, T. UEDA, K. NICHIKAWA, K. WATANABE and R. THOMAS. 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.-I., U. TAKUYA, G. FELDMAIER-FUCHS, S. PÄÄBO, R. UESHIMA, A. KONDOW, K. NISHIKAWA, and K. WATANABE. 1999. Complete DNA sequence of the mitochondrial genome of the ascidian *Halocynthia roretzi* (Chordata, Urochordata). Genetics **153:**1851-1862.

YOKOGAWA, T., Y. -I. WATANABE, Y. KUMAZAWA, T. UEDA, I. HIRAO, K. -I. MIURA and K. WATANABE. 1991. A novel cloverleaf structure found in mammalian mitochondrial $tRNA^{ser}$(UCN). Nucl. Acids Res. **19(22):**6101-6105.

Table 1

Complete mitchondrial gene arrangements for the taxa considered in this phylogenetic analysis, with those matching partial *Phascolopsis gouldii* arrangements shown in boldface.

Taxon (classification)	Gene arrangement	Reference
Alligator[a, b] (Chordata: Vertebrata)	cox1[c], –S2[d], D, cox2, K, atp8, atp6, cox3, G, nad3, R, nad4L, nad4, S1, H, L1, nad5, –nad6, –E, cob, T, –P, F, **rrnS, V, rrnL**, L2, nad1, I, –Q, M, nad2, W, –A, –N, –C, –Y	Janke and Arnason 1997
Branchiostoma (Chordata: Cephalochordata)	cox1, –S2, D, cox2, K, atp8, atp6, cox3, nad3, R, nad4L, nad4, H, S1, L1, nad5, G, –nad6, –E, cob, T, –P, rrnS, F, **V, rrnL**, L2, nad1, I, M, –Q, nad2, –N, W, –A, –C, –Y	Boore, Daehler, and Brown 1999
Balanoglossus (Hemichordata)	cox1, –S2, –D, cox2, K, atp8, atp6, cox3, G, nad3, R, nad4L, nad4, H, S1, nad5, cob, E, T, –P, –nad6, F, **rrnS, V, rrnL**, L1, L2, nad1, –Q, I, M, nad2, –N, W, –A, –C, –Y	Castresana et al. 1998
Asterina (Echinodermata: Asteroidea)	cox1, R, nad4L, cox2, K, atp8, atp6, cox3, –S2, nad3, nad4, H, S1, nad5, –nad6, cob, F, rrnS, E, T, –rrnL, –nad2, –I, –nad1, –L2, **–G, –Y**, D, –M, V, –C, –W, A, –L1, –N, Q, –P	Asakawa et al. 1995
Paracentrotus (Echinodermata: Echinoidea)	cox1, R, nad4L, cox2, K, atp8, atp6, cox3, –S2, nad3, nad4, H, S1, nad5, –nad6, cob, F, rrnS, E, T, P, –Q, N, L1, –A, W, C, –V, M, –D, **Y, G**, L2, nad1, I, nad2, rrnL	Cantatore et al. 1989
Locusta (Arthropoda: Hexapoda)	cox1, L2, **cox2, D**, K, atp8, atp6, cox3, G, nad3, A, R, N, S1, E, –F, –nad5, –H, –nad4, –nad4L, T, –P, **nad6, cob**, S2, –nad1, –L1, **–rrnL, –V, –rrnS**, I, –Q, M, nad2, W, –C, –Y	Flook, Rowell, and Gellissen 1995

Drosophila (Hexapoda), Daphnia (Crustacea)	cox1, L2, cox2, K, **D, atp8**, atp6, cox3, G, nad3, A, R, N, S1, E, –F, –nad5, –H, –nad4, –nad4L, T, –P, **nad6, cob**, S2, –nad1, –L1, **–rrnL, –V, –rrnS**, I, –Q, M, nad2, W, –C, –Y	Clary and Wolstenholme 1985; Crease 1999
Artemia (Arthropoda: Crustacea)	cox1, L2, cox2, K, **D, atp8**, atp6, cox3, G, nad3, A, R, N, S1, E, –F, –nad5, –H, –nad4, –nad4L, T, –P, **nad6, cob,** S2, –nad1, –L1, **–rrnL, –V, –rrnS**, M, nad2, W, –I, –Q, –C, –Y	Valverde et al. 1994
Limulus, Ixodes (Arthropoda: Chelicerata)	cox1, cox2, K, **D, atp8**, atp6, cox3, G, nad3, A, R, N, S1, E, –F, –nad5, –H, –nad4, –nad4L, T, –P, **nad6, cob**, S2, –nad1, –L2, –L1, **–rrnL, –V, –rrnS**, I, –Q, M, nad2, W, –C, –Y	Staton, Daehler and Brown 1997; Black and Rohrdanz 1998
Katharina (Mollusca: Polyplacophora)	cox1, D, cox2, atp8, atp6, –F, –nad5, –H, –nad4, –nad4L, T, –S2, **–cob, –nad6**, P, –nad1, –L2, –L1, **–rrnL, –V, –rrnS, –M, –C**, –Y, –W, –Q, –G, –E, cox3, K, A, R, N, I, nad3, S1, nad2	Boore and Brown 1994
Loligo (Mollusca: Cephalopoda)	cox1, –C, –Y, –E, **N, cox2**, –M, R, –F, –nad5, –nad4, –nad4L, T, –L2, –G, A, **D, atp8**, atp6, –H, –L1, cox3, nad3, –S2, **–cob, –nad6**, –P, –nad1, –Q, I, **–rrnL, –V, –rrnS**, –W, K, S1, nad2	Sasuga et al. 1999; Genbank AB009838
Terebratulina (Brachiopoda)	cox1, **cox2, D, atp8**, atp6, Y, **C, M, rrnS, V, rrnL**, L1, A, L2, nad1, nad6, P, cob, K, N, S2, nad4L, nad4, Q, W, H, nad5, F, E, **G, cox3**, T, R, I, nad3, S1, nad2	Stechmann and Schlegel 1999
Lumbricus (Annelida: Oligochaeta)	**cox1, N, cox2, D, atp8, Y, G, cox3, Q, nad6, cob**, W, atp6, R, H, nad5, F, E, P, T, nad4L, nad4, **C, M, rrnS, V, rrnL**, L1, A, S2, L2, nad1, I, K, nad3, S1, nad2	Boore and Brown 1995
Platynereis (Annelida: Polychaeta)	**cox1, N, cox2**, G, Y, atp8, M, D, **cox3, Q, nad6, cob**, W, atp6, R, H, nad5, F, E, P, T, nad4L, nad4, **rrnS, V, rrnL**, L1, S2, A, L2, nad1, I, K, nad3, S1, nad2, C	Boore and Brown 2000; Genbank AF178678

[a] Alligator mtDNA differs from the gene arrangement typical of vertebrates (see Boore 1999) by only the position of *trnH*, which is not considered for this analysis since it is not in the sequenced portion of *P. gouldii* mtDNA.

[b] Only the genus is named here for brevity. Full binomens appear in figures 4 and 5.

[c] All of these genomes are circular but have been graphically linearized at the arbitrarily chosen *cox1* gene.

[d] Transfer RNA genes are abbreviated by the one letter code for the corresponding amino acid, with the two each for leucine and serine being differentiated by numerals. A minus sign indicates opposite (i.e. right-to-left as shown) transcriptional orientation.

Table 2

Codon usage for the protein-encoding genes in the sequenced portion of *Phascolopsis gouldii* mtDNA (1591 codons) compared with that of the homologous portions of *Lumbricus terrestris* mtDNA.

		P. gouldii					*L. terrestris*				
Amino acid	Codon family	All	Third codon position				All	Third codon position			
			T	C	A	G		T	C	A	G
Phe (F)	TTY	135	93	42	----	----	123	74	49	----	----
Leu (L)	TTR	80	----	---	80	0	61	----	----	56	5
Leu (L)	CTN	163	74	34	53	2	150	39	32	72	7
(Total Leu)-------		243	---	---	---	---	211	---	---	---	---
Ile (I)	ATY	123	85	38	---	---	136	93	43	----	----
Met (M)	ATR	66	---	---	57	9	70	----	----	55	15
Val (V)	GTN	83	26	16	38	3	101	19	15	45	22
Ser (S)	TCN	94	46	25	23	0	91	26	30	33	2
Ser (S)	AGN	31	4	5	22	0	39	9	6	19	5
(Total Ser)--------		125	---	---	---	---	130	---	---	---	---
Pro (P)	CCN	90	37	15	36	2	85	30	24	25	6
Thr (T)	ACN	90	33	20	36	1	106	31	30	42	3
Ala (A)	GCN	124	49	29	42	4	121	41	37	35	8
Tyr (Y)	TAY	56	36	20	---	---	56	31	25	----	----
His (H)	CAY	64	35	29	---	---	50	18	32	----	----
Gln (Q)	CAR	33	---	---	31	2	27	----	----	24	3

Asn (N)	AAY	61	38	23	---	---	57	26	31	---	---
Lys (K)	AAR	30	---	---	28	2	26	----	----	19	7
Asp (D)	GAY	32	18	14	---	---	39	21	18	----	----
Glu (E)	GAR	28	---	---	25	3	32	----	----	21	11
Cys (C)	TGY	14	8	6	---	---	10	2	8	----	----
Trp (W)	TGR	56	---	---	51	5	56	----	----	44	12
Arg (R)	CGN	36	13	5	18	0	34	9	4	16	5
Gly (G)	GGN	102	22	22	43	15	102	21	22	29	30
Total ----------		1591	617	343	583	48	1572	490	406	535	141

Figure legends

FIG 1.—Gene map of the sequenced portion of the mtDNA of the sipunculid *Phascolopsis gouldii* compared with the complete map of the oligochaete annelid *Lumbricus terrestris* (Boore and Brown 1995), which has been graphically linearized at *cox1*. All genes are transcribed left to right as depicted. Transfer RNA genes are designated by a single letter for the corresponding amino acid. There are two tRNAs for each of leucine and serine that are differentiated by numerals: L1 and L2 recognize the codons CUN and UUR, respectively, and S1 and S2 recognize the codons AGN and UCN, respectively. All other genes are designated by standard annotation and are not to scale. Lines connect homologous genes or blocks of genes for each mtDNA.

FIG 2.—Abbreviated representation of the 7470 nts determined of the mtDNA of the sipunculid *Phascolopsis gouldii*. Center portions of each protein- or rRNA-encoding gene have been replaced with a numeral indicating the number of omitted nucleotides. Nucleotides forming stop codons, partial or complete (see text) are marked with a caret (^).

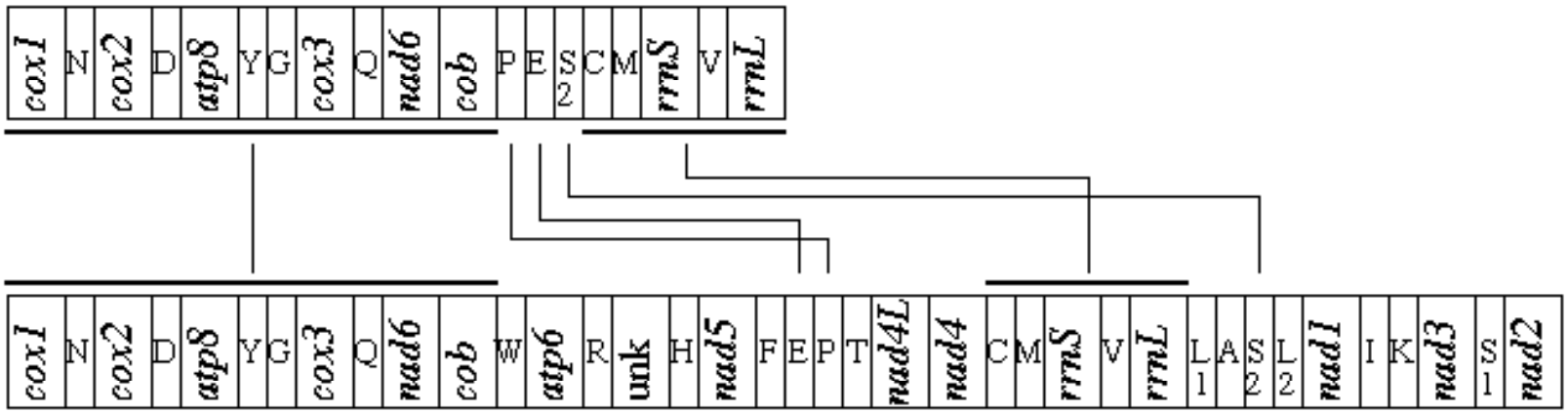
FIG 3.—DNA sequences of the 11 tRNA genes proposed for *Phascolopsis gouldii* mtDNA folded into the standard cloverleaf structure. Two serine-specifying tRNAs are normally encoded by animal mtDNAs and it is the one predicted to recognize the codon UCN that is found here; this is designated S2 for consistency with our earlier work, although there is no universal convention for this. Potential base pairing for an unusual DHU arm of tRNA(S2) that would have only one nucleotide after the amino-acyl

acceptor arm and two nucleotides before the anticodon arm (see text) is indicated by connecting lines. Structural elements are identified for tRNA(V).

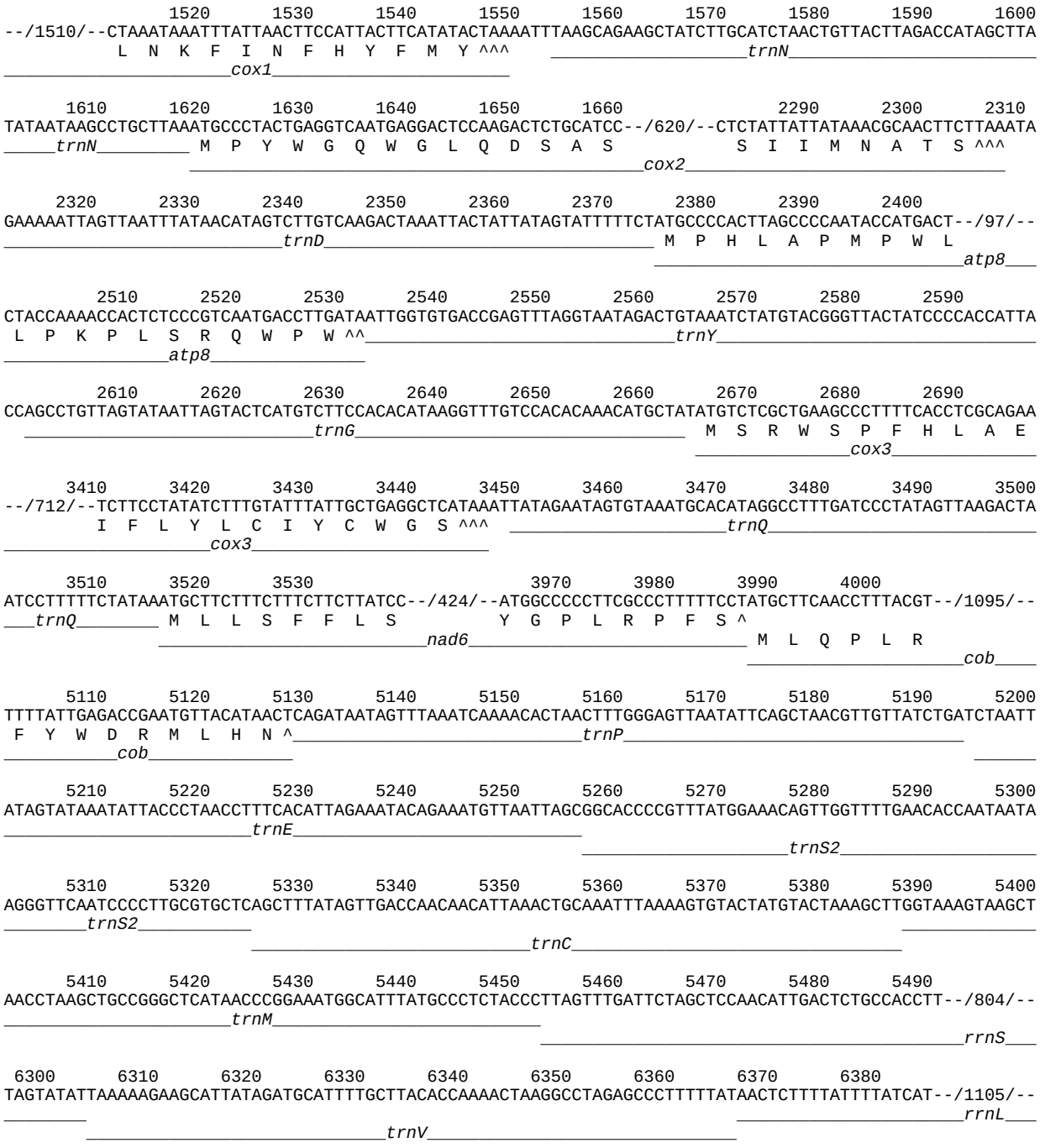
FIG 4.—Phylogenetic tree based on comparisons of amino acid sequences for Cob, Cox1, Cox2, and Cox3 for 17 taxa. The tree was rooted with the deuterostome sequences and branch lengths were assigned using Tree-Puzzle (Strimmer and von Haeseler 1996). Numbers above branches refer to percent of puzzle quartets supporting this branch, followed by bootstrap support using distance criteria, then parsimony criteria (a dash denotes < 50% bootstrap support), then the parsimony branch support value (Bremer 1994). The two sets of relationships where the methods give differing results (monophyly vs. paraphyly of the crustaceans *D. pulex* and *A. franciscana* and the placement of *T. retusa* as sister to *L. bleekeri* vs. as sister to both mollusks) were subsequently collapsed to polytomies, so this is a strict consensus tree of the three phylogenetic methods employed (maximum likelihood, parsimony, neighbor joining; see text).

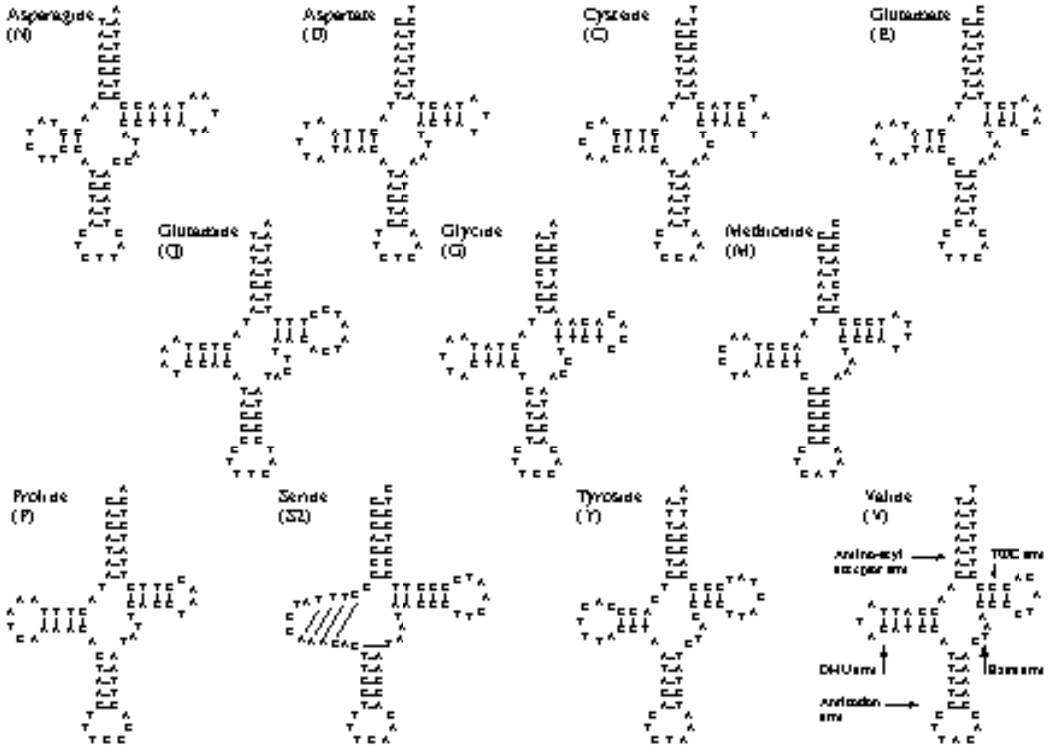
FIG 5.—Phylogenetic tree based on gene arrangement comparisons using an adjacency matrix. This is the strict consensus of 30 equally parsimonious trees. The unusual placement of *A. mississippiensis* is poorly supported (see text).

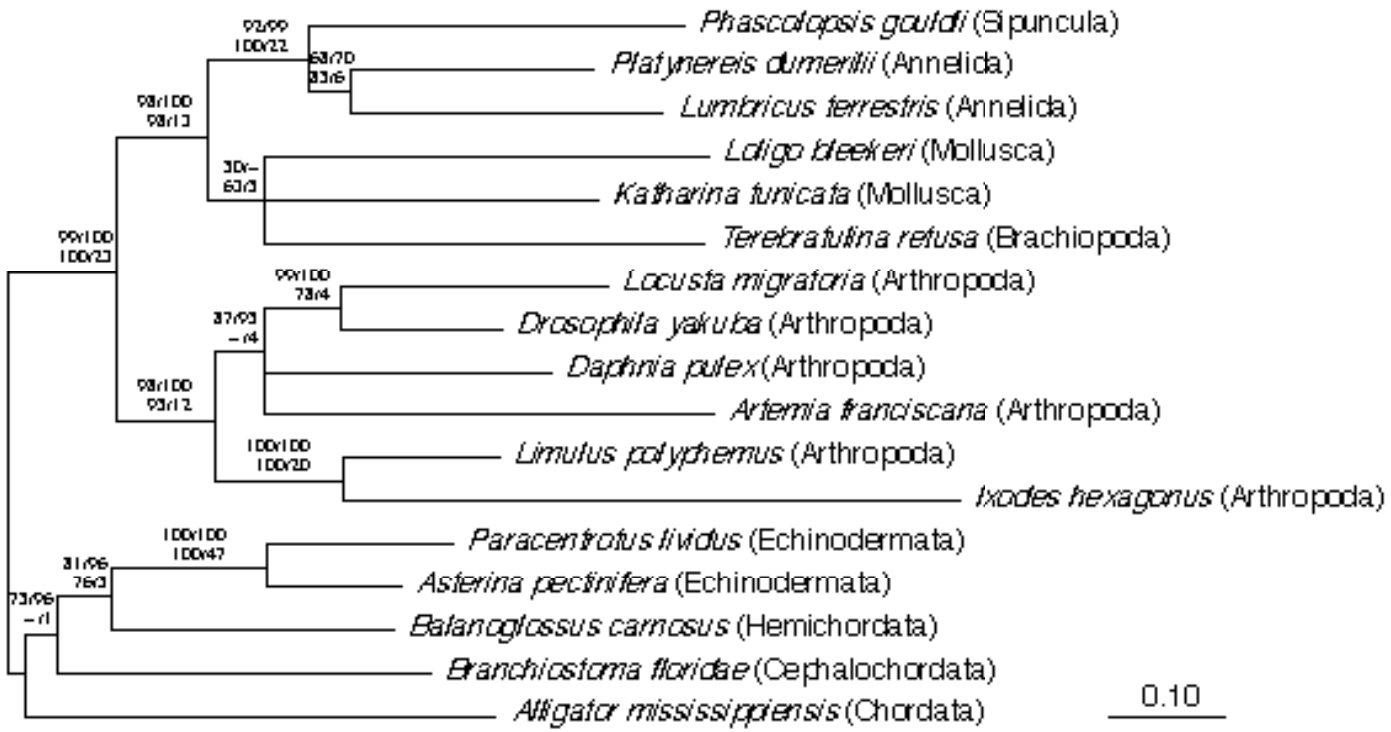
Phascolopsis gouldii (Sipuncula)

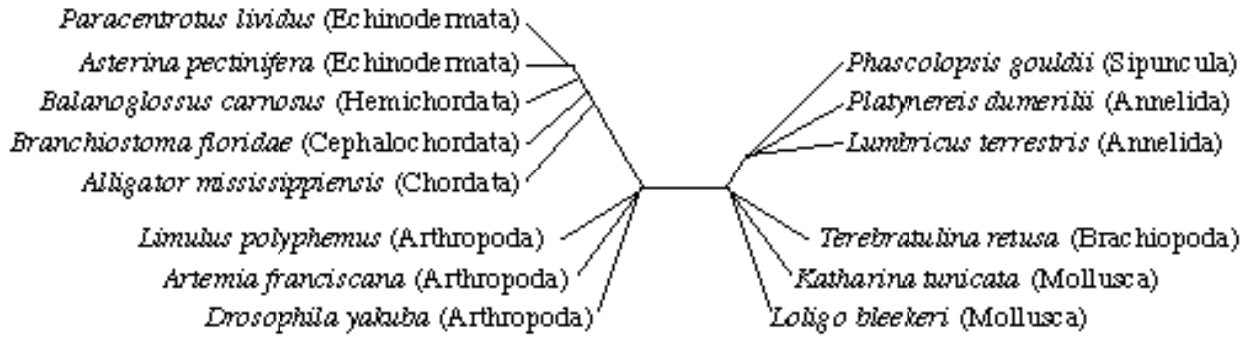


Lumbricus terrestris (Annelida, Oligochaeta)


1520 1530 1540 1550 1560 1570 1580 1590 1600
--/1510/--CTAAATAAATTTATTAACTTCCATTACTTCATATACTAAAATTTAAGCAGAAGCTATCTTGCATCTAACTGTTACTTAGACCATAGCTTA
L N K F I N F H Y F M Y ^^^
trnN
cox1
1610 1620 1630 1640 1650 1660 2290 2300 2310
TATAATAAGCCTGCTTAAATGCCCTACTGAGGTCAATGAGGACTCCAAGACTCTGCATCC--/620/--CTCTATTATTATAAACGCAACTTCTTAAATA
trnN M P Y W G Q W G L Q D S A S S I I M N A T S ^^^
cox2
2320 2330 2340 2350 2360 2370 2380 2390 2400
GAAAAATTAGTTAATTTATAACATAGTCTTGTCAAGACTAAATTACTATTATAGTATTTTTCTATGCCCCACTTAGCCCCAATACCATGACT--/97/--
trnD M P H L A P M P W L
atp8
2510 2520 2530 2540 2550 2560 2570 2580 2590
CTACCAAAACCACTCTCCCGTCAATGACCTTGATAATTGGTGTGACCGAGTTTAGGTAATAGACTGTAAATCTATGTACGGGTTACTATCCCCACCATTA
L P K P L S R Q W P W ^^
trnY
atp8
2610 2620 2630 2640 2650 2660 2670 2680 2690
CCAGCCTGTTAGTATAATTAGTACTCATGTCTTCCACACATAAGGTTTGTCCACACAAACATGCTATATGTCTCGCTGAAGCCCTTTTCACCTCGCAGAA
trnG M S R W S P F H L A E
cox3
3410 3420 3430 3440 3450 3460 3470 3480 3490 3500
--/712/--TCTTCCTATATCTTTGTATTTATTGCTGAGGCTCATAAATTATAGAATAGTGTAAATGCACATAGGCCTTTGATCCCTATAGTTAAGACTA
I F L Y L C I Y C W G S ^^^
trnQ
cox3
3510 3520 3530 3970 3980 3990 4000
ATCCTTTTTCTATAAATGCTTCTTTCTTTCTTCTTATCC--/424/--ATGGCCCCCTTCGCCCTTTTTCCTATGCTTCAACCTTTACGT--/1095/--
trnQ M L L S F F L S Y G P L R P F S ^
nad6 M L Q P L R
cob
5110 5120 5130 5140 5150 5160 5170 5180 5190 5200
TTTTATTGAGACCGAATGTTACATAACTCAGATAATAGTTTAAATCAAAACACTAACTTTGGGAGTTAATATTCAGCTAACGTTGTTATCTGATCTAATT
F Y W D R M L H N ^
trnP
cob
5210 5220 5230 5240 5250 5260 5270 5280 5290 5300
ATAGTATAAATATTACCCTAACCTTTCACATTAGAAATACAGAAATGTTAATTAGCGGCACCCCGTTTATGGAAACAGTTGGTTTTGAACACCAATAATA
trnE
trnS2
5310 5320 5330 5340 5350 5360 5370 5380 5390 5400
AGGGTTCAATCCCCTTGCGTGCTCAGCTTTATAGTTGACCAACAACATTAAACTGCAAATTTAAAAGTGTACTATGTACTAAAGCTTGGTAAAGTAAGCT
trnS2
trnC
5410 5420 5430 5440 5450 5460 5470 5480 5490
AACCTAAGCTGCCGGGCTCATAACCCGGAAATGGCATTTATGCCCTCTACCCTTAGTTTGATTCTAGCTCCAACATTGACTCTGCCACCTT--/804/--
trnM
rrnS
6300 6310 6320 6330 6340 6350 6360 6370 6380
TAGTATATTAAAAAGAAGCATTATAGATGCATTTTGCTTACACCAAAACTAAGGCCTAGAGCCCTTTTTATAACTCTTTTATTTTATCAT--/1105/--
rrnL
trnV


Asparagine
(N)
Aspartate
(D)
Cysteine
(C)
Glutamate
(E)
Glutamine
(Q)
Glycine
(G)
Methionine
(M)
Proline
(P)
Serine
(S2)
Tyrosine
(Y)
Valine
(V)
Amino-acyl acceptor arm
TΨC arm
DHU arm
Base arm
Anticodon arm


92/99
100/22
68/70
33/6
98/100
98/13
30/–
60/3
99/100
100/23
99/100
78/4
87/93
–/4
98/100
93/12
100/100
100/20
100/100
100/47
81/96
76/3
73/96
–/1
Phascolopsis gouldii (Sipuncula)
Platynereis dumerilii (Annelida)
Lumbricus terrestris (Annelida)
Loligo bleekeri (Mollusca)
Katharina tunicata (Mollusca)
Terebratulina retusa (Brachiopoda)
Locusta migratoria (Arthropoda)
Drosophila yakuba (Arthropoda)
Daphnia pulex (Arthropoda)
Artemia franciscana (Arthropoda)
Limulus polyphemus (Arthropoda)
Ixodes hexagonus (Arthropoda)
Paracentrotus lividus (Echinodermata)
Asterina pectinifera (Echinodermata)
Balanoglossus carnosus (Hemichordata)
Branchiostoma floridae (Cephalochordata)
Alligator mississippiensis (Chordata)
0.10


Paracentrotus lividus (Echinodermata)
Asterina pectinifera (Echinodermata)
Balanoglossus carnosus (Hemichordata)
Branchiostoma floridae (Cephalochordata)
Alligator mississippiensis (Chordata)
Limulus polyphemus (Arthropoda)
Artemia franciscana (Arthropoda)
Drosophila yakuba (Arthropoda)
Phascolopsis gouldii (Sipuncula)
Platynereis dumerilii (Annelida)
Lumbricus terrestris (Annelida)
Terebratulina retusa (Brachiopoda)
Katharina tunicata (Mollusca)
Loligo bleekeri (Mollusca)