Molecular Phylogenetics and Evolution 55 (2010) 1–17

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

# Nearly complete rRNA genes assembled from across the metazoan animals: Effects of more taxa, a structure-based alignment, and paired-sites evolutionary models on phylogeny reconstruction

Jon Mallatt [a,*], Catherine Waggoner Craig [a], Matthew J. Yoder [b]

[a] School of Biological Sciences, Washington State University, Pullman, WA 99164-4236, USA
[b] Department of Entomology, Texas A&M University, College Station, TX 77843, USA

ARTICLE INFO

*Article history:*
Received 9 December 2008
Revised 28 August 2009
Accepted 22 September 2009
Available online 26 September 2009

*Keywords:*
28S rRNA
Ribosomal rRNA genes
Metazoa
Ecdysozoa
Molecular phylogeny
Secondary structure
PHASE program

ABSTRACT

This study (1) uses nearly complete rRNA-gene sequences from across Metazoa (197 taxa) to reconstruct animal phylogeny; (2) presents a highly annotated, manual alignment of these sequences with special reference to rRNA features including paired sites (http://purl.oclc.org/NET/rRNA/Metazoan_alignment) and (3) tests, after eliminating as few disruptive, rogue sequences as possible, if a likelihood framework can recover the main metazoan clades. We found that systematic elimination of ~6% of the sequences, including the divergent or unstably placed sequences of cephalopods, arrowworm, symphylan and pauropod myriapods, and of myzostomid and nemertodermatid worms, led to a tree that supported Ecdysozoa, Lophotrochozoa, Protostomia, and Bilateria. Deuterostomia, however, was never recovered, because the rRNA of urochordates goes (nonsignificantly) near the base of the Bilateria. Counterintuitively, when we modeled the evolution of the paired sites, phylogenetic resolution was not increased over traditional tree-building models that assume all sites in rRNA evolve independently. The rRNA genes of non-bilaterians contain a higher % AT than do those of most bilaterians. The rRNA genes of Acoela and Myzostomida were found to be secondarily shortened, AT-enriched, and highly modified, throwing some doubt on the location of these worms at the base of Bilateria in the rRNA tree—especially myzostomids, which other evidence suggests are annelids instead. Other findings are marsupial-with-placental mammals, arrowworms in Ecdysozoa (well supported here but contradicted by morphology), and Placozoa as sister to Cnidaria. Finally, despite the difficulties, the rRNA-gene trees are in strong concordance with trees derived from multiple protein-coding genes in supporting the new animal phylogeny.



## 1. Introduction

For several decades, ribosomal RNA genes have been used to reconstruct the phylogeny of the multicellular animals (Field et al., 1988; Halanych, 2004; Hillis and Dixon, 1991; Paps et al., 2009). They remain useful today, although phylogenies are now being calculated from multiple protein-coding genes, microRNA sequences, and even from whole genomes (Dunn et al., 2008; Lartillot and Philippe, 2008; Philippe and Telford, 2006; Phillipe et al., 2009; Schierwater et al., 2009; Sempere et al., 2007). Most rRNA studies have used the small-subunit gene (SSU, 18S: Giribet et al., 2000), and many sequences of the longer, large-subunit gene (LSU, 28S plus 5.8S) are now available: enough to yield nearly complete LSU+SSU rRNA for hundreds of taxa spanning the Metazoa. These complete rRNA sequences have been analyzed piecemeal; that is, to reconstruct relations in subgroups of Metazoa such as in Deuterostomia (Mallatt and Winchell, 2007), in Lophotrochozoa (Passamaneck and Halanych, 2006), in Ecdysozoa (Mallatt and Giribet, 2006; von Reumont et al., 2009), in Acoelomorph worms (Telford et al., 2003; Wallberg et al., 2007), and in Cnidaria (Collins et al., 2006; Medina et al., 2001). Recently, Paps et al. (2009) broadened the analysis by using complete rRNA sequences from about 100 taxa, focusing on Lophotrochozoa but sampling most other metazoan phyla as well (also see their preliminary report: Baguñà et al., 2008). Here, we take the next logical step of doubling this sample to about 200 rRNA sequences from the full range of Metazoa.

Seeking still more improvement, we constructed the largest secondary structure-based alignment of the rRNA genes to date, and used the meta data on paired and unpaired sites in a mixed evolutionary model for reconstructing phylogeny (Gillespie et al., 2005, 2006; Gowri-Shankar and Jow, 2006). Using paired sites should be an improvement, at least in theory, over the vast majority of studies that unrealistically assume all sites in the rRNA gene

* Corresponding author. Address: Box 644236, School of Biological Sciences, Washington State University, Pullman, WA 99164-4236, USA. Fax: +1 509 335 3184.
*E-mail address:* jmallatt@mail.wsu.edu (J. Mallatt).

1055-7903/$ - see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.ympev.2009.09.028

evolve independently (Dixon and Hillis, 1993; Kjer, 2004; Swofford et al., 1996; Wheeler and Honeycutt, 1988). Some groundbreaking studies have already used mixed models of rRNA evolution to reconstruct phylogeny, but these studies either used smaller clades of metazoans, or only the SSU or the LSU gene (Erpenbeck et al., 2007; Kjer, 2004; Kober and Nichols, 2007; Rix et al., 2008; Telford et al., 2005; Ware et al., 2008). Here, we use a mixed model for the entire rRNA-gene family across the metazoans.

## 2. Materials and methods

For a complete record of the rRNA sequences used here, and their alignment, the reader is referred to our master NEXUS file, which is built on structural considerations of paired sites and is presented online at http://purl.oclc.org/NET/rRNA/Metazoan_alignment. The alignment is also available in Stockholm format on request. Table 1 is an index, with definitions, of the names of some metazoan clades that are considered in this study.

### 2.1. Sequences

All the LSU and SSU sequences were obtained from GenBank. Only nearly complete sequences were chosen; that is, those over 90% complete, with several exceptions down to 80% (e.g., from frog *Rana*, bug *Poppea*, krill *Meganyctiphanes*, and wasp *Labena*). The 197 taxa used, whose names and GenBank numbers are given on the above-mentioned website, comprise 33 chordates, 5 hemichordates, 6 echinoderms, 4 nemerteans, 16 annelids (including sipunculans and echiurans), 3 bryozoans, 1 entoproct, 1 phoronid, 4 brachiopods, 9 molluscs, 1 chaetognath, 1 cycliophoran, 1 myzostomid, 2 rotifers, 2 acanthocephalans, 5 platyhelminths, 2 priapulids, 10 chelicerate arthropods, 26 hexapod arthropods, 3 pycnogonid arthropods, 14 myriapod arthropods, 20 crustacean arthropods, 1 tardigrade, 2 onychophorans, 2 nematomorphs, 6 nematodes, 1 kinorhynch, 1 placozoan, 1 nemertodermatid, 1 acoelan, 8 cnidarians, 1 ctenophore, 2 poriferans, 1 choanoflagellate, and 1 fungus.

A few of the 197 sequences were never before included in a large-scale phylogenetic analysis: those of *Ornithorhynchus* (duck-billed platypus), *Monodelphis* (opossum), *Plumatella* (bryozoan), and *Heterorhabditis* (rhabditid nematode: Chilton et al., 2006). Also, the LSU gene of the lizard *Anolis*, which had only partly been sequenced (Mallatt and Winchell, 2007), was completed here by adding several hundred bases from the Anolis Genome Sequencing Project (http://www.ensembl.org/index.html).

**Table 1**
Some metazoan clades defined, for reference in this paper.

1. Acoelomorpha: two clades of small worms, Acoela and Nemertodermatida
2. Ambulacraria: Hemichordata and Echinodermata
3. Bilateria: the metazoan clades with primitively bilaterian bodies; all metazoans except Cnidaria, Ctenophora, Placozoa, and Porifera
4. Deuterostomia: a group of Bilateria phyla, in most of which the anus develops from the embryonic blastopore. Currently defined to include the Chordata, Hemichordata, Echinodermata, and the worm *Xenoturbella* (Bourlat et al., 2003)
5. Diploblast phyla: The Porifera, Cnidaria, Ctenophora, and Placozoa. Generally not considered to form a clade, but see Schierwater et al. (2009)
6. Ecdysozoa (Aguinaldo et al., 1997; Telford et al., 2008): a major group of Protostomia; animals that shed a cuticle: Arthropoda, Tardigrada, Onychophora, Nematoda, Nematomorpha (Schmidt-Rhaesa, 1996), Priapulida, Kinorhyncha, Loricifera
7. Lophotrochozoa (Halanych et al., 1995), or Spiralia (*sensu* Giribet, 2008): a major group of Protostomia; originally defined to include Mollusca, Annelida, Nemertea, Brachiopoda, Phoronida, Platyhelminthes, and Bryozoa, and now thought also to include Entoprocta, Rotifera, Acanthocephala, Gastrotricha, Gnathostomulida, and Micrognathozoa (Paps et al., 2009); most have spiral cleavage and some have trochophore larvae or lophophore feeding structures
8. Malacostraca: the subgroup of crustacean arthropods that includes lobsters, crabs, shrimps, isopods, and some lesser-known forms; united by such characters as a thorax with eight segments and an abdomen with seven segments, all of which have appendages (Ruppert et al., 2004)
9. Metazoa: all multicellular animals
10. Myriapoda: centipedes (Chilopoda), millipedes (Diplopoda), symphylans and pauropods: land arthropods that share a long trunk with many appendage-bearing segments, two tagmata, no median eye, ommatidia without crystalline cone (Ruppert et al., 2004)
11. Panarthropoda: Ecdysozoa with appendages and claws: Arthropoda, Tardigrada, and Onychophora
12. Pancrustacea: Hexapoda and the (paraphyletic) crustacea
13. Platyhelminthes: here, the catenulid and rhabditophoran flatworms, but not the Acoelomorph worms
14. Protostomia: the Ecdysozoa and Lophotrochozoa (and perhaps, the Chaetognatha)
15. Syndermata: Rotifera plus Acanthocephala
16. Urochordata: the tunicate chordates

### 2.2. Alignment

Our alignments were done manually and by visual inspection, with reference to the conserved-structural diagrams of SSU and LSU rRNA from the frog *Xenopus laevis*, yeast *Saccharomyces cerevisae*, and bee *Apis mellifera* (Gillespie et al., 2006; Gutell, 1994; Gutell et al., 1993; Schnare et al., 1996). We know of no computerized algorithmic approach that can produce annotated alignments of this nature *de novo*, though such programs are on the horizon (Nawrocki et al., 2009; von Reumont et al., 2009; and http://www.zfmk.de/web/Forschung/Abteilungen/AG_Wgele/Software/RNAsalsa/index.en.html).

In previous studies, we aligned rRNA genes by secondary structure but we did not proceed in a complete way (Mallatt and Sullivan, 1998; Mallatt et al., 2001, 2004; Mallatt and Winchell, 2002, 2007; Mallatt and Giribet, 2006; Winchell et al., 2002, 2004). That is, we delimited half-stems and loops but did not consider or match the paired sites of the bonding half-stems. In the present study, by contrast, we did identify the paired sites by matching the canonical base-pairs (C-G and A-T) and wobble pair (G-T), and by using the comparative method to identify compensatory base changes (see Gillespie et al., 2005; Kjer, 2004; Kjer et al., 2007). For especially difficult and length-variable sub-segments, such as the D2 and D8 divergent domains of 28S and the V2 and V4 variable regions of 18S (Gillespie et al., 2006; Hassouna et al., 1984), we took a different approach to the alignment. That is, we used the mfold program (as implemented at http://mobyle.pasteur.fr/) to estimate secondary structure based on free-energy minimization. mfold generates alternate configurations, so we selected the configuration most congruent with the *Xenopus* and *Apis* structural diagrams.

The fastest-evolving regions of the genes, where alignment was ambiguous, were bracketed out and excluded from the phylogenetic analyses. As a rule of thumb, we excluded sites where fewer than 70% of all taxa agree or align. This led to a final alignment with 4682 sites, of which 2612 were paired (from stems) and 2070 were unpaired (from loops or bulges). This alignment contained 442 nucleotides in the divergent domains of 28S genes, as well as 4240 nt in the 28S-core, 5.8S, and the 18S gene.

The alignment was constructed as a Microsoft Word file in NEXUS format, using a simplified version of the notation scheme of Gillespie et al. (2005). Then, this NEXUS file was saved in a text-only format that can be executed by the PAUP* and MrBayes programs for tree reconstruction. This is the aligned file on the website mentioned above.

### 2.3. Phylogenetic analyses

A goal of this study was to obtain a broad taxonomic alignment that includes even the most-divergent rRNA sequences. To make full use of this alignment, we avoided the common practice (Bourlat et al., 2008; Jenner and Littlewood, 2008; Paps et al., 2009) of excluding divergent sequences *a priori* from the phylogenetic analyses.

#### 2.3.1. Independent-sites models (Garli and RAxML)

Maximum-likelihood (ML) trees were calculated from the 197 sequences using Garli 0.95 and 0.96 (http://www.molecularevolution.org/mbl/software/garli/ and Zwickl, 2006), and RAxML 7.0.4 (http://icwww.epfl.ch/~stamatak/index-Dateien/Page433.htm and Stamatakis et al., 2008). Garli has good accuracy (Morrison, 2007) and RAxML computes fastest. The model of nucleotide substitution was specified as GTR+I+Γ4. The use of an 'I' parameter has been questioned (Mayrose et al., 2005; Stamatakis, 2008), but it was the only way to select a preferred model objectively with the AIC test in ModelTest (Posada and Buckley, 2004).

In Garli, starting trees were set either to 'random' or to the best parsimony tree calculated in PAUP*. The other parameters were set to default. To find the best ML tree, we ran the search 15 times with different starting seeds. Then, we ran a 100 ML-bootstrap-replicate analysis, also in Garli. For this bootstrap analysis, the starting tree was specified as stepwise and the number of searchreps per replicate was 8.

In RAxML, the rapid hill-climbing function was specified (−f d), with the GTR+I+Γ model (−m GTRMIXI) and four gamma categories (−c 4), and the number of initial rearrangements set to the default (−i 10). The best ML tree was obtained from 50 runs (−# 50), then 500 bootstrap replicates were executed (−# 500).

#### 2.3.2. Mixed models, paired and unpaired sites (PHASE)

The PHASE program (Versions 1.0 and 2.0: Gowri-Shankar and Jow, 2006; http://www.bioinf.man.uk/resources/phase/) was used to calculate trees with mixed models. As was done previously (http://purl.oclc.org/NET/rRNA/jRNA and Gillespie et al., 2005), Perl scripts were constructed that parsed the alignment, a stem-index file, and the pairing statements to generate input files formatted for PHASE (this scripting framework is available from M.J.Y.). Model 1, used for the unpaired sites, was GTR (=REV)+I+Γ4. Model 2, for the paired sites, was RNA7A+I+Γ4. Justification for using the RNA7A model, which is used more often than other models (Jow et al., 2002; Kober and Nichols, 2007; Savill et al., 2001; Ware et al., 2008), is given by Gillespie et al. (2005).

The PHASE commands were those in the sample file named REVdG6-7DdG6.mcmc in the PHASE control folder, except that 'Hyperpriors proposal priority = 1' was added to the {PERTURBATION} commands, and 'Branch lengths, prior = exponential (uniform (0, 100))' was added to the {PERTURBATION_TREE} commands. In constructing the mcmc chain, sampling iterations were set at 8 million, sampling period was 1000, and burnin iterations were 2 million generations. This burnin value was selected from where test runs had shown stabilization to occur in the LnL and model parameters.

This PHASE analysis was repeated three times, each with a different random seed. The resulting 18,000 trees were combined into a majority-rule consensus tree from which posterior probabilities (PP) were taken.

PHASE also analyzed the alignment with a traditional, independent-sites model in which REV+I+Γ4 was applied to all the sites. That is, in this version all sites were treated as unpaired. This allowed comparison with the findings of the mixed model. Again, three analyses were run, then combined into a consensus tree with PP values.

Throughout the study, we accepted ML-bootstrap values above 60–70%, and PHASE-derived PP values $\geqslant$95%, as significant.

#### 2.3.3. Identifying unstable taxa and clades

Divergent and unstable sequences can disrupt tree topologies (Thorley and Wilkinson, 1999; Wilkinson, 2006). To tell if this was happening here, we identified which sequences (or taxa, or "leafs") were unstable in position on our tree. The leaves-stability (LS) values were calculated with the Phyutility program (Smith and Dunn, 2008) from the 100 bootstrap-replicate trees that we had generated with Garli with the ordinary, independent-sites model. A mixed model could not be used because the LS index (Thorley and Wilkinson, 1999) was not designed for mixed models. The bootstrap trees were treated as unrooted. Because our calculated values of LS fit a beta distribution, we placed them into quantiles using R statistical software (R Development Core Team, 2008; Wessa, 2008); then, we rejected the taxa with low LS values, in the lower 0.10 quantile, as unstable at $P < 0.10$. We chose 0.10 because it was used by Dunn et al. (2008) in a comparable analysis.

#### 2.3.4. Hypothesis testing

We formally tested some literature-based hypotheses of metazoan relations that were not recovered in our trees. All methods available for hypothesis testing use independent-sites models, not mixed models.

First, Garli calculated the absolute-best ML tree, and then it calculated the best trees that were constrained to each alternate hypothesis. Next, these alternate trees were evaluated with the approximately unbiased (AU) test in Consel (Shi et al., 2005; Shimodaira, 2002; Shimodaira and Hasegawa, 2001).

## 3. Results

### 3.1. Erroneous and valid sequences

A few taxa came out in biologically unlikely positions in test trees, and their rRNA sequences did not resemble or BLAST with those of their relatives, so these sequences seem wrong and are suspected contaminants. These taxa, which were not included in our 197-taxon set, are: the bryozoan *Alcyonidium*, whose 28S gene (AY210453) BLASTs closest to urochordates; the squid *Histioteuthis*, whose 28S (AY145410) BLASTs with teleost fish; the aplacophoran mollusc *Helicoradomenia*, whose 18S (AY145377) BLASTs closest to polychaetes; and another aplacophoran *Chaetoderma*, whose 28S (AY145397) BLASTs close to vestimentiform worms (see Passamaneck and Halanych, 2006; Passamaneck et al., 2004). Also, one 18S sequence reported for the nemertodermatid, *Meara* sp. (AF051328: Ruiz-Trillo et al., 1999) seems to be a hybrid, because its 5' half differs from that of valid *Meara* sequences (AM747473 and AF119085) and BLASTs instead with copepod arthropods. Finally, an 18S sequence from the platypus *Ornithorhynchus* (AJ311679: Janke et al., 2002) BLASTs with a marsupial instead, and differs markedly from the platypus 18S that we assembled and used from the Duck-Billed Platypus Genome Sequencing project (Warren et al., 2008; http://www.ensembl.org/).

Sequences of two taxa that we formerly feared were pseudogenes (Mallatt and Giribet, 2006) are probably valid after all. First, the LSU and SSU sequences of the symphylan myriapod *Hanseniella* (AY210821–AY210823) not only match those of another symphylan, *Scutigerella*, which were obtained more recently by Gai et al. (2006), but they also can be mfolded into a reasonable secondary structure that contains the universal stems and loops. Second, the long 28S sequence of the millipede Sphaerotheriidae sp. (AY859605–AY859607) also folded rationally with mfold; and

some conserved, universally alignable, parts of its divergent domains were detected here for the first time.

### 3.2. Heterogeneity of base composition

A chi-square test of uniformity of the nucleotide frequencies across the rRNA genes of our 197 taxa, performed in PAUP*, revealed a strong heterogeneity (chi-square value = 2966.9, $P = 0.00000000$). This is shown in Supplementary material, S1, a table that also tells the sizes and the CG-versus-AT content of these genes. The compositional heterogeneity of the nucleotides violates an assumption of homogeneity that is made by all the tree-building methods used in this study (Swofford et al., 1996). In our previous studies, which used narrower ranges of taxa, compositional heterogeneity was present but not so problematic, because it was caused by just a few AT- and CG-rich sequences that were scattered across the tree (e.g., Mallatt and Giribet, 2006; Mallatt and Winchell, 2007). There, the problem was easy to fix by removing these few sequences. Here, by contrast, all the non-bilaterian sequences are AT-rich compared to those of typical bilaterians (see the taxa marked by asterisks in Supplementary material, S1), introducing a more systematic bias. Thus, we could not remove the heterogeneous sequences here because that would eliminate the outgroup to Bilateria and thereby erase all evidence of basal–bilaterian relationships. As will be seen soon, a Bilateria clade still holds together well in our trees, but the compositional bias means that some of the AT-rich bilaterian sequences may attract artifactually toward the AT-rich non-bilaterians.

### 3.3. Variations in base composition and gene lengths

rRNA genes vary in length across taxa, mostly due to different lengths of the divergent domains of the 28S genes (Gillespie et al., 2005). The longer genes tend to have longer versions of the largest domains (compare columns 1 and 14 in S1). The divergent domains also tend to be more CG-rich than the rest of the gene (see column 14), so we hypothesized that the longer rRNA genes among the metazoans are more CG-rich. To test this hypothesis, we ran a linear-regression analysis on the length versus % CG for the whole gene family (column 1 versus column 10 in S1). As shown in Fig. 1, a positive correlation was indeed detected ($\rho = 0.51$, $P < 0.01$). Although this correlation was not too strong, the figure does show that all the shortest genes (at the far left) are CG-poor and AT-rich.

Supplementary material, S1 also shows that when a taxon's whole gene family is CG-rich (or CG-poor), then each of its subparts (SSU, LSU, divergent domains) likewise tends to be CG-rich (or CG-poor). Additionally, in nearly every taxon, SSU has the lowest CG-content, followed by the core of 28S, followed by the CG-rich divergent domains.

### 3.4. Trees from the independent-sites model

#### 3.4.1. 197 taxa

Fig. 2 shows the best ML tree calculated by Garli for the 197 rRNA sequences of this study. Garli and RAxML yielded essentially identical trees and bootstrap-support values. The "mid-level" phylogenetic relations obtained here were similar to those found by the previous rRNA-based studies of metazoan subgroups; that is, we found similar relations within Vertebrata (see Mallatt and Winchell, 2007), within Pancrustacea (Mallatt and Giribet, 2006), within Annelida (Struck et al., 2007), within Platyhelminthes (Lockyer et al., 2003), and in Cnidaria (Collins et al., 2006). However, compared to the previous studies we found less support for some low-level clades near the end branches of the tree (right side of Fig. 2). This loss of resolution happened, for example, within jawed vertebrates, Malacostraca, and Insecta, in comparison to

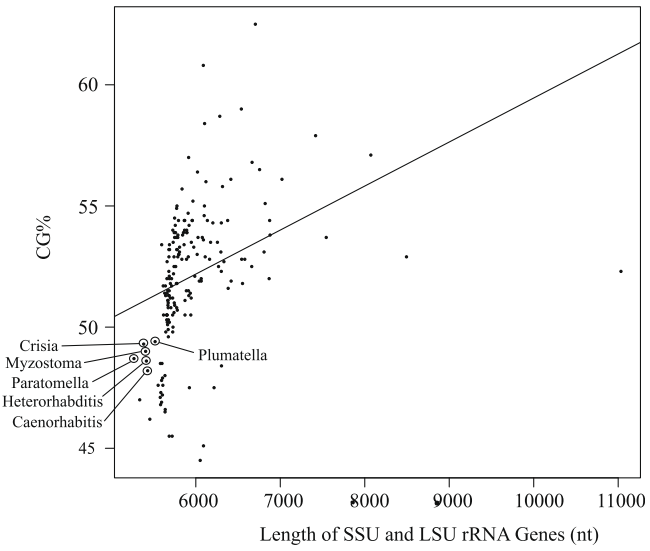


**Fig. 1.** Relation between the estimated total lengths of rRNA genes and their percent CG-content in the animal taxa of this study. For the raw data, see Supplementary material, S1. The regression line shows a positive correlation ($\rho = 0.51$, $P < 0.01$). The circled dots at lower left are from the shortest bilaterian sequences, listed in Table 7A: *Paratomella*, *Crisia*, *Myzostoma*, *Heterorhabditis*, *Caenorhabditis*, and *Plumatella* (but *Bugula*, whose 28S gene has a large, unique insert, is not included).

the findings of Mallatt and Giribet (2006) and Mallatt and Winchell (2007). Actually, this was expected because here we had to trim the rRNA-gene sequences down to their more-conserved parts that are alignable across all the Metazoa.

This 197-taxon tree supports some of the currently accepted, major clades, such as Bilateria, Vertebrata, and Ambulacraria, but a few oddly placed taxa prevent it from supporting the basic clades of the new animal phylogeny (Halanych, 2004), namely the Ecdysozoa, Lophotrochozoa, and Protostomia. We suspected the problem stemmed from the disruptive effect of the most-divergent sequences with the longest branches, and we noticed that many of these sequences go together in a "mongrel assemblage" that is located within 'Ecdysozoa (mostly)' below the arthropods in Fig. 2. This strange assemblage consists of cephalopod molluscs (*Vampyroteuthis* and *Nautilus*), the chaetognath arrowworm (*Sagitta*), the symphylan and pauropod myriapods (e.g., *Hanseniella* and *Allopauropus*) and the onychophoran worms (*Peripatus* and *Peripatoides*), with only the onychophorans occupying their accepted position (as relatives of arthropods: Ruppert et al., 2004). Besides being divergent, most of the nine sequences in this assemblage are CG-rich (double daggers near the bottom of Supplementary material, S1).

#### 3.4.2. 194 taxa

We proceeded to test whether eliminating these divergent sequences yielded support for Ecdysozoa, Lophotrochozoa, and Protostomia (Table 2). First we omitted the three sequences that had received positive support for their biologically unlikely positions near the onychophorans (with bootstrap values above 60%): squid *Vampyroteuthis*, *Nautilus,* and chaetognath *Sagitta*. That is, we assumed the cephalopod molluscs and the limbless arrowworms are not in Panarthropoda. The resulting 194-taxon tree supported Ecdysozoa (76%: Table 2).

#### 3.4.3. 193 taxa

To decide which of the "mongrel sequences" to eliminate next, we first had to discover which ones are in this assemblage naturally, as opposed to being drawn in by long-branch-attraction. To

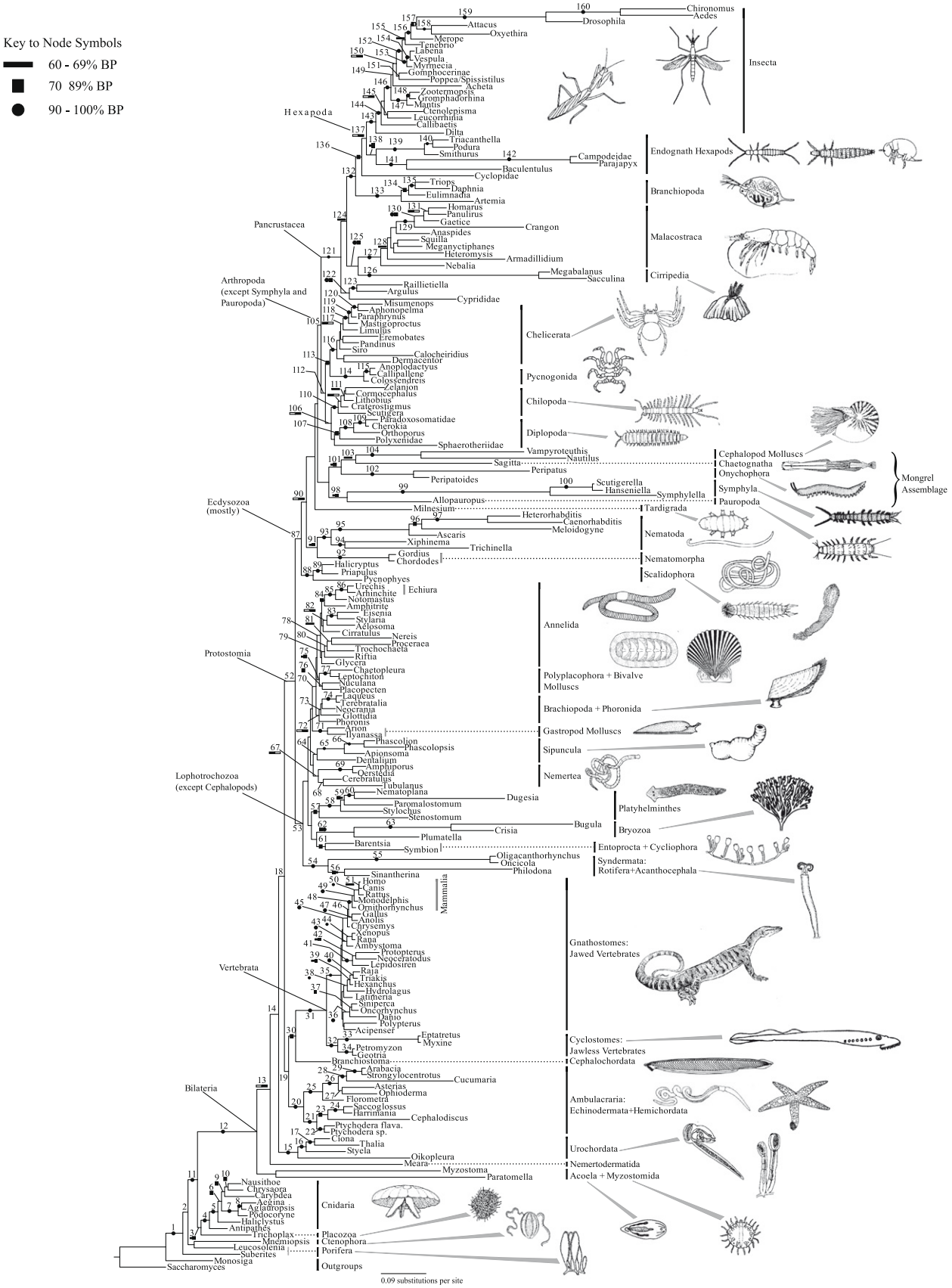


**Fig. 2.** Maximum likelihood tree calculated from nearly complete LSU+SSU rRNA genes of 197 taxa and 4862 nt, by the Garli program and the independent-sites model. The LnL value of this optimal tree is −194,134.0. The numbered nodes, 1–160, are those with bootstrap support above 50%, or those of interest. The symbols on the nodes indicate clades with bootstrap probabilities (BP) of 60–69% (rectangles), 70–89% (squares), 90–100% (circles), and <50% (no symbol), as calculated by Garli (left half of the symbol) and then by RAxML (right half of the symbol). The precise bootstrap values at the nodes are given in Supplementary material, S2. Most of the animal pictures were taken from Pimentel (1967), with the lizard picture from Colbert (1969), reproduced with the permission of R. Pimentel and John Wiley & Sons, Inc.

**Table 2**
Bootstrap support increases for protostome clades as divergent and unstable taxa are removed.

| # Taxa/ sequences | Ecdysozoa | Lophotrochozoa | Protostomia | Bilateria |
|---|---|---|---|---|
| 197[a] | <50%,<50% | <50%,<50% | <50%,<50% | 100%,100% |
| 194[b] | 76% | 58% | <50% | 100% |
| 193[c] | 82% | 60% | 53% | 100% |
| 190[d] | 80% | 63% | 53% | 100% |
| 186[e] | 97%,97% | 80%,87% | 82%,94% | 100%,100% |

*Note*. Where two numbers are given, the first was calculated by Garli and the second by RAxML. All other values are from RAxML. For a complete list of all the bootstrap-support values see Supplementary material, S2.

[a] 197, all sequences were included.

[b] 194, three sequences were omitted, from cephalopods (*Vampyroteuthis*, *Nautilus*) and the chaetognath (*Sagitta*).

[c] 193, four sequences were omitted, from cephalopods (*Vampyroteuthis*, *Nautilus*), chaetognath (*Sagitta*), and pauropod (*Allopauropus*).

[d] 190, seven sequences were omitted, from cephalopods (*Vampyroteuthis*, *Nautilus*), chaetognath (*Sagitta*), pauropod (*Allopauropus*), and symphylans (*Hanseniella*, *Scutigerella*, and *Symphylella*).

[e] 186, the 11 unstable sequences in Table 4 were omitted.

do this, we calculated the position of each mongrel sequence after all the others were removed (Table 3). Results showed that only the onychophoran, chaetognath, and pauropod sequences truly go here near the base of the arthropods, but that the cephalopod and symphylan sequences had been attracted here artifactually, because the latter two naturally join, much more logically, with another mollusc and with the other myriapods, respectively (this was also found by Giribet and Wheeler, 2001). Given this finding, we knew to eliminate *Allopauropus* next because it was the only sequence that had gone naturally in a biologically incorrect place (as a myriapod that incorrectly went outside of arthropods) and had not already been eliminated (as the chaetognath had: see previous paragraph). After *Allopauropus* was removed (193 taxa, Table 2), the bootstrap support for Ecdysozoa was higher (82%) and the support for Lophotrochozoa crept up toward significance (60%).

### 3.4.4. 190 taxa

Then, we tested the effect of eliminating the remaining out-of-place taxa from the assemblage by removing the symphylan sequences, of *Hanseniella*, *Symphylella*, and *Scutigerella*. However, the bootstrap-support values remained about the same (Table 2: 190 taxa): 80% for Ecdysozoa and 63% for Lophotrochozoa.

**Table 3**
True positions of the divergent rRNA sequences of the mongrel, CG-rich assemblage (Fig. 2): the tree position of each of these sequences, after all the other "mongrels" had been removed.

| Sequence/taxon | Position in the new ML tree (trees not shown) |
|---|---|
| 1. Onychophoran velvetworms (*Peripatoides*, *Peripatus*) | In Ecdysozoa, sister to Arthropoda |
| 2. Cephalopod molluscs (*Vampyroteuthis*, *Nautilus*) | In Lophotrochozoa, with mollusc *Dentalium*[a] |
| 3. Chaetognath arrowworm (*Sagitta*) | In Ecdysozoa, in or sister to Panarthropoda[b] |
| 4. Pauropod myriapod (*Allopauropus*) | In Ecdysozoa, sister to other arthropods[b] |
| 5. Symphylan myriapods (*Hanseniella*, *Scutigerella*, *Symphylella*) | In the myriapod arthropods[a] |

[a] These are biologically reasonable locations, meaning the original positions of the cephalopods and symphyla in the mongrel assemblage were artifacts of long-branch-attraction.

[b] These two are likely to be erroneous indicators of true relations, clashing with strong morphological evidence that says chaetognaths are not panarthropods and pauropods are actually true arthropods.

**Table 4**
Eleven taxa with unstable positions in the main tree (Fig. 2)[a], leaf-stability values <0.572, $P < 0.10$.[b]

*Meara* (nemertodermatid: 0.373), *Myzostoma* (myzostomid worm: 0.529), *Pycnophyes* (kinorhynch: 0.536), *Priapulus* (priapulan: 0.537), *Halicryptus* (priapulan: 0.537), *Nautilus* (cephalopod mollusc: 0.552), *Vampyroteuthis* (cephalopod mollusc: 0.552), *Gordius* (nematomorph: 0.569), *Chordodes* (nematomorph: 0.569), *Oncicola* (acanthocephalan: 0.571), *Oligacanthorhynchus* (acanthocephalan: 0.571)

[a] Based on low leaf-stability values (Thorley and Wilkinson, 1999) in the left tail ($P < 0.1$) of the following beta distribution: 197 leaf-stability values from 0.373 to 0.79, $\mu = 0.679$, $\alpha = 21.96$, $\beta = 10.38$, $q_{0.1} = 0.572$.

[b] For a list of all 197 leaf-stability values, see Supplementary material, S3.

### 3.4.5. 186 taxa

Finally, we took an entirely different approach to removing disruptive sequences, by identifying those with the most unstable positions on the tree. The leaf-stability test found 11 sequences to have low stability values with $P < 0.1$ (Table 4 and Supplementary material, S3). Eliminating these 11 from the 197 taxa yielded a 186-taxon tree (Fig. 3) that supported Protostomia (82% and 94% bootstrap values with Garli and RAxML), as well as Lophotrochozoa (80% and 87%) and Ecdysozoa (97% and 97%) (Table 2).

Thus, trimming progressively more of the most-divergent or unstable sequences produced successively more support for Ecdysozoa, Lophotrochozoa, and Protostomia (Table 2).

Although the trimmed, 186-taxon tree of Fig. 3 fits the new animal phylogeny, it has a few "flaws." First, it continues to show bad resolution within the Lophotrochozoa, a difficult clade for molecular phylogeny in general (Dunn et al., 2008; Passamaneck and Halanych, 2006). Second, it still shows the symphylans and pauropod outside of arthropods, but this is actually not a flaw because it is not significantly supported (the bootstrap value is under 50%). Third, the tree shows paraphyly of the chordates by supporting the urochordates (tunicates) as the sister clade of all the protostomes and other deuterostomes (85% by RAxML, 77% by Garli). Note that this position for tunicates was also present with 197, 194, 193, and 190 taxa, but it was not significantly supported there (all <50%: clade 18 in Supplementary material, S2).

The bootstrap-support values for all clades in the five analyses, from 197 to 186 taxa, are listed in Supplementary material, S2. These values are generally quite uniform across the analyses (except for Ecdysozoa, Lophotrochozoa, and Protostomia, as shown in Table 4). This uniformity implies that the rogue sequences did not often disrupt the positions of other, stable taxa in the trees.

## 3.5. Hypothesis testing

Another way to determine the accuracy of the 197-taxon tree in Fig. 2 is to use the statistical AU test to find if the dataset incorrectly rejects some reasonable or widely accepted phylogenetic hypotheses (Table 5, Hypotheses 2-6). These hypotheses explore the unlikely polyphyly of myriapods, of Ecdysozoa, of Chordata, and of Mollusca. Note that one of these, number 6, was constrained to include all four of these clades together, so Hypothesis 6 fits current views the best.

We also tested eight more hypotheses of current taxonomic interest, such as the relations of the Acoelomorph worms (Hypotheses 7–8: Telford et al., 2003; Wallberg et al., 2007); the reality of a clade of platyhelminths and syndermatids (Syndermata = rotifers with acanthocephalan worms) (Hypothesis 9: Giribet, 2008; Passamaneck and Halanych, 2006; Struck and Fisse, 2008); whether myzostomid worms are annelids (Hypothesis 10: Bleidorn et al., 2007); relations of the simple Placozoa animals (Hypotheses 11–12: Srivastava et al., 2008); and of chaetognaths (Hypotheses 13–14: Marlétaz et al., 2006; Matus et al., 2006; Philippe et al., 2007).

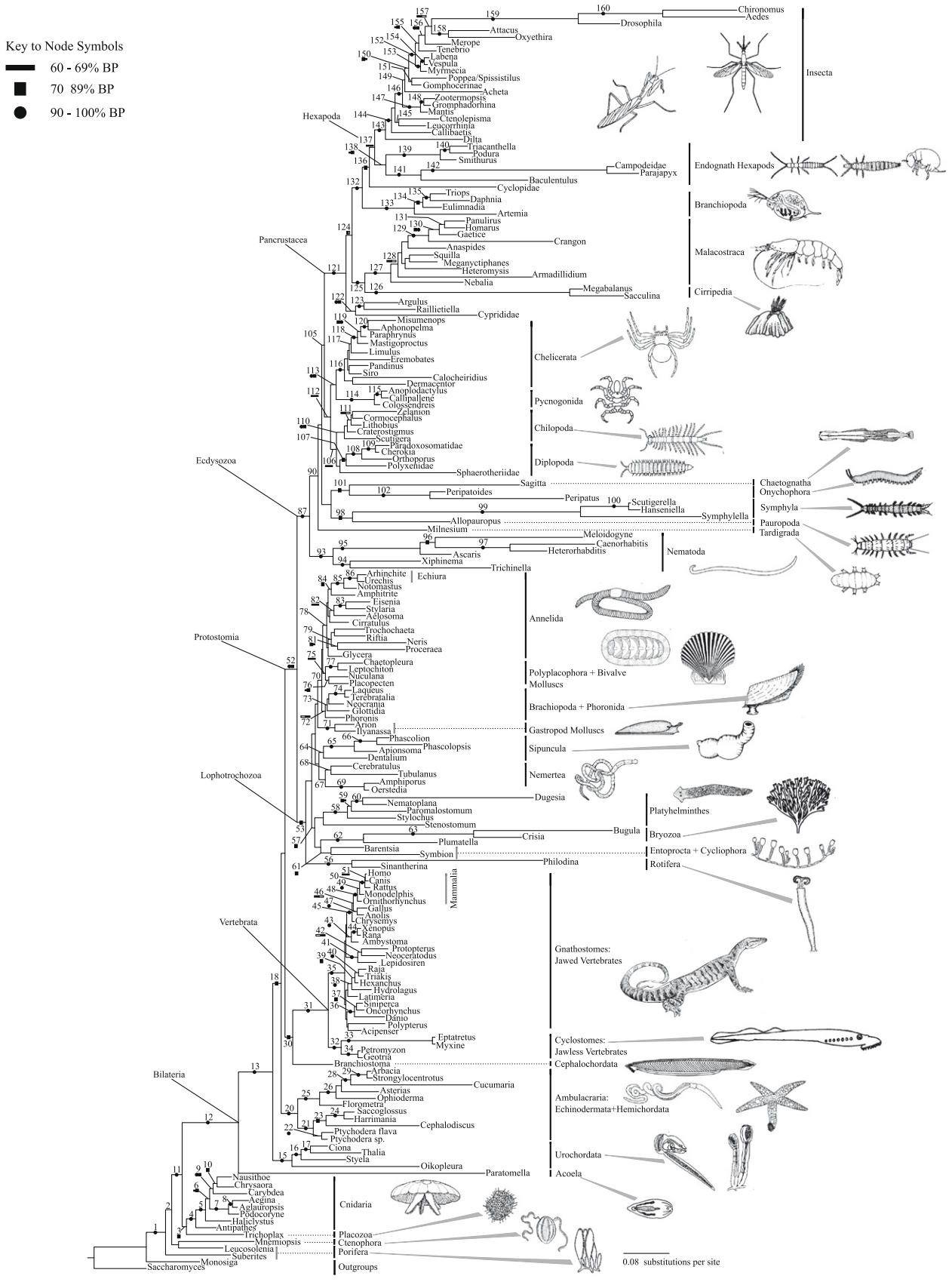


**Fig. 3.** Maximum likelihood tree calculated by Garli from the rRNA genes of just 186 taxa. That is, the 11 taxa with the lowest leaf-stability values in Table 4 were removed from the parent dataset of 197 taxa. The LnL value of this tree is −179,686.7. The numbers and symbols at the nodes mean the same as in Fig. 2.

**Table 5**
Hypothesis testing: results of the AU test, with rejection at $P < 0.05$ ($^*$ means rejected).

| | Hypothesis that was tested | LnL | Probability |
|---|---|---|---|
| 1. | Best tree (Fig. 1) | −194,134.0 | 0.908 |
| 2. | Myriapod monophyly | −194,151.0 | 0.347 |
| 3. | Ecdysozoa[a] | −194,154.1 | 0.397 |
| 4. | Chordate monophyly | −194,145.0 | 0.365 |
| 5. | Molluscan monophyly | −194,190.9 | 0.134 |
| 6. | All of the above (#2–#5) | −194,237.0 | 0.012* |
| 7. | Acoelomorpha is with Platyhelminthes[b] | −194,215.4 | 0.002* |
| 8. | Acoelomorpha is monophyletic, and it branches from base of Bilateria | −194,153.2 | 0.177 |
| 9. | Platyhelminthes + Syndermata | −194,142.6 | 0.506 |
| 10. | *Myzostoma* in Annelida | −194,165.3 | 0.122 |
| 11. | Placozoa (*Trichoplax*) is sister to all metazoans except Porifera | −194,177.2 | 0.060 |
| 12. | Placozoa (*Trichoplax*) is sister to all other Metazoa | −194,179.7 | 0.044* |
| 13. | Chaetognath (*Sagitta*) as basally branching lophotrochozoan | −194,192.9 | 0.012* |
| 14. | Chaetognath (*Sagitta*) as basally branching protostome | −194,199.0 | 0.008* |

[a] This is Ecdysozoa excluding *Sagitta*, *Nautilus*, and *Vampyroteuthis*, which likely are not in Ecdysozoa.
[b] Classical, morphology-based idea: see Ruppert et al. (2004).

As shown by asterisks in Table 5, five of the hypotheses were rejected: 7. Acoelomorphs with the platyhelminth flatworms; 12. Placozoa as sister to all other metazoans; 13. Chaetognath as the basally branching lophotrochozoan; 14. Chaetognath as the basally branching protostome. Hypothesis 6, the monophyly of myriapods, of Ecdysozoa, of chordates, and of molluscs, was rejected, although none of these clades was rejected individually. That is, Hypotheses 2–5 were not rejected.

To explore why the four, combined clades of Hypothesis 6 were rejected, we ran new analyses that were constrained to just two of the four clades, in all possible combinations: myriapods and Ecdysozoa each monophyletic, then chordates and molluscs each monophyletic, then Ecdysozoa and molluscs, etc. We found that whenever molluscs were in a pair, the *P* value for the resulting hypothesis declined to near or below the rejection level (0.05), but otherwise the values did not come close to rejection (i.e., not for myriapods and Ecdysozoa, for myriapods and chordates, nor for Ecdysozoa and chordates). This implies that the mollusc sequences—likely the divergent sequences of the cephalopods—contributed most heavily to our likelihood tree rejecting the combined, conventional Hypothesis 6. Mollusc paraphyly remains one of the most difficult problems faced by rRNA-based phylogeny.

### 3.6. Trees from the PHASE analyses, paired and unpaired sites

Fig. 4A shows the best mcmc tree that was calculated by PHASE using the mixed model that considered both paired and unpaired sites. For comparison, Fig. 4B shows the corresponding mcmc tree that PHASE calculated with the traditional independent-sites model, in which all sites were considered unpaired. Neither tree seems to differ much from the ML-unpaired tree of Fig. 2. That is, both still show the mongrel assemblage of cephalopods, chaetognath *Sagitta*, onychophorans, pauropod and symphylans; and both still fail to place the urochordates among the chordates. The mixed model put the mongrel assemblage and the bryozoans in different positions on its tree than did the unpaired model, but all these positions had low, nonsignificant support values. The few supported differences (at the four asterisks in Fig. 4A) included the positions of horse-shoe crab *Limulus*, of the brachiopod *Glottidia* versus *Phoronis* worm, of the tardigrade *Milnesium* and of the scaphopod mollusc *Dentalium*—but all these are difficult taxa that have not been placed stably in past rRNA studies (Mallatt and Giribet, 2006; Passamaneck and Halanych, 2006). One is tempted entirely to dismiss these few differences between the mixed and the unpaired PHASE trees, by attributing them to the tendency of mcmc-based methods to give exaggerated support for short internal branches (see p. 1006 in Mallatt and Winchell, 2007).

Nonetheless, when we compared the likelihood scores of the mixed and unpaired PHASE trees, we found a difference (Table 6). More specifically, we took the Newick formula of each PHASE tree of Fig. 4 and compared it to the formula of the ML tree of Fig. 2, using the AU test. The scores of the two unpaired-sites trees (PHASE and ML) did not differ from one another, but they both differed from the score of the mixed-sites tree ($P = 0.035$). Note that this AU test involved translating a mixed-sites tree into a different, unpaired-sites model, from which the mixed tree itself had not originally been calculated, so the AU results cannot say the mixed tree is *worse* than the unpaired-sites trees. It can only say that these two trees are *different*—and, as argued in the previous paragraph, the differences seem trivial in a qualitative, biological sense.

## 4. Discussion

This study assembled almost 200 nearly complete rRNA-gene sequences from across Metazoa, provided a new alignment based on paired sites, and adopted mixed-sites models to compute phylogeny. The most encouraging finding is that no more than 6% of the sequences, or ~11 of 197, had to be omitted to generate a fairly reasonable tree (Fig. 3). However, the results reveal some weaknesses as well.

### 4.1. Persistent problem of difficult sequences

The difficult rRNA sequences found in our broad study are the same ones found by past studies that used narrower subsets of metazoan sequences. However, with the full range of metazoan taxa now included as its outgroups, our study found different and presumedly more-accurate (but still taxonomically wrong) positions for the rogues than did the other studies, which used fewer and shallower outgroups. In general, these rogues proved to be farther out-of-place on the rRNA tree (Fig. 2) than was previously suspected. That is, they are now in *more*-incorrect positions taxonomically. Specifically, the cephalopod sequences are long-branch-attracted into Ecdysozoa and are not in the Lophotrochozoa where Passamaneck et al. (2004) found them; pauropods do not go in Myriapoda and symphylans are long-branch-attracted away from Myriapoda, in contrast to the findings of Gai et al. (2006); the urochordate sequences are farther away from those of chordates and deuterostomes than was indicated before (by Mallatt and Winchell, 2007); the unlikely position of the non-molting chaetognath in the Ecdysozoa on the tree is here shown to be real and not an artifact of long-branch-attraction (cf., Mallatt and Winchell, 2002); and the *Myzostoma* sequence does not go in Lophotrochozoa after all (cf., Passamaneck and Halanych, 2006). This finding that the rogues are more "roguish" is disappointing. However, only by identifying where the rogue sequences truly lie can we hope to solve their misplacement problem in the future.

### 4.2. Better models are still needed

#### 4.2.1. Paired sites did not help

It is also disappointing that mixed models, which are supposed to be more realistic than unpaired-sites models, did not improve our tree (Fig. 4) or correctly position the dozen or so rogue rRNA sequences. By contrast, most other phylogenetic studies that modeled base pairing in rRNA did report improvement (Dohrmann et al., 2008; Erpenbeck et al., 2007; Hudelot et al., 2003; Kjer,

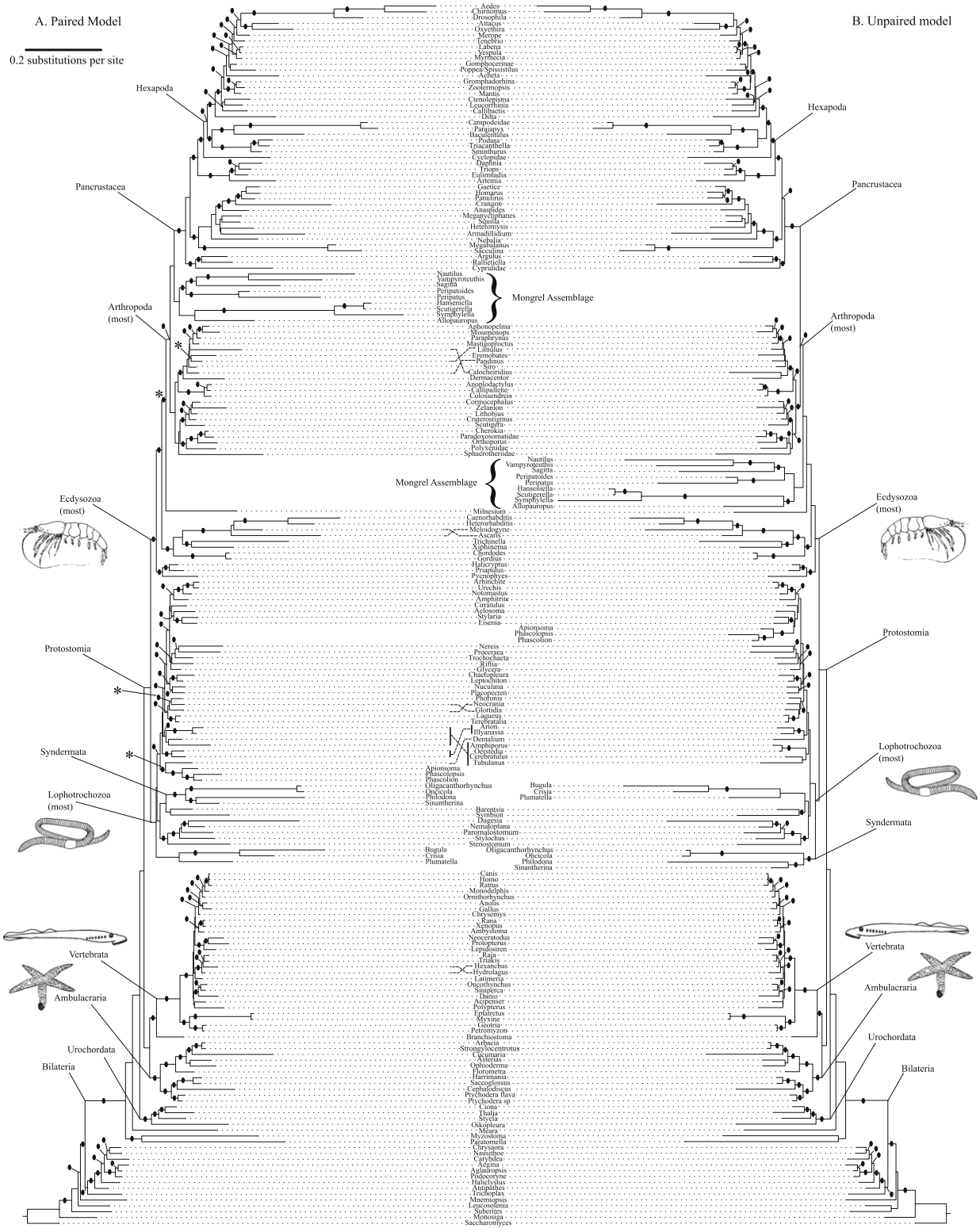


**Fig. 4.** Best mcmc trees calculated by PHASE from the nearly complete LSU + SSU rRNA genes of 197 taxa and 4862 nt. (A) From a mixed model, accounting for substitutions at both paired and unpaired sites. (B) From a traditional, independent-sites model that considers all sites as unpaired. The dark dots at the nodes indicate clades with posterior-probability support of $\geqslant 0.95$. The few clades that are marked in (A) with an asterisk (*) were supported at $\geqslant 0.95$ by one of the two models but had no support (<0.50) from the other model. The LnL values of the best trees in the mcmc chains are −167,816.3 (A, mixed model) and −194,267.5 (B, unpaired model). Notice that the topologies and posterior-probability values of both trees are similar, and that both trees resemble the ML tree in Fig. 2.

2004; Telford et al., 2005; Ware et al., 2008; but not Rix et al., 2008). Although our study used more taxa or more nucleotides than did those studies, and thus may not be comparable to them, the question remains of why our mixed model did not help. Per-

**Table 6**
Mixed- versus unpaired-sites analyses: AU test of the best mixed-sites tree versus unpaired-sites trees (* means rejected at $P < 0.05$).

| Trees | LnL score | Difference | Probability |
|---|---|---|---|
| 1. Mixed, paired and unpaired (PHASE tree: Fig. 4A) | −194,199.7 | 65.7 | 0.035* |
| 2. Unpaired (PHASE tree: Fig 4B) | −194,137.9 | 3.9 | 0.494 |
| 3. Unpaired (Garli-ML tree: Fig. 2) | −194,134.0 | (Best) | 0.602 |

*Note*: To allow these comparisons, both of the best PHASE-tree topologies first had to be loaded into the Garli program, and then recalculated as Garli-ML trees.

haps any improvement we achieved from more-accurately modeling the substitutions of paired bases was canceled by a loss of resolution from having fewer characters. That is, using nucleotide pairs means the stems of rRNA have 50% fewer characters than if the nucleotides in the pairs had been treated as separate sites, and fewer characters bring deflated support values. In other words, the *un*paired model overoptimistically *inflates* the support values because it wrongly assumes the two sites in a pair evolve independently of one another (Jow et al., 2002; Rix et al., 2008).

Truly determining why paired/mixed-sites models did not help will require further study. Adjusting the number of model parameters, from the rRNA7 used here to rRNA16 (Savill et al., 2001; von Reumont et al., 2009), for example, might reveal and reduce any problems of over- or under-parameterization in the paired-sites framework. Then, ultimately, more-detailed models must be built and used, models that account for the evolution of the structural motifs in rRNA tertiary structure (Leontis and Westhof, 2003).

In the meantime, the best next-step will be to use "heterotachy" or "time-heterogeneous" models of tree reconstruction. These models let the evolutionary-model parameters change across lineages (Galtier and Gouy, 1998), especially the base-frequency parameters (Blanquart and Lartillot, 2006; Foster, 2004; Gowri-Shankar and Rattray, 2007). Besides potentially solving the problem of fast-evolving rogue taxa, time-heterogeneous models should also solve the problem of base-compositional heterogeneity in the rRNA dataset (Supplementary material, S1).

#### 4.2.2. Heterotachy models may be better

Recently, von Reumont et al. (2009) ran a pioneering study that used a time-heterogeneous, mixed-sites model on the 18S+28S rRNA genes of arthropods alone. Their findings are very encouraging. The first good result is that their heterogenous model neatly recovered and upheld most of the clades obtained previously from rRNA genes (see Fig. 2), such as a well-supported Pancrustacea, entognath hexapods, Nonoculata (proturan and dipluran hexapods), copepods as the sister clade of hexapods, pauropods and symphylans together as a clade (after Gai et al., 2006), and the same internal relationships within sea spiders, millipedes, centipedes, chelicerates, and holometabolous insects. Beyond such confirmation, their second, better result is that the time-heterogeneous model improved resolution in several places, such as recovering silverfish as the sister clade of winged insects. Von Reumont et al. (2009) added about 60 more arthropod sequences than we had, including some extremely divergent new rogues (e.g., a fast-evolving silverfish)—and recovered these rogues in more-reasonable positions than did the model without time-heterogeneity.

Despite its successes, the time-heterogeneous method still placed symphylans and pauropods artifactually outside the arthropods and near the onychophorans, and it had trouble resolving relations among the basic groups of crustaceans. Von Reumont et al. (2009) considered these problems to be a remaining limitation of the otherwise-improved method, but we suspect they were partly due to the fact that just two, divergent, outgroups were used to root their arthropod tree: the moderately divergent and only borderline-complete tardigrade sequence, from *Milnesium* (see columns 1 and 5 in Supplementary material, S1) and the highly divergent onychophoran sequences (*Peripatoides* and *Peripatus*). Adding more ecdysozoan outgroups, especially the ultra-conserved sequences of priapulid worms, might improve resolution at all the deeper nodes of arthropods in future time-heterogeneous studies.

In summary, the time-heterogeneous approach shows promise and should be used to analyze relationships of all the metazoans.

Although we have mapped future directions, it is too early to abandon the traditional, stationary, unpaired-sites models. To date, these models allow the most elaborate testing of phylogenetic hypotheses (see Table 5 and Mallatt and Winchell, 2007), and they take much less time and effort than do the paired-sites and time-heterogeneous approaches.

### 4.3. Comparison to other broad studies of metazoan relationships

#### 4.3.1. rRNA study of Paps et al. (2009)

Recently, Paps et al. (2009) used 18S and 28S rRNA genes to construct the relationships across Metazoa, with an alignment based on secondary structure and unpaired-sites models. Both our and their studies sought to overcome problems experienced by rRNA genes in resolving deep splits in Metazoa, problems stemming from the most-divergent sequences. But the two studies used different strategies to achieve this goal. Paps et al. (2009) used only the least-divergent sequences within each phylum to get relationships *between* phyla, whereas we used the full range of available sequences and got many relations *within* the phyla as well. Paps et al. (2009) got more resolution within Lophotrochozoa, finding, for example, that rRNA joins sipunculan worms significantly with annelids and joins brachiopods with phoronids. By contrast, we got more resolution in the Ecdysozoa and Vertebrata.

Still, the basic-metazoan relationships found by Paps et al. (2009) were much like those we found. Both studies recovered Ecdysozoa, Lophotrochozoa, and Protostomia, but not Deuterostomia or Chordata. Both obtained chaetognaths in Ecdysozoa, and in both the AU tests rejected chaetognaths as basal protostomes or basal lophotrochozoans.

By successively removing the most-divergent sequences from their analysis, Paps et al. (2009) showed that these rogues had not disrupted the relations of any neighboring taxa in the original tree. That is, they skillfully revealed the lack of any long-branch-attraction (LBA) in their rRNA-gene tree. Although our study also found little LBA artifact, we did demonstrate some: namely, that the divergent onychophoran, chaetognath, and pauropod sequences attracted the cephalopod and symphylan sequences into the mongrel assemblage (this is the logical interpretation of our findings in Table 3).

At the base of the Bilateria, our study did not resolve the relationships between the Acoel (*Paratomella*) and nemertodermatid (*Meara*) worms as well as Paps et al. (and Wallberg et al., 2007) did. Presumedly, the reason those researchers did better was because they included multiple sequences from each Acoelomorph phylum, whereas we used just one.

Despite any weaknesses in our study, the fact that it only had to exclude about 6% of its sequences to get a reasonable tree remains a plus. This could mean that the *a priori* screening and the chimeric assembly of conserved 18S with conserved 28S sequences, as performed by Paps et al. (2009) to reach this same end, may not be necessary in the future.

#### 4.3.2. Protein-gene study of Dunn et al. (2008)

Our rRNA-based results can be compared to those from the largest study that was based on multiple protein genes (Dunn et al., 2008). That study used 77 taxa and 21,152 concatenated amino

acids to reconstruct metazoan relations. After removing the 11–13 most unreasonably or unstably positioned sequences from the trees, both their study and ours found significant support for the basic clades Bilateria, Ecdysozoa, Lophotrochozoa, Protostomia, and Ambulacraria. The multigene study did better in some ways. It alone found support for Deuterostomia and Chordata (accepted clades that are well-defined by morphological evidence: Kardong, 2006; Ruppert, 2005), it found better resolution within the Lophotrochozoa, and it alone identified onychophorans as the sister clade of arthropods. The rRNA study, however, did better in other ways, placing the entoproct (*Barentsia*) significantly within Lophotrochozoa (Fig. 3), and, because of its wider taxon sampling, it put more taxa into likely positions: the cycliophoran *Symbion* with the entoproct *Barentsia* (also supported by Funch and Kristensen, 1995; Giribet et al., 2004b; Passamaneck and Halanych, 2006); the proturan, diplurans, and collembolans together as endognath hexapods (discussed in Giribet et al., 2001, 2004a; Luan et al., 2005); acanthocephalans with rotifers in Syndermata (García-Varela and Nadler, 2006; Passamaneck and Halanych, 2006); and perhaps of significance, it put Placozoa together with cnidarians (see below).

The good overall agreement between phylogenies obtained from protein genes and from rRNA genes is remarkable, synergistic evidence that the new animal phylogeny, with Ecdysozoa and Lophotrochozoa, is correct.

### 4.4. Additional positive findings

#### 4.4.1. Extreme shortening means extreme evolution

In some bilaterians the 28S gene, which is larger than the 18S and contains more phylogenetic information (Mallatt and Winchell, 2002), has been shortened during evolution. As shown in Table 7A, these shortened sequences are from the acoel, myzostomid, bryozoans, and rhabditid nematodes. In *non*-bilaterian metazoans, the 28S genes are almost as long (averaging 3626 nt) as those of typical bilaterians (averaging 3700 or 3800 nt), and non-bilaterian rRNAs have the same complex secondary structure as rRNAs of bilaterians, with all the same stems and loops. This means there was an ancestral-metazoan length and secondary structure. Thus, it stands to reason that any bilaterian 28S genes shorter than 3626 nt do not reflect some primitive state but instead have been reduced in length (Table 7A). Likewise, the bilaterians whose *complete* rRNA is shorter than 5592 nt (the non-bilaterian average) have mostly shortened their 28S, because the lengths of the SSU and 5.8S genes do not vary much across taxa (see Supplementary material, S1). As mentioned, size changes in rRNA genes mostly involve their variable regions (Gillespie et al., 2005) and the size reductions here were mostly due to deletions in the large divergent domains of 28S.

The rRNA sequences of the taxa in Table 7A must have experienced extreme evolutionary modification. They not only are shortened but also are divergent, with relatively long branches in Fig. 2. Furthermore, most are comparatively AT-rich and CG-poor (Fig. 1, Supplementary material, S1). Their lower overall-CG-content must be due to two factors: (1) decreased CG throughout the whole LSU of these taxa (because their CG-content is low in all the subparts of LSU: columns 12–14 in S1); and (2) the shortening of the most CG-rich part, the divergent domains, as documented above. Consistent with this claim for extensive evolutionary reorganization, their divergent genes were frustratingly difficult to align with those of other bilaterians, much more so than were the sequences of the non-bilaterian metazoans, most of which are conserved and short-branched (Medina et al., 2001; Figs. 2 and 3). These findings increase the danger that the AT-enriched sequences of *Paratomella* and *Myzostoma* are displaced and long-branch-attracted down the tree toward the AT-rich non-bilaterian sequences (Fig. 2). Also, we found that the *Bugula* and *Crisia* sequences went to such a basal–bilaterian position when we omitted the other, less-divergent bryozoan, *Plumatella*, and then recalculated the tree (not shown).

We also re-explored the positions of the acoel *Paratomella* and of *Myzostoma*, by omitting one of these, then the other, from the analysis. We found both of these sequences went genuinely at the base of the Bilateria, but we still suspect they are so modified that this basal location is not to be trusted. Adding to this suspicion, the 28S molecules of *Paratomella* and *Myzostoma* have lost some stems that are present in almost all other metazoans, both bilaterian and non-bilaterian: For example, the *Myzostoma* 28S lacks stem D2-3b and the *Paratomella* 28S lacks stems D2-3b and D8c. This is seen in Fig. 5, which shows how modified these divergent domains of the *Myzostoma* and *Paratomella* rRNA really are.

A shortening and reorganization of the genes by evolution might also explain why the urochordates, whose 28S and total-rRNA genes are the shortest among deuterostomes (Table 7, part B), incorrectly branch from near the base of Bilateria instead of with the other chordates (though generally without significant support: see Figs. 2 and 4, and Supplementary material, S2). The secondary simplification and divergent nature of the entire urochordate genome has been amply documented (Delsuc et al., 2006; Putnam et al., 2008).

It is important to acknowledge that not all the AT-rich sequences among the bilaterians go in unlikely places on the rRNA tree. For example, the unshortened sequences of the flies (*Chironomus*, *Aedes*, *Drosophila*) and the shortened sequences of the rhabditid nematodes (*Caenorhabditis*, *Heterorhabditis*) go, as expected, in the insects and nematode clades—and with good bootstrap support. However, both these groups of AT-rich sequences have some tendency to move artifactually down the tree because that is what happened to them in past studies that used less-complete data, namely few taxa or only the 18S gene (flies: p. 182 in Mallatt et al., 2004; rhabditid nematode: Winnepinninckx et al., 1995).

#### 4.4.2. Groundwork for rRNA evolution

Our alignment is designed to be a valuable scientific resource. Laid out by secondary structure with homologous sites in the same columns for hundreds of diverse taxa, the alignment provides a template to which scientists can readily add and align more sequences. It can also serve as a reference to help improve computer programs that align rRNA by secondary structure, such as RNAsalsa (http://www.zfmk.de/web/Forschung/Abteilungen/AG_Wgele/Software/

**Table 7**
Taxa with shortened 28S and total-rRNA sequences.

| | |
|---|---|
| *Part A: Bilaterian sequences shorter than 3626 nt (28S rRNA) and 5592 nt (LSU + SSU) long; that is, shorter than the average length of non-bilaterian metazoan rRNA* | |
| 1. *Paratomella* | 3343 nt, 5268 nt |
| 2. *Bugula* | 3385[a], 5392[a] |
| 3. *Crisia* | 3388, 5383 |
| 4. *Myzostoma* | 3424, 5406 |
| 5. *Heterorhabditis* | 3493, 5415 |
| 6. *Caenorhabditis* | 3508, 5430 |
| 7. *Plumatella* | 3543, 5518 |
| *Part B: Shortest deuterostome sequences, compared to the 6123 nt average length of LSU+SSU for bilaterians* | |
| 1. *Oikopleura* | 3577 nt, 5554 nt |
| 2. *Thalia* | 3639, 5614 |
| 3. *Ciona* | 3676, 5648 |
| 4. *Styela* | 3700, 5660 |

*Note*. The total lengths of the rRNA genes of all 197 taxa are given at the start of the NEXUS file on the web site, http://purl.oclc.org/NET/rRNA/Metazoan_alignment, and in Supplementary material, S1.

[a] For *Bugula*, the 28S sequence (AY210457) has a unique insert of 433 nt near D10, which is not included in this total.

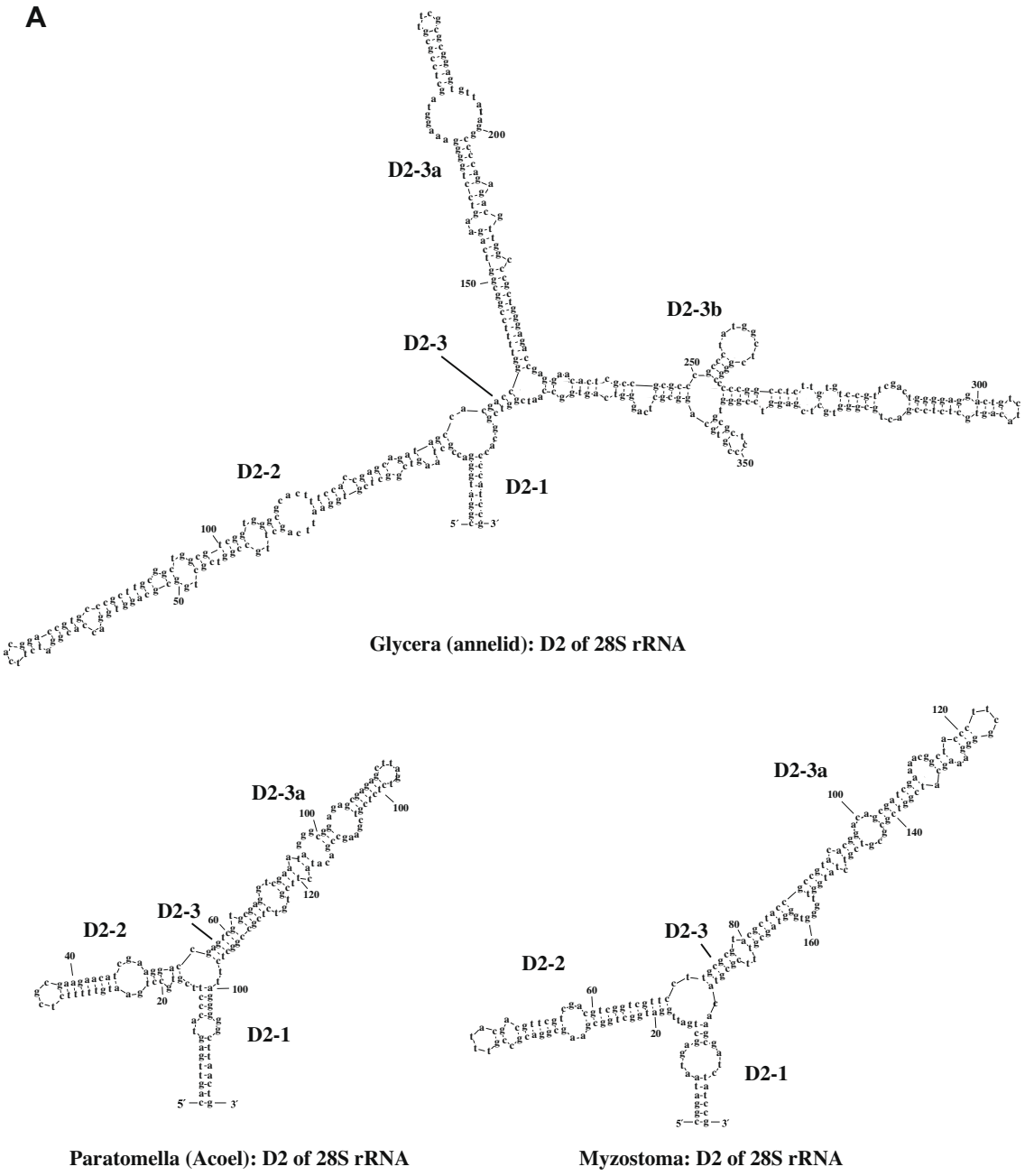


**Fig. 5.** Secondary structure of typical (*Glycera*) and shortened (*Paratomella*, *Myzostoma*) divergent domains of 28S rRNA, as reconstructed by mfold. (A) The D2 domains. (B) The D8 domains. For more such diagrams, of a conserved vertebrate, ecdysozoan, echinoderm, and cnidarian, see Supplementary material, S4.

RNAsalsa/index.en.html) and Infernal (Nawrocki et al., 2009). This is because our alignment has already subdivided the stems and loops into neat partitions, even in the difficult variable regions.

Furthermore, all this comparative information on rRNA genes will benefit non-phylogenetic studies of the higher-order structure of rRNA molecules and of how rRNA structure evolves (Bokov and Steinburg, 2009; Gutell et al., 2002; Noller, 2005).

### 4.5. Individual taxa, new and controversial

#### 4.5.1. New taxa and sequences

Ours is the first broad phylogenetic study to incorporate complete rRNA sequences from a marsupial or monotreme mammal, and it supports the accepted relations in Mammalia, of: (monotreme *Ornithorhynchus*, (marsupial *Didelphis*, (placentals *Homo*, *Canis*, *Rattus*))). These relations were first constructed from reproductive characters and other synapomorphies (Binenda-Edmonds et al., 2007; Kardong, 2006).

Upon completing the 28S rRNA sequence of the *Anolis* lizard, our study upheld the accepted view that birds are diapsid reptiles by producing 99% + support for the clade of '*Anolis* + bird *Gallus*' (see Fig. 2 and #47 in Supplementary material, S2). This supercedes an earlier, controversial, placement of birds with mammals, which was based on 18S genes alone (Hedges et al., 1990; for more, see Mallatt and Winchell, 2007).

The position of bryozoans is controversial because the different types of molecular and morphological evidence yield inconsistent pictures of bryozoan relations (Dunn et al., 2008; Giribet et al., 2000; Hausdorf et al., 2007; Peterson and Eernisse, 2001). A big debate is whether they are lophotrochozoans or not

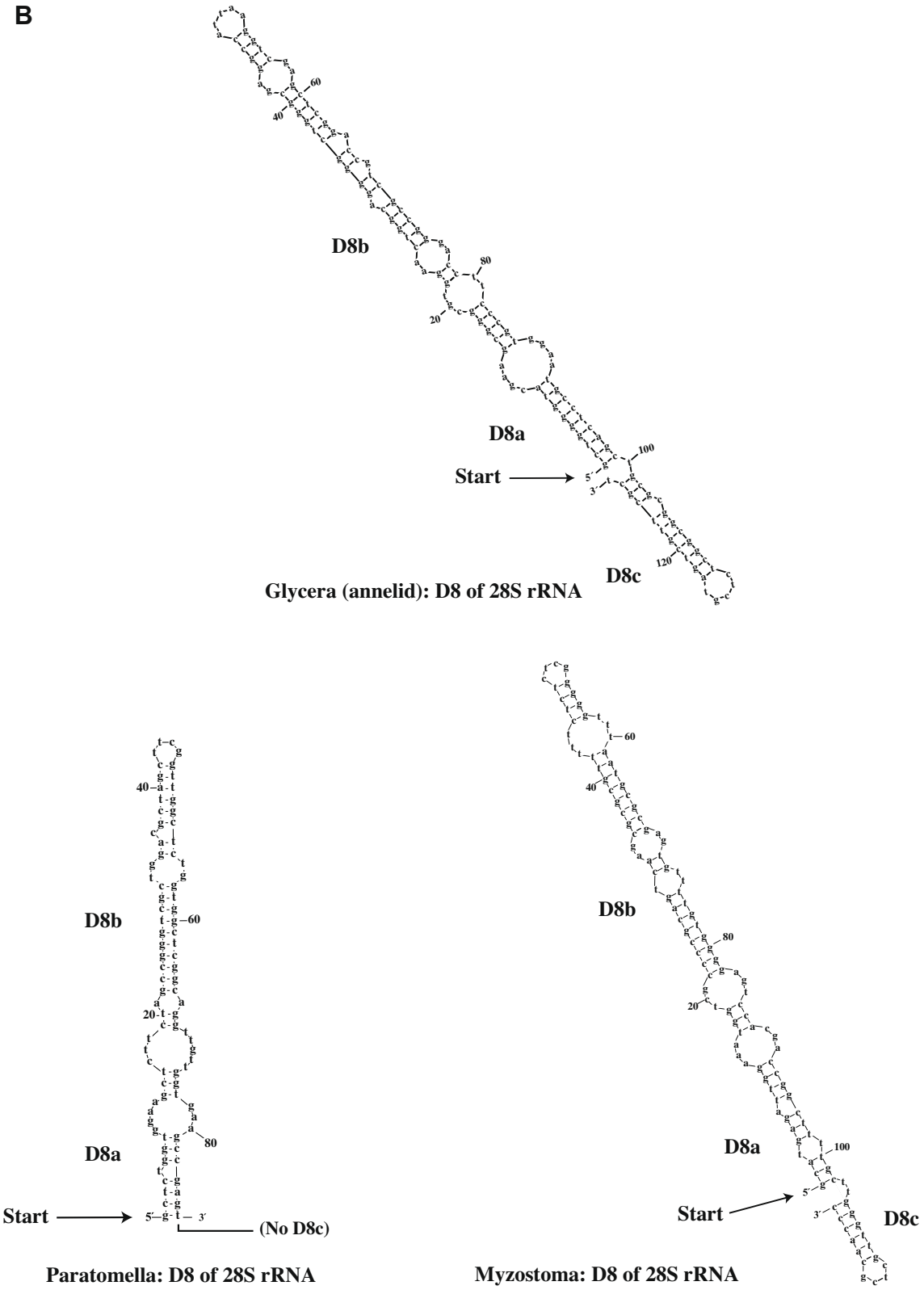


**Fig. 5** (*continued*)

(Halanych, 2004; Halanych et al., 1995; Paps et al., 2009). Here, we added the bryozoan *Plumatella* (H. Zhao, Q. Yang: DQ333339) and recovered Bryozoa as Lophotrochozoa. By this interpretation, the *Plumatella* sequence, as the least divergent among the reported bryozoans, with the shortest branch (Fig. 2), retains enough signal to show Bryozoa are lophotrochozoans.

The nematode rRNA sequence we added, from *Heterorhabditis*, joined with that of the other rhabditid nematode, *Caenorhabditis* (*C. elegans*), inside Ecdysozoa (Figs. 2 and 3). This finding seems

trivial, but it emphasizes that complete rRNA sequences now support rhabditids as Ecdysozoa, rather than as the sister to 'protostomes + deuterostomes.' That is, rRNA speaks against the Coelomata hypothesis (Rogozin et al., 2007; Zheng et al., 2007).

*4.5.2. Controversial taxa*

Myzostomids, small oval worms that live on echinoderms, are of uncertain affinity (Eeckhaut et al., 2000; Fauchald and Rouse, 1997; Giribet, 2008; Zrzavy et al., 2001). However, more and more evidence from protein-coding genes and morphology is placing them with annelids (Bleidorn et al., 2007; Giribet, 2008; Ruppert et al., 2004). Our trees place the shortened, AT-rich, divergent, and phylogenetically unstable rRNA sequence of *Myzostoma* far from the annelids, but with little bootstrap support (Fig. 2; Supplementary material, S2). The AU test did not reject *Myzostoma* as an annelid (Table 5, Hypothesis 10), so an annelid affinity does seem most likely when the bulk of the evidence is considered.

Acoelomorpha, consisting of acoels and nemertodermatids, is a controversial group. Debate has raged over whether the Acoelomorphs are the most basally arising Bilateria, and whether they are diphyletic (Dunn et al., 2008; Hejnol and Martindale, 2008a,b; Jondelius et al., 2002; Paps et al., 2009; Ruiz-Trillo et al., 2002, 2004; Telford et al., 2003; Wallberg et al., 2007). Although our results do not resolve these questions (Table 5), we favor the evidence from microRNA that shows acoels do represent the basal bilaterians (Sempere et al., 2007), and favor the morphological evidence that Acoelomorpha is monophyletic (with their synapomorphies being special cilia, a frontal organ, and unique body-wall musculature: Ruiz-Trillo et al., 2004).

For the present study, we used only one acoel (*Paratomella*) and one nemertodermatid (*Meara*). However, the rRNA sequences of these taxa are among the least divergent of the known sequences from their phyla (see Fig. 3 in Wallberg et al., 2007), so they should be good representatives. They contribute something new in showing that the Acoelomorph positions on the rRNA tree are less certain than was previously assumed. That is, the rRNA of acoels (*Paratomella*) has been simplified (Fig. 5) and thus could have lost its phylogenetic signal, and the rRNA of nemertodermatids (*Meara*) is unstably located on the tree (Table 4).

To reconcile as much of the evidence as possible, we accept Acoelomorph monophyly but propose that the divergent *Paratomella* sequence was pulled away from that of its true sister *Meara* by a long-branch-attraction toward the similarly AT-rich non-bilaterians nearby (Fig. 2).

Placozoa (*Trichoplax*) are small, flat, benthic marine animals with a simple body plan. They have only a few cell types, no mouth or digestive tube, no nervous system, and no left-right or anterior-posterior polarization of the body (Ruppert et al., 2004; Syed and Schierwater, 2002). Different sorts of evidence support different phylogenetic positions for placozoans: either as the sister group of all other Metazoa (structure of their mitochondrial genes: Dellaporta et al., 2006); or as the sister of the Eumetazoa, which are all the metazoans except poriferans (anatomical characters: Ruppert et al., 2004; multiple protein genes: Srivastava et al., 2008); or in or sister to the Cnidaria (one 18S-rRNA study: Cavalier-Smith and Chao, 2003); or as sister to the Bilateria (another 18S-rRNA study: Wallberg et al., 2004). For more on the extensive literature, see Schierwater et al. (2009).

Several past studies used nearly complete rRNA genes to explore placozoan relations (Collins et al., 2005; da Silva et al., 2007; Voigt et al., 2004). They found, however, that adding 28S to 18S actually reduced resolution in the tree, yielding no definite position for *Trichoplax*. By contrast, the present study, which used more taxa and a different alignment procedure, supported *Trichoplax* as the sister clade of cnidarians. It also rejected the hypothesis of Placozoa as sister of all other Metazoa (Table 5, Hypothesis 12). Thus, our trees favor the idea that placozoans are secondarily simplified animals rather than indicators of the primitive-metazoan body plan. This conclusion is by no means definitive, however, because rRNA genes are just a single line of evidence, because we sampled relatively few poriferans and single-celled outgroups, and because big new genomic and total-evidence studies indicate a more-basal position of placozoans, as the sister clade of diploblasts (Schierwater et al., 2009) or of Eumetazoa (Phillipe et al., 2009).

## 5. Summary

1. We built a structure-based alignment of nearly complete rRNA genes from across Metazoa, as a reference for future studies. Also, structural models were built for the D8 and D2 divergent domains of 28S rRNA from a range of metazoans (Fig. 5, Supplementary material, S4).
2. The tree calculated from all 197 taxa (Fig. 2) recovers many accepted clades but fails to show key groups of the new animal phylogeny. However, removing between 7 and 11 of the most-divergent or unstable sequences (Table 2; Fig. 3) did yield the key groups: Ecdysozoa, Lophotrochozoa and Protostomia.
3. Our use of sequences from across all of Metazoa provided a wider range of outgroups for the metazoan subclades of Protostomia, Ecdysozoa, Bilateria, etc. This demonstrated that the misplaced, rogue, sequences are concentrated in just two places on the tree (Fig. 2): 1. near the base of the panarthropods (cephalopods, chaetognath, symphylans and pauropod), and 2. near the base of the Bilateria (urochordates, myzostomid and possibly acoels).
4. The statistical AU test (Table 5) allowed major clades that were not present in the original 197-taxon tree: Myriapoda, Ecdysozoa, Chordata, and Mollusca. However, it rejected the presence of these four clades together in the same tree, largely because of the poor signal for molluscan monophyly in the dataset.
5. The rRNA genes of non-bilaterian animals have a higher AT content than do those of most bilaterians (Supplementary material, S1). Future phylogenetic reconstructions should account for this compositional heterogeneity by using "heterotachy" or "time-heterogeneous" models (Gowri-Shankar and Rattray, 2007; von Reumont et al., 2009).
6. Neither aligning by secondary structure nor the use of paired-sites evolutionary models yielded improved tree resolution, compared to non-paired sites methods (Fig. 4A versus Fig. 4B). Again, heterotachy models may one day solve this problem.
7. The 28S rRNA sequences of acoels and myzostomids are secondarily shortened (Table 7; Fig. 5), highly evolved, and experienced a preferential loss of C and G nt. Thus, their "basal–bilaterian" positions in rRNA trees are suspicious. Still, we ultimately accepted the result of the AU test that allows acoels arising from the base of Bilateria. Myzostomids, however, are not confined there by the AU test, and are better interpreted as annelids.
8. rRNA genes firmly support chaetognaths in Ecdysozoa, which contradicts their morphology and is an enigma of rRNA-based phylogenetics.
9. Our rRNA trees support Placozoa as the sister group of Cnidaria, and the AU test rejects them as the sister of all other Metazoa.
10. This study and that of Paps et al. (2009) show that rRNA genes match multiple-protein-coding genes (Dunn et al., 2008) in supporting the new animal phylogeny.

## Acknowledgments

Thanks are extended to Karl Kjer for teaching us the alignment method, to Meade Snoddy for computer support, to Gary Thorgaard in whose lab we sequenced many rRNA genes, and to Jack Sullivan who taught us much about molecular phylogenetics. Dr. Yoder was funded in part by NSF PEET Grant DEB-0328922 to Robert Wharton.

## Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ympev.2009.09.028.

## References

Aguinaldo, A.M.A., Turbeville, J.M., Lindford, L.S., Rivera, M.C., Garey, J.R., Raff, R.A., Lake, J.A., 1997. Evidence for a clade of nematodes, arthropods and other moulting animals. Nature 387, 489–493.

Baguñà, J., Martinez, P., Paps, J., Riutort, M., 2008. Back in time: a new systematic proposal for the Bilateria. Philos. Trans. R. Soc. B 363, 1481–1491.

Binenda-Edmonds, O.R.P., Cardillo, M., Jones, K.E., MacPhee, R.D.E., Beck, R.M.D., Grenyer, R., Price, S.A., Vos, R.A., Gittleman, J.L., Purvis, A., 2007. The delayed rise of present-day mammals. Nature 446, 507–512.

Blanquart, S., Lartillot, N., 2006. A Bayesian compound stochastic process for modelling nonstationary and nonhomogeneous sequence evolution. Mol. Biol. Evol. 23, 2058–2071.

Bleidorn, C., Eeckhaut, I., Podsiadlowski, L., Schult, N., McHugh, D., Halanych, K.M., Milinkovitch, M.C., Tiedemann, R., 2007. Mitochondrial genome and nuclear sequence data support myzostomids as part of the annelid radiation. Mol. Biol. Evol. 24, 1690–1701.

Bokov, K., Steinburg, S.V., 2009. A hierarchal model for evolution of 23S ribosomal RNA. Nature 457, 977–980.

Bourlat, S.J., Nielsen, C., Economou, A.D., Telford, M.J., 2008. Testing the new animal phylogeny: a phylum level molecular analysis of the animal kingdom. Mol. Phylogenet. Evol. 49, 23–31.

Bourlat, S.J., Nielsen, C., Lockyer, A.E., Littlewood, D.T.J., Telford, M.J., 2003. *Xenoturbella* is a deuterostome that eats molluscs. Nature 424, 925–928.

Cavalier-Smith, T., Chao, E. E.-Y., 2003. Phylogeny of Choanozoa, Apusozoa, and other Protozoa and early eukaryote megaevolution. J. Mol. Evol. 56, 540–563.

Chilton, N.B., Huby-Chilton, F., Gasser, R.B., Beveridge, I., 2006. The evolutionary origins of nematodes within the order Strongylida are related to predeliction sites within hosts. Mol. Phylogenet. Evol. 40, 118–128.

Colbert, E.H., 1969. Evolution of the Vertebrates, second ed. John Wiley & Sons, New York.

Collins, A.G., Cartwright, P., McFadden, C.S., Schierwater, B., 2005. Phylogenetic context and basal metazoan model systems. Inter. Comp. Biol. 45, 585–594.

Collins, A.G., Schuchert, P., Marques, A.C., Jankowski, T., Medina, M., Schierwater, B., 2006. Medusozoan phylogeny and character evolution clarified by new large and small subunit rDNA data and an assessment of the utility of phylogenetic mixture models. Syst. Biol. 55, 97–115.

da Silva, F.B., Muschner, V.C., Bonatto, S.L., 2007. Phylogenetic position of Placozoa based on large subuint (LSU) and small subunit (SSU) rRNA genes. Genet. Mol. Biol. 30, 127–132.

Dellaporta, S.L., Xu, A., Sagasser, S., Jakob, W., Moreno, M.A., Buss, L.W., Schierwater, B., 2006. Mitochondrial genome of *Trichoplax adhaerans* supports Placozoa as the basal lower metazoan phylum. Proc. Natl. Acad. Sci. USA 103, 8751–8756.

Delsuc, F., Brinkmann, H., Chourrout, D., Philippe, H., 2006. Tunicates and not cephalochordates are the closest living relatives of vertebrates. Nature 439, 965–968.

Dixon, M.T., Hillis, D.M., 1993. Ribosomal RNA secondary structure: compensatory mutations and implications for phylogenetic analysis. Mol. Biol. Evol. 10, 256–267.

Dohrmann, M., Janussen, D., Reitner, J., Collins, A.G., Worheide, G., 2008. Phylogeny and evolution of glass sponges (Porifera, Hexactinellida). Syst. Biol. 57, 388–405.

Dunn, C.W., Hejnol, A., Matus, D.Q., Pang, K., Browne, W.E., Smith, S.A., Seaver, E., Rouse, G.W., Obst, M., Edgecombe, G.D., Sorensen, M.V., Haddock, S.H.D., Schmidt-Rhaesa, A., Okusu, A., Kristensen, R.M., Wheeler, W.C., Martindale, M.Q., Giribet, G., 2008. Broad phylogenomic sampling improves resolution of the animal tree of life. Nature 452, 745–749.

Eeckhaut, I., McHugh, D., Mardulyn, P., Tiedemann, R., Monteyn, D., Jangoux, M., Milinkovitch, M.C., 2000. Myzostomida: a link between trochozoans and flatworms? Proc. R. Soc. Lond. B. 267, 1383–1392.

Erpenbeck, D., Nichols, S.A., Voigt, O., Dohrmann, M., Degnan, B.M., Hooper, J.N.A., Worheide, G., 2007. Phylogenetic analyses under secondary structure-specific substitution models outperform traditional approaches: case studies with diploblast LSU. J. Mol. Evol. 64, 543–557.

Fauchald, K., Rouse, G., 1997. Polychaete systematics: past and present. Zool. Scr. 26, 71–138.

Field, K.G., Olsen, G.J., Lane, D.J., Giovannoni, S.J., Ghiselin, M.T., Raff, E.C., Pace, N.R., Raff, R.A., 1988. Molecular phylogeny of the animal kingdom. Science 239, 748–753.

Foster, P.G., 2004. Modelling compositional heterogeneity. Syst. Biol. 53, 485–495.

Funch, P., Kristensen, R.M., 1995. Cycliophora is a new phylum with affinities to Entoprocta and Ectoprocta. Nature 378, 711–714.

Gai, Y.-H., Song, D.-X., Sun, H.-Y., Zhou, K.-Y., 2006. Myriapod monophyly and relationships among myriapod classes based on nearly complete 28S and 18S rDNA sequences. Zool. Sci. 23, 1101–1108.

Galtier, N., Gouy, M., 1998. Inferring pattern and process: maximum-likelihood implementation of a non-homogeneous model of DNA sequence evolution for phylogenetic analysis. Mol. Biol. Evol. 15, 871–879.

García-Varela, M., Nadler, S.A., 2006. Phylogenetic relationships among Syndermata inferred from nuclear and mitochondrial gene sequences. Mol. Phylogenet. Evol. 40, 61–72.

Gillespie, J.J., Yoder, M.J., Wharton, R.A., 2005. Predicted secondary structure of 28S and 18S rRNA from Ichneumonoidea (Insecta: Hymenoptera: Apocrita): impact on sequence alignment and phylogeny estimation. J. Mol. Evol. 61, 114–137.

Gillespie, J.J., Johnston, J.S., Cannone, J.J., Gutell, R.R., 2006. Characteristics of nuclear (18S, 58S, 28S and 5S) and mitochondrial (12S and 16S) rRNA genes of *Apis mellifera* (Insecta: Hymenoptera): structure, organization, and retrotransposable elements.. Insect Mol. Biol. 15, 657–686.

Giribet, G., 2008. Assembling the lophotrochozoan (=spiralian) tree of life. Philos. Trans. R. Soc. B 363, 1513–1522.

Giribet, G., Distel, D.L., Polz, M., Sterrer, W., Wheeler, W.C., 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.

Giribet, G., Edgecombe, G.D., Carpenter, J.M., D'Haese, C.A., Wheeler, W.C., 2004a. Is Ellipura monophyletic? A combined analysis of basal–hexapod relationships with emphasis on the origin of insects. Org. Divers. Evol. 4, 319–340.

Giribet, G., Edgecombe, G.D., Wheeler, W.C., 2001. Arthropod phylogeny based on eight molecular loci and morphology. Nature 413, 157–161.

Giribet, G., Sorensen, M.V., Funch, P., Kristensen, R.M., Sterrer, W., 2004b. Investigations into the phylogenetic position of Micrognathozoa using four molecular loci. Cladistics 20, 1–13.

Giribet, G., Wheeler, W.C., 2001. Some unusual small-subunit ribosomal RNA sequences of metazoans. Am. Mus. Novit. 3337, 1–16.

Gowri-Shankar, V., Jow, H., 2006. PHASE: a Software Package for Phylogenetics and Sequence Evolution. Version 2.0 manual available from: <http://www.bioinf.man.uk/resources/phase/>.

Gowri-Shankar, V., Rattray, M., 2007. A reversible jump method for Bayesian phylogenetic inference with a nonhomogeneous substitution model. Mol. Biol. Evol. 24, 1286–1299.

Gutell, R.R., 1994. Collection of small subunit (16S- and 16S-like) ribosomal RNA structures: 1994. Nucleic Acids Res. 22, 3502–3507.

Gutell, R.R., Gray, M.W., Schnare, M.N., 1993. A compilation of large subunit (23S and 23S-like) ribosomal RNA structures: 1993. Nucleic Acids Res. 21, 3055–3074.

Gutell, R.R., Lee, J.C., Cannone, J.J., 2002. The accuracy of ribosomal RNA comparative structure models. Curr. Opin. Struct. Biol. 12, 301–310.

Halanych, K.M., 2004. The new view of animal phylogeny. Annu. Rev. Nat. Ecol. Evol. Syst. 35, 229–256.

Halanych, K.M., Bacheller, J.D., Aguinaldo, A.M.A., Liva, S.M., Hillis, D.M., Lake, J.A., 1995. Evidence from 18S ribosomal DNA that lophophorates are protostome animals. Science 267, 1641–1643.

Hassouna, N., Michot, B., Bachellerie, J.P., 1984. The complete nucleotide sequence of mouse 28S rRNA gene. Implications for the process of size increase of the large subunit rRNA in higher eukaryotes.. Nucleic Acids Res. 12, 3563–3583.

Hausdorf, B., Helmkampf, M., Meyer, A., Witek, A., Herlyn, H., Bruchhaus, I., Hankeln, T., Struck, T.H., Lieb, B., 2007. Spiralian phylogenomics supports the resurrection of Bryozoa comprising Ectoprocta and Entoprocta. Mol. Biol. Evol. 24, 2723–2729.

Hedges, S.B., Moberg, K.D., Maxson, L.R., 1990. Tetrapod phylogeny inferred from 18S and 28S ribosomal RNA sequences and a review of the evidence for amniote relationships. Mol. Biol. Evol. 7, 607–633.

Hejnol, A., Martindale, M.Q., 2008a. Acoel development supports a simple planula-like urbilaterian. Philos. Trans. R. Soc. B 363, 1493–1502.

Hejnol, A., Martindale, M.Q., 2008b. Acoel development indicates the independent evolution of the bilaterian mouth and anus. Nature 546, 382–386.

Hillis, D.M., Dixon, M.T., 1991. Ribosomal DNA: molecular evolution and phylogenetic inference. Quart. Rev. Biol. 66, 411–453.

Hudelot, C., Gowri-Shankar, V., Jow, H., Rattray, M., Higgs, P.C., 2003. RNA-based phylogenetic methods: application to mammalian mitochondrial RNA sequences. Mol. Phylogenet. Evol. 28, 241–252.

Janke, A., Magnell, O., Wieczorek, G., Westerman, M., Arnason, U., 2002. Phylogenetic analysis of 18S rRNA and the mitochondrial genomes of the wombat, *Vombatus ursinus*, and the spiny anteater *Tachyglossus aculeatus*: increased support for the Marsupionta. J. Mol. Evol. 54, 71–80.

Jenner, R.A., Littlewood, D.T.J., 2008. Problematica old and new. Philos. Trans. R. Soc. B 363, 1503–1512.

Jondelius, U., Ruiz-Trillo, I., Baguñà, J., Riutort, M., 2002. The Nemertodermatida are basal bilaterians and not members of the Platyhelminthes. Zool. Scr. 31, 201–215.

Jow, H., Hudelot, C., Rattray, M., Higgs, P.G., 2002. Bayesian phylogenetics using an RNA substitution model applied to early mammalian evolution. Mol. Biol. Evol. 19, 1591–1601.

Kardong, K.V., 2006. Vertebrates. Comparative Anatomy, Function, Evolution, fourth ed. McGraw Hill Higher Education, Dubuque, IA.
Kjer, K.M., 2004. Aligned 18S and insect phylogeny. Syst. Biol. 53, 506–514.
Kjer, K.M., Gillespie, J.J., Ober, K.A., 2007. Opinions of multiple sequence alignment and an empirical comparison of repeatability and accuracy between POY and structural alignment. Syst. Biol. 56, 133–146.
Kober, K.M., Nichols, S.A., 2007. On the phylogenetic relationships of hadromerid and poecilosclerid sponges. J. Mar. Biol. Ass. UK 87, 1585–1598.
Lartillot, N., Philippe, H., 2008. Improvement of molecular phylogenetic inference and the phylogeny of Bilateria. Philos. Trans. R. Soc. B 363, 1463–1472.
Leontis, N.B., Westhof, E., 2003. Analysis of RNA motifs. Curr. Opin. Struct. Biol. 13, 300–308.
Lockyer, A.E., Olson, P.D., Littlewood, D.T.J., 2003. Utility of complete large and small subunit rRNA genes in resolving the phylogeny of the Neodermata (Platyhelminthes): implications and a review of the carcomer theory. Biol. J. Linn. Soc. 78, 155–171.
Luan, Y.X., Mallatt, J.M., Xie, R.D., Yang, Y.M., Yin, W.Y., 2005. The phylogenetic positions of three basal–hexapod groups (Protura, Diplura, and Collembola) based on ribosomal RNA gene sequences. Mol. Biol. Evol. 22, 1579–1592.
Mallatt, J., Giribet, G., 2006. Further use of nearly complete 28S and 18S rRNA genes to classify Ecdysozoa: 37 more arthropods and a kinorhynch. Mol. Phylogenet. Evol. 40, 772–794.
Mallatt, J., Sullivan, J., 1998. 28S and 18S rDNA sequences support the monophyly of lampreys and hagfishes. Mol. Biol. Evol. 15, 1706–1718.
Mallatt, J., Winchell, C.J., 2002. Testing the new animal phylogeny: first use of combined large-subunit and small-subunit rRNA gene sequences to classify the protostomes. Mol. Biol. Evol. 19, 289–301.
Mallatt, J., Winchell, C.J., 2007. Ribosomal RNA genes and deuterostome phylogeny revisited: more cyclostomes, elasmobranchs, reptiles, and a brittle star. Mol. Phylogenet. Evol. 43, 1005–1022.
Mallatt, J.M., Garey, J.R., Shultz, J.W., 2004. Ecdysozoan phylogeny and Bayesian inference: first use of nearly complete 28S and 18S rRNA gene sequences to classify the arthropods and their kin. Mol. Phylogenet. Evol. 31, 178–191.
Mallatt, J., Sullivan, J., Winchell, C.J., 2001. The relationships of lampreys to hagfishes: a spectral analysis of ribosomal DNA sequences. In: Ahlberg, P.E. (Ed.), Major Events in Early Vertebrate Evolution: Palaeontology, Phylogeny, and Development. Taylor and Francis, London, pp. 106–118.
Marlétaz, F., Martin, E., Perez, Y., Papillon, D., Caubit, X., Lowe, C.J., Freeman, B., Fasano, L., Dossat, C., Wincker, P., Weissenbach, J., Le Parco, Y., 2006. Chaetognath phylogenomics: a protostome with a deuterostome-like development. Curr. Biol. 16, R577–R578.
Matus, D.Q., Copley, R.R., Dunn, C.W., Hejnol, A., Eccleston, H., Halanych, K.M., Martindale, M.Q., Telford, M.J., 2006. Broad taxon and gene sampling indicate that chaetognaths are protostomes. Curr. Biol. 16, R575–R576.
Mayrose, I., Friedman, N., Pupko, T., 2005. A Gamma mixture model better accounts for among-site rate heterogeneity. Bioinformatics 21 (Suppl. 2), ii151–ii158.
Medina, M., Collins, A.G., Silberman, J.D., Sogin, M.L., 2001. Evaluating hypotheses of basal animal phylogeny using complete sequences of large and small subunit rRNA. Proc. Natl. Acad. Sci. USA 98, 9707–9712.
Morrison, D.A., 2007. Increasing the efficiency of searches for the maximum likelihood trees in a phylogenetic analysis of up to 150 nucleotide sequences. Syst. Biol. 56, 988–1010.
Nawrocki, E.P., Kolbe, D.L., Eddy, S.R., 2009. Infernal 1.0: Inference of RNA alignments. Bioinformatics 25, 1335–1337.
Noller, H.F., 2005. RNA structure: reading the ribosome. Science 309, 1508–1514.
Paps, J., Baguñà, J., Riutort, M., 2009. Lophotrochozoa internal phylogeny: new insight from an up-to-date analysis of nuclear ribosomal genes. Proc. R. Soc. B. 276, 1245–1254.
Passamaneck, Y., Halanych, K.M., 2006. Lophotrochozoan phylogeny assessed with LSU and SSU data: evidence of lophophorate polyphyly. Mol. Phylogenet. Evol. 40, 20–28.
Passamaneck, Y.J., Schander, C., Halanych, K.M., 2004. Investigation of molluscan phylogeny using large-subunit and small-subunit nuclear rRNA sequences. Mol. Phylogenet. Evol. 32, 25–38.
Peterson, K.J., Eernisse, D.J., 2001. Animal phylogeny and the ancestry of bilaterians: inferences from morphology and 18S rDNA gene sequences. Evol. Dev. 3, 170–205.
Philippe, H., Telford, M.A., 2006. Large-scale sequencing and the new animal phylogeny. Trends Ecol. Evol. 21, 614–620.
Philippe, H., Brinkmann, H., Martinez, P., Riutort, M., Baguñà, J., 2007. Acoel flatworms are not Platyhelminthes: evidence from phylogenomics. PLoS ONE 2(8), e717, 1–4.
Phillipe, H., Derille, R., et al., 2009. Phylogenomics revives traditional views on deep animal relationships. Curr. Biol. 19, 706–712.
Pimentel, R.A., 1967. Invertebrate Identification Manual. Van Nostrand Reinhold, New York.
Posada, D., Buckley, T., 2004. Model selection and model averaging in phylogenetics: advantages of Akaike information criterion and Bayesian approaches over likelihood ratio tests. Syst. Biol. 53, 793–808.
Putnam, N.H. et al., 2008. The amphioxus genome and the evolution of the chordate genotype. Nature 453, 1064–1071.
R Development Core Team, 2008. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. Available from: <http://www.R-project.org/>.
Rix, M.G., Harvey, M.S., Roberts, J.D., 2008. Molecular phylogenetics of the spider family Micropholcommatidae (Arachnida: Araneae) using nuclear rRNA genes (18S and 28S). Mol. Phylogenet. Evol. 46, 1031–1048.
Rogozin, I.B., Wolf, Y.I., Carmel, L., Koonin, E.V., 2007. Ecdysozoan clade rejected by genome-wide analysis of rare amino acid replacements. Mol. Biol. Evol. 25, 1080–1090.
Ruiz-Trillo, I., Paps, J., Loukota, M., Ribera, C., Jondelius, U., Baguñà, J., Riutort, M., 2002. A phylogenetic analysis of myosin heavy chain type II sequences corroborates that Acoela and Nemertodermatida are basal bilaterians. Proc. Natl. Acad. Sci. USA 99, 11246–11251.
Ruiz-Trillo, I., Riutort, M., Fourcade, H.M., Baguñà, J., Boore, J.L., 2004. Mitochondrial genome data support the basal position of Acoelomorpha and the polyphyly of the Platyhelminthes. Mol. Phylogenet. Evol. 33, 321–332.
Ruiz-Trillo, I., Riutort, M., Littlewood, D.T.J., Herniou, E.A., Baguñà, J., 1999. Acoel flatworms: earliest extant bilaterian metazoans, not members of the Platyhelminthes. Science 283, 1919–1923.
Ruppert, E.E., 2005. Key characters uniting hemichordates and chordates: homologies or homplasies? Can. J. Zool. 83, 8–23.
Ruppert, E.E., Fox, R.S., Barnes, R.D., 2004. Invertebrate Zoology, A Functional Evolutionary Approach, seventh ed. Thomson, Brooks/Cole, Belmont, CA.
Savill, N.J., Hoyle, D.C., Higgs, P.G., 2001. RNA sequence evolution with secondary structure constraints: comparison of substitution rate models using maximum-likelihood methods. Genetics 157, 399–411.
Schierwater, B., Eitel, M., Jakob, W., Osigus, H., Hadrys, H.-J., Dellaporta, S.L., Kolokotronis, S.-O., DeSalle, R., 2009. Concatenated analysis sheds light on early metazoan evolution and fuels a modern "urmetazoan" hypothesis. PLoS Biol. 7, 36–43.
Schnare, M.N., Damberger, S.H., Gray, M.W., Gutell, R.R., 1996. Comprehensive comparison of structural characteristics in eukaryotic cytoplasmic large subunit (23 S-like) ribosomal RNA. J. Mol. Evol. 256, 701–719.
Schmidt-Rhaesa, A., 1996. Ultrastructure of the anterior end in three ontogenetic stages of *Nectonema minudae* (Nematomorpha). Acta Zool. 77, 267–278.
Sempere, L.F., Martinez, P., Cole, C., Baguñà, J., Peterson, K.J., 2007. Phylogenetic distribution of microRNAs supports the basal position of acoel flatworms and the polyphyly of Platyhelminthes. Evol. Dev. 9, 409–415.
Shi, X., Gu, H., Susko, E., Field, C., 2005. The comparison of confidence regions in phylogeny. Mol. Biol. Evol. 22, 2285–2296.
Shimodaira, H., 2002. An approximately unbiased test of phylogenetic tree selection. Syst. Biol. 51, 492–508.
Shimodaira, H., Hasegawa, M., 2001. CONSEL: for assessing the confidence of phylogenetic tree selection. Bioinformatics 17, 1246–1247.
Smith, S.A., Dunn, C.W., 2008. Phyutility: a phyloinformatics tool for trees, alignments, and molecular data. Bioinformatics 24, 715–716.
Srivastava, M. et al., 2008. The *Trichoplax* genome and the nature of placozoans. Nature 454, 955–960.
Stamatakis, A., 2008. The RAxML 7.0.0 Manual. Available from: <http://icwww.epfl.ch/~stamatak/index-Dateien/Page433.htm/>.
Stamatakis, A., Hoover, P., Rougemont, J., 2008. A rapid bootstrap algorithm for the RAxML web servers. Syst. Biol. 75, 758–771.
Struck, T.H., Fisse, F., 2008. Phylogenetic position of nemertea derived from phylogenomic data. Mol. Biol. Evol. 25, 728–736.
Struck, T.H., Schult, N., Kusen, T., Hickman, E., Bleidorn, C., McHugh, D., Halanych, K.M., 2007. Annelid phylogeny and the status of Sipuncula and Echiura. BMC Evol. Biol. 2007, 7:57.
Swofford, D.L., Olsen, G.J., Waddell, P.J., Hillis, D.M., 1996. Phylogenetic inference. In: Hillis, D.M., Moritz, C., Mable, B.K. (Eds.), Molecular Systematics, second ed. Sinauer Associates, Sunderland, pp. 407–514.
Syed, T., Schierwater, B., 2002. *Trichoplax adhaerens*: a phylum discovered, forgotten, and rediscovered. Vie Milieu 52, 177–187.
Telford, M.J., Bourlat, S.J., Economou, A., Papillon, D., Rota-Stabelli, O., 2008. The evolution of the Ecdysozoa. Philos. Trans. R. Soc. B 363, 1529–1538.
Telford, M.J., Lockyer, A.E., Cartwright-Finch, C., Littlewood, D.T.J., 2003. Combined large and small subunit ribosomal phylogenies support a basal position of the acoelomorph flatworms. Proc. R. Soc. Lond. B 270, 1077–1083.
Telford, M.J., Wise, M.J., Gowri-Shankar, V., 2005. Consideration of RNA secondary structure improves likelihood-based estimates of phylogeny: examples from the Bilateria. Mol. Biol. Evol. 22, 1129–1136.
Thorley, J.L., Wilkinson, M., 1999. Testing the phylogenetic stability of early tetrapods. J. Theor. Biol. 200, 343–344.
Voigt, O., Collins, A.G., Buchsbaum Pearse, V., Pearse, J.S., Ender, A., Hadrys, H., Schierwater, B., 2004. Placozoa—no longer a phylum of one. Curr. Biol. 14, R944–R945.
von Reumont, B.M., Meusemann, K., Szucsich, N.U., Dell 'Ampio, E., Gowri-Shankar, V., Bartel, D., Simon, S., Letsch, H.O., Stocsits, R.R., Luan, Y-x., Waegele, J.W., Pass, G., Hadrys, H., Misof, B., 2009. Can comprehensive background knowledge be incorporated into substitution models to improve phylogenetic analyses? A case study on major arthropod relationships. BMC Evol. Biol. 9, 119.
Wallberg, A., Curini-Galletti, M., Ahmadzadeh, A., Jondelius, U., 2007. Dismissal of Acoelomorpha: Acoela and Nemertodermatida are separate early bilaterian clades. Zool. Scripta 36, 509–523.
Wallberg, A., Thollesson, M., Farris, J.S., Jondelius, U., 2004. The phylogenetic position of the comb jellies (Ctenophora) and the importance of taxonomic sampling. Cladistics 20, 558–578.
Ware, J.L., Ho, S.Y.W., Kjer, K., 2008. Divergence dates of libelluloid dragonflies (Odonata: Anisoptera) estimated from rRNA using paired-sites substitution models. Mol. Phylogenet. Evol. 47, 426–432.
Warren, W.C. et al., 2008. Genome analysis of the platypus reveals unique signatures of evolution. Nature 453, 175–183.

Wessa, P., 2008. Free Statistics Software, Office of Research Development and Education, version 1.1.23-r2. Available from: <http://www.wessa.net/>.
Wheeler, W.C., Honeycutt, R.L., 1988. Paired sequence difference in ribosomal RNAs: evolutionary and phylogenetic implications. Mol. Biol. Evol 5, 90–96.
Wilkinson, M., 2006. Identifying stable reference taxa for phylogenetic nomenclature. Zool. Scripta 35, 109–112.
Winchell, C.J., Martin, A.P., Mallatt, J., 2004. Phylogeny of elasmobranches based on LSU and SSU ribosomal RNA genes. Mol. Phylogenet. Evol. 31, 214–224.
Winchell, C.J., Sullivan, J., Cameron, C.B., Swalla, B.J., Mallatt, J., 2002. Evaluating hypotheses of deuterostome phylogeny and chordate evolution with new LSU and SSU ribosomal DNA data. Mol. Biol. Evol. 19, 762–776.
Winnepinninckx, B., Backeljau, T., Mackey, L., Brooks, J.M., De Wachter, R., Kumar, S., Garey, J.R., 1995. 18S rRNA data indicate that Aschelminthes are polyphyletic in origin and consist of at least three distinct clades. Mol. Biol. Evol. 12, 1132–1137.
Zheng, J., Rogozin, I.B., Koonin, E.V., Przytycka, T.M., 2007. Support for the Coelomata clade of animals from a rigorous analysis of the pattern of intron conservation. Mol. Biol. Evol. 24, 2583–2592.
Zrzavy, J., Hypsa, V., Tietz, D., 2001. Myzostomida are not annelids: molecular and morphological support for a clade of animals with anterior sperm flagella. Cladistics 17, 1–29.
Zwickl, D.J., 2006. Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion. Ph.D. Dissertation. The University of Texas at Austin.