Ecdysozoan phylogeny and Bayesian inference: first use of nearly complete 28S and 18S rRNA gene sequences to classify the arthropods and their kin

Jon M. Mallatt, a,* James R. Garey, b and Jeffrey W. Shultz c

a School of Biological Sciences, Washington State University, Pullman, WA 99164-4236, USA
b Department of Biology, University of South Florida, 4202 East Fowler Ave. SCA110, Tampa, FL 33620, USA
c Department of Entomology, University of Maryland, College Park, MD 20742, USA

Received 4 March 2003; revised 18 July 2003

Abstract

Relationships among the ecdysozoans, or molting animals, have been difficult to resolve. Here, we use nearly complete 28S + 18S ribosomal RNA gene sequences to estimate the relations of 35 ecdysozoan taxa, including newly obtained 28S sequences from 25 of these. The tree-building algorithms were likelihood-based Bayesian inference and minimum-evolution analysis of LogDet-transformed distances, and hypotheses were tested with parametric bootstrapping. Better taxonomic resolution and recovery of established taxa were obtained here, especially with Bayesian inference, than in previous parsimony-based studies that used 18S rRNA sequences (or 18S plus small parts of 28S). In our gene trees, priapulan worms represent the basal ecdysozoans, followed by nematomorphs, or nematomorphs plus nematodes, followed by Panarthropoda. Panarthropoda was monophyletic with high support, although the relationships among its three phyla (arthropods, onychophorans, tardigrades) remain uncertain. The four groups of arthropods—hexapods (insects and related forms), crustaceans, chelicerates (spiders, scorpions, horseshoe crabs), and myriapods (centipedes, millipedes, and relatives)—formed two well-supported clades: Hexapoda in a paraphyletic crustacea (Pancrustacea), and Chelicerata + Myriapoda (a clade that we name Paradoxopoda). Pycnogonids (sea spiders) were either chelicerates or part of the chelicerate + myriapod clade, but not basal arthropods. Certain clades derived from morphological taxonomy, such as Mandibulata, Atelocerata, Schizoramia, Maxillopoda and Cycloneuralia, are inconsistent with these rRNA data. The 28S gene contained more signal than the 18S gene, and contributed to the improved phylogenetic resolution. Our findings are similar to those obtained from mitochondrial and nuclear (e.g., elongation factor, RNA polymerase, Hox) protein-encoding genes, and should revive interest in using rRNA genes to study arthropod and ecdysozoan relationships.

2003 Elsevier Inc. All rights reserved.

Keywords: Large-subunit ribosomal RNA; 28S rRNA; Small-subunit ribosomal RNA; Phylogeny; Protostome; Ecdysozoa; Arthropod

1. Introduction

The Ecdysozoa concept says that all phyla of animals that grow by molting a cuticular exoskeleton are united in a monophyletic clade (Ecdysozoa = Arthropoda, Tardigrada, Onychophora, Nematoda, Nematomorpha, Priapulida, Kinorhyncha, and Loricifera). Originally derived from a phylogenetic study using 18S rRNA genes (Aguinaldo et al., 1997), this controversial concept has been supported by subsequent molecular studies based on several genes (de Rosa et al., 1999; Giribet, 2002; Manuel et al., 2000; but see Hausdorf, 2000). However, many morphologists do not accept it (Brusca et al., 2003; Nielsen, 2001; Scholtz, 2002; Wägele and Misof, 2001; but see Peterson and Eernisse, 2001). We accept the Ecdysozoa because we recently obtained strong support for it from 28S rRNA genes (Mallatt and Winchell, 2002). The present paper explores the relationships among the different ecdysozoan lineages, which have also generated controversy (Giribet et al., 2000; Schmidt-Rhaesa et al., 1998). For example, some
morphological studies unite priapulans, kinorhynchs, loriciferans, nematodes, and nematomorphs into a monophyletic Cycloneuralia (Schmidt-Rhaesa et al., 1998) or Introverta (Nielsen, 1995), but some 18S-based studies indicate this group is paraphyletic (Garey, 2001; Peterson and Eernisse, 2001). Other controversies involve the relationships within Panarthropoda (Arthropoda + Onychophora + Tardigrada) (Dewel and Dewel, 1997; Giribet, 2002; Peterson and Eernisse, 2001), and especially, relationships within Arthropoda. Fig. 1 shows four major hypotheses of the interrelationships of the arthropod subgroups: hexapods, myriapods, crustaceans, and chelicerates (also see Giribet and Ribera, 2000). Finally, the position of pycnogonids within arthropods is uncertain (Blaxter, 2001).

The goal of this study was to sequence nearly complete 28S and 18S rRNA genes from a large sample of ecdysozoans in order to resolve their relationships. This was pursued because of encouraging preliminary results (Mallatt and Winchell, 2002), and because rRNA seems to contain more signal than other genes used for higher-order animal phylogeny (Giribet, 2002).

Analysis of rRNA gene sequences from 35 ecdysozoan and eight outgroup taxa using likelihood-based Bayesian inference and minimum evolution of LogDet-transformed distances provided important phylogenetic insights. These include the recovery of a basal position for Priapulida (and presumably for the related Kinorhyncha and Loricifera), the monophyly and internal relationships of Nematoda, paraphyly of Cycloneuralia, the monophyly of Panarthropoda and of Arthropoda, and a sister-group relationship between Pancrustacea (= Crustacea + Hexapoda) and a lineage comprising Chelicera and Myriapoda. These results emphasize the value of adding 28S sequences to 18S sequences for phylogenetic reconstruction and suggest that purported weaknesses of rRNA genes are attributable to inadequate sequence lengths and sub-optimal methods of phylogeny reconstruction used in the past. Bayesian inference performed extremely well on this data set, recovering established clades that are not obtained by other tree-building algorithms.

2. Materials and methods

2.1. Specimens and sequences

The names of the taxa sampled are shown in Figs. 2–4, whereas more detailed information, including voucher-specimen numbers and GenBank-accession numbers (AY210803-45), is presented in the Supplementary material (S1). Among the 35 ecdysozoans sampled, the large-subunit rRNA genes (28S + 5.8S) of 25 species were newly sequenced here. The taxa provide broad representation within Ecdysozoa, and comprise two priapulans, two nematomorphs, five nematodes, one tardigrade, two onychophorans, five chelicerates, four myriapods, eight crustaceans, and six hexapods. No kinorhynchs or loriciferans were available. Eight non-ecdysozoan bilaterians were used as outgroups, and provide a wide range of conserved lophotrochozoan and deuterostome sequences. DNA extraction, primers, PCR amplification, DNA purification, and sequence and fragment assembly followed protocols described in past studies (Mallatt and Sullivan, 1998; Winchell et al., 2002). The gene sequences were imported into SeqLab (Smith et al., 1994). There, the concatenated 28S and 18S genes were aligned by eye, with the alignment rigidly based on the large-subunit rRNA secondary-structure model of the frog Xenopus laevis (Schnare et al., 1996) and the small-subunit models of X. laevis and the sea urchin Strongylocentrotus purpuratus (Guttell, 1994). For the phylogenetic analyses, we used only the 28S and 18S genes, excluding the small 5.8S gene because it could not be sequenced across all ecdysozoan groups.

The 28S gene has a conserved core and 12 variable divergent domains (Hassouna et al., 1984; Hillis and Dixon, 1991; Mallatt et al., 2001). As in our past studies of deep-level phylogeny, we excluded the divergent domains and used only the core. For 18S, about 13% of all sites in the gene could not be aligned and were excluded. Overall, in the analyses we used 2313 and 1544 aligned sites from the 28S and 18S genes, respectively, for a total of 3857 sites. The alignment is available at http://chuma.cas.usf.edu/~garey/alignments/alignment.html.
2.2. Phylogenetic analyses

Phylogenetic algorithms that model the processes of sequence evolution, such as nucleotide substitution rates and among-site rate heterogeneity, are often preferred over algorithms that are not explicitly model-based (Bollback, 2002; Douady et al., 2003; Huelsenbeck et al., 2001b; Sullivan and Swofford, 2001). Here, we used two model-based algorithms for tree reconstruction: minimum evolution using LogDet-Paralinear distances (ME-LogDet) and Bayesian inference based on the likelihood function. These algorithms were executed using PAUP* 4.0 beta 10 (Swofford, 2001) or MrBayes 2.01 (Huelsenbeck and Ronquist, 2001).

LogDet is a model-based distance method for estimating phylogenetic trees (Lake, 1994; Lockhart et al., 1994; Swofford et al., 1996) that can be used with the ME algorithm. The ME-LogDet method models nucleotide-substitution rates better than other methods do and is designed to perform best when nucleotide frequencies are not stationary across taxa. By contrast, LogDet distances do not account for among-site rate heterogeneity. PAUP* crudely handles such heterogeneity by dividing the gene into two parts: a proportion of sites that cannot vary, called $P_I$; and all other sites, which are assumed to have evolved at the same rate. Following our previous studies, we chose a value of $P_I$ (0.45) slightly lower than the proportion of sites that was constant across all 43 taxa ($\approx 0.5$) (see Mallatt and Winchell, 2002; Winchell et al., 2002). This choice should be acceptable because trial analyses showed that all $P_I$ values from 0.3-0.5 produced the same tree topology and similar bootstrap-support values. After the $P_I$ was chosen, optimal ME-LogDet trees were calculated, and nonparametric bootstrap analyses were performed with 1000 replicates.

Bayesian inference based on the likelihood function (Huelsenbeck et al., 2001a,b; Larget and Simon, 1999) was used instead of the better-known maximum likelihood procedure (ML: Swofford et al., 1996) because it computes more efficiently and could analyze our 43 se-
quences in a reasonable amount of time (Hall, 2001; Huelsenbeck et al., 2002; Leach and Reeder, 2002). However, it is more susceptible than ML to the error of assigning high confidence to short, incorrect nodes (Alfaro et al., 2003; Suzuki et al., 2002), so we used ML on subsets of sequences to check some of the Bayesian results (see below).

When running Bayesian analyses in MrBayes 2.01, we specified a general-time-reversible (GTR) model of nucleotide substitution with a proportion of invariant sites (I) and a gamma (Γ) distribution of among-site rate heterogeneity with four rate categories (see Mallatt and Sullivan, 1998; and Mallatt and Winchell, 2002). No initial values were assigned to the model parameters, and empirical nucleotide frequencies were used. Four Markov chains were run for a million generations and sampled every 100 generations to yield a posterior probability distribution of 10,000 trees. After eliminating the first 2000 trees as “burn-in,” we constructed a 50% majority-rule consensus tree, with nodal values representing the probability (‘posterior probability’) that the recovered clades exist, given the aligned sequence data. We accepted a clade in the Bayesian tree at around 92% posterior probability (Murphy et al., 2001; Wilcox et al., 2002), while accepting values around 60–70% in the nonparametric bootstrap trees (Hillis and Bull, 1993).

A weakness of Bayesian inference, and of other algorithms based on the likelihood function, is that they assume stationarity of nucleotide frequency across all taxa in the data set, so their trees can be inaccurate if nonstationarity exists (Omilian and Taylor, 2001).

Finally, for comparison with past studies, we constructed trees using unweighted maximum parsimony (MP) (Swofford et al., 1996), including nonparametric bootstrapping with 1000 replicates. MP is not explicitly model-based and its resolving capacity may be compromised by a high susceptibility to long-branch attraction artifacts (Felsenstein, 1978; Lewis, 2001; Swofford et al., 2001).

2.3 Statistical tests of alternate hypotheses

Parametric bootstrapping based on ML (Efron, 1985; Huelsenbeck et al., 1996) was used to test if previous, alternate hypotheses of arthropod relations are consistent with present findings. Parametric bootstrapping can be extremely powerful and makes fewer simplifying assumptions than do other techniques of hypothesis testing. However, it is heavily reliant on the accuracy of the
evolutionary model it uses to simulate sequence data (Goldman et al., 2000) and it may reject hypotheses too readily (Antezana, 2003).

To accommodate the high computational demands of ML-based parametric bootstrapping, we trimmed the data set to 26 conserved sequences by removing divergent sequences. Care was taken to assure that the retained sequences represented all major ecdysozoan groups and yielded a topology similar to that of the 43-taxon tree. Initially, the optimal, unconstrained, ML tree was estimated from the 26-taxon data, using an iterative search strategy (Sullivan et al., 1997), in which the \( \text{GTR} + \Gamma \) model was found to fit the data best. After this, the actual parametric bootstrap tests were performed, involving 100 simulations per hypothesis tested, as described elsewhere (Mallatt and Sullivan, 1998; Wilcox et al., 2002). Data were simulated with the Seq-Gen v1.2.3 program (Rambaut and Grassly, 2001).

3. Results

3.1. Unusual sequences and nucleotide frequencies

As the 43 gene sequences were being aligned, two things became evident: The rRNA genes of the two nematomorphs, Gordius and Chordodes, are extremely similar, and those of the symphylan myriapod, Hanseniella, are highly divergent. The similarity between the two nematomorphs was so high (>98% for their entire 28S genes, 95% for their entire 18S genes) that it is doubtful whether they can be considered two distinct taxa in the phylogenetic analysis. The Hanseniella sequence is so different from all others that the first 350 nucleotides at the 5' end of its 18S gene were almost not alignable with the other sequences (but the rest of this 18S gene, and the 28S gene, were alignable); this divergence is reflected in the long branches for Hanseniella in the subsequent phylogenetic trees.\(^1\)

1 The 18S sequence we obtained for Hanseniella was >1970 nucleotides long, whereas that obtained by Giribet and Wheeler (2001), from the same Hanseniella specimens, was only 1350 nucleotides. We cannot fully explain this discrepancy, but stand by our results. We obtained this sequence, without difficulty, by our usual method of amplifying the entire gene with two universal end-primers for 18S (5' primer: CTGGTTGTATCCGCCAGT; 3' primer: TATGATCCCTCCGAGGTTCACCT). Furthermore, the shared parts of our and Giribet and Wheeler’s sequences (AY210823 and AF173237) are fully identical in nucleotide sequence. Perhaps Giribet and Wheeler, who amplify their 18S genes in several pieces instead of one, did not amplify the ultra-divergent 5' end in this case?
When the $\chi^2$ test of stationarity of nucleotide frequencies was applied to the 28S+18S data, the frequencies were found to be nonstationary across the 43 taxa ($\chi^2 = 336.5; P < 0.000000005$: see Supplementary material, S2). The 18S sequences were stationary, but the 28S were not, so the nonstationarity must arise from the 28S genes.

Given the nonstationarity of the entire data set, we identified the largest subset of sequences that was both stationary for nucleotide frequencies and biologically informative. To this end, we first removed Chordodes because it was nearly identical to Gordius. Then we removed the five divergent sequences with the most atypical proportions of nucleotides (Supplementary material, S2): C-rich Hanseniella, AT-rich Drosophila and Aedes; T-rich Caenorhabditis, and CG-rich Peripatus. The remaining 37 sequences were stationary ($\chi^2 = 124; P = 0.139$).

### 3.2. Combined 28S + 18S trees

Figs. 2 and 3 show the Bayesian and ME-LogDet trees, respectively, for all 43 taxa and also for the stationary, 37-taxon subset. The 37-taxon trees are included because they best fit the assumptions of both the Bayesian method (for stationarity) and the ME-LogDet method (for the fewest divergent sequences likely to have experienced extreme evolutionary-rate heterogeneity among sites). The 43-taxon set is included because it is more complete, even though it violates stationarity and contains divergent “long-branched” sequences from Hanseniella, Drosophila, Aedes, Caenorhabditis, and Peripatus (branch lengths are indicated in Fig. 3 and in Supplementary material, S3).

Although not identical, the four trees in Figs. 2 and 3 share many similarities:
1. Priapulans are basal to the other Ecdysozoa, and Cycloneuralia (priapulans, nematomorphs, and nematodes) are paraphyletic.
2. All nematodes go together, with one subgroup consisting of Trichinella and Xiphinema and the other containing Ascaris and Meloidogyne (and Caenorhabditis).
3. The nematodes go with the nematomorphs, although support for this ‘Nematoida’ node did not reach the level of significance in both Bayesian trees (e.g., 79% in Fig. 2B).
4. All panarthropods go together.
5. Chelicerates go with millipede and centipede myriapods.
6. The euchelicerates go together (scorpion *Pandinus*, xiphosuran *Limulus*, spider *Aphonopelma*).
7. Pycnogonids, *Callipallen* and *Colossendeis*, are in the ‘chelicerate + myriapod’ clade.
8. The Pancrustacea (hexapods and crustaceans) are united.
9. In Pancrustacea, the branchiopods go together (*Artemia*, *Daphnia*, *Triops*), as do the malacostracans (*Squilla* and *Pandalus*), but the maxillopods do not (*Argulus*, *Cyprididae* sp., *Cyclopidae* sp.).
10. The collembolean (*Podura*) goes with other hexapods.

On the other hand, there are several differences between the Bayesian and ME-LogDet trees, such as slightly different placements of pycnogonids within the ‘chelicerate + myriapod’ clade and different arrangements of traditional crustacean taxa in Pancrustacea. The biggest difference, however, is that ME-LogDet puts the long-branched sequences of the myriapod *Hanseniella*, and of the hexapods *Drosophila* and *Aedes*, outside the myriapod and hexapod clades, respectively (Fig. 3A), but Bayesian inference puts these sequences in their traditional clades (Fig. 2A). We checked the Bayesian result by individually adding *Hanseniella* and *Drosophila* to a 26-taxon subset of conserved sequences that was small enough for ML to handle (taxa listed at end of Table 1), and found that ML likewise put these two divergent sequences in their traditional clades.

### 3.3. 28S tree versus 18S tree

This section compares a tree derived from the 28S sequences alone with a tree from the 18S sequences.
(see Fig. 4, with their branch lengths shown in Supplementary material, S4–S5). Because Bayesian inference seems to place long-branch sequences more accurately than does ME-LogDet (see above), we show only the Bayesian trees here. In this section, tree topology is considered more important than high nodal support because the goal is to find which of the two gene trees most closely resembles the combined 28S + 18S tree. Both the 28S (Fig. 4A) and 18S (Fig. 4B) trees are roughly similar to the combined-gene tree (Fig. 2A) and to one another, sharing many nodes. For example, both 28S and 18S trees show strong support for priapulans as basal ecdysozoans, for paraphyly of Cycloneuralia, for two groups of nematodes (Trichinella and Xiphinema versus Ascaris, Meloidogyne and Caenorhabditis), and for monophyletic euchelicerates, Pancrustacea, branchiopods, and malacostracans. On the other hand, the 28S and 18S trees differ in the following ecdysozoan clades (Fig. 4: D1–D12):

D1. Nematomorphs. In the 28S tree, the nematomorphs (Chordodes and Gordius) are basal to nematodes and panarthropods, but in the 18S tree they group with the nematodes.

D2. Nematodes. Within the nematodes, the relative positions of Ascaris, Meloidogyne, and Caenorhabditis differ.

D3. Strong Panarthropoda. The 28S tree has more support for Panarthropoda (100% versus 61%).

D4. Pycnogonids. The 28S tree places the pycnogonids (Callipallen and Colossendeis) in the ‘chelicerate + myriapod’ clade, but the 18S tree shows them as distinct from euchelicerates and from a ‘myriapod + pancrustacean’ line.

D5. Myriapods. The 28S tree groups myriapods with chelicerates, but the 18S tree places centipede and millipede myriapods as the sister to Pancrustacea.

D6. Hanseniella’s 28S sequence goes with the myriapods, but its 18S sequence forms a distinct lineage in Panarthropoda.

D7. In crustaceans, 28S places Argulus at the base of the Pancrustacea, whereas 18S puts the branchiopods in this basal position.

D8. In branchiopods, 28S joins Triops with Daphnia, but 18S joins Triops with Artemia.

D9. Cyclopidae sp. 28S groups this copepod with hexapods, 18S puts it with malacostracans.

D10. Podura. The 28S sequence of Podura groups with the other hexapod sequences, but its 18S sequence falls outside the hexapods.

D11. 28S fails to unite Gromphadorhina and Tenebrio, but 18S does unite these pterygote insects.

D12. Diptera hexapods. Although neither gene correctly places the dipterans Drosophila and Aedes with the other pterygote insects Gromphadorhina and Tenebrio, the 18S gene joins them with silverfish Ctenolepisma, an apterygote.

The above list indicates that the 28S gene has the stronger influence on the combined 28S + 18S tree. That is, for eight of the 12 differences (D3–D10), the 28S condition is seen in the combined tree of Fig. 2A. By contrast, the 18S condition dominates in only four cases (D1–D2, D11–D12).

3.4. Maximum parsimony tree

Maximum-parsimony analysis of the 28S + 18S sequences produced trees that were more like the ME-LogDet trees than the Bayesian trees, so parsimony-bootstrap values are included on the ME-LogDet trees in Fig. 3. The MP bootstrap-consensus trees (not shown) resembled the ME trees in misplacing the divergent Aedes, Drosophila and Hanseniella sequences, but were unique in that they did not unite chelicerates with myriapods, nor centipedes with the millipede.

3.5. Hypothesis testing

Parametric-bootstrapping of a 26-taxon subset of the 28S+18S sequences tested four alternative hypotheses of arthropod interrelationships (Table 1). As planned, the 26 sequences in this subset showed stationarity of nucleotide frequencies ($\chi^2 = 68.51$; $P = 0.69$) and they produced an ML tree (shown in Supplementary material, S6) that is similar to the Bayesian tree derived from the larger, stationary set of 37 sequences (Fig. 2B) in that it recovers the same major groups of arthropods. Three of the alternate hypotheses tested in Table 1 have received support from a whole-evidence study (Giribet et al., 2001): hypothesis 1. The basic groups of arthropods are Mandibulata and Chelicera, with Mandibulata consisting of Pancrustacea and Myriapoda (Fig. 1C); hypothesis 2. Chelicerates + myriapods do not form a clade; hypothesis 3. Pycnogonids are not chelicerates as traditionally claimed, but instead are basal arthropods. Finally, we tested the traditional hypothesis 4 that pycnogonids are basal chelicerates (Wolszczek and Dunlop, 2002). Results show that our rRNA data are consistent only with #4, pycnogonids as basal chelicerates ($\delta = 1.3$, $P = 0.06$), and are highly inconsistent with the other three hypotheses (in which $\delta$ values were $> 7.8$ and $P$ was $<< 0.01$). The time-consuming computations of parametric bootstrapping kept us from using it to test other, more dissimilar, hypotheses of arthropod relations. However, we found that both Classical Mandibulata (Fig. 1A: $\delta = 184$) and Atelocerata + Schizoramia (Fig. 1B: $\delta = 202$) were strongly rejected ($P < 0.01$) by the SH test, which is far more conservative than parametric bootstrapping (Shimodaira and Hasegawa, 1999; Goldman et al., 2000).
4. Discussion

4.1. General comparison to previous studies

Most previous analyses of ecdysozoan phylogeny using 18S genes, with or without a small part of 28S, used MP or the parsimony-compatible ‘direct optimization method’ (Giribet and Ribera, 2000; Peterson and Eernisse, 2001; Spears and Abele, 1997; Wheeler, 1997). These analyses often yielded low resolution or questionable results, such as paraphyly of established clades and terminal taxa in unlikely positions, including failure to show the monophyly of Panarthropoda or Arthropoda, recovery of onychophorans and tardigrades within other ecdysozoan phyla, paraphyletic myriapoda in which centipedes did not unite with millipedes, placement of collembolans outside Hexapoda, placement of highly divergent sequences (e.g., symphylans, dipterans) outside their traditional groups, and uncertain and variable placement of the nematode clade (the latter reviewed by Garey, 2001). Consequently, some authorities expressed discouragement at the performance of rRNA genes (e.g., Giribet and Ribera, 2000; Spears and Abele, 1997), especially when the situation was not improved by adding more 18S sequences. Confirming these shortcomings, the 18S MP tree for our own 43 taxa also showed low resolution (Supplementary material, S7).

The shortcomings noted above have been largely eliminated from our combined 28S + 18S analysis. Although our taxon sampling remains low relative to the diversity (even hyper-diversity) of the represented groups, the present Bayesian trees are better resolved (Fig. 2) than trees from earlier studies and show little paraphyly of accepted clades. Specifically, our analyses recovered monophyletic Panarthropoda, Arthropoda, Myriapoda and Hexapoda, and a monophyletic Nematoda that fit cleanly in the Ecdysozoa. True resolving power is also suggested by the fact that ours is the only approach that makes no assumptions about the nature of ecdysozoan rRNA sequences. Nevertheless, some day a LogDet method that does model rate heterogeneity may resolve the divergent dipteran and symphylan sequences in more reasonable positions than did ME-LogDet. Thus, for our data set, the strength of Bayesian inference in accounting for rate heterogeneity among sites in a gene seems to overshadow any weakness caused by nucleotide nonstationarity and overconfidence of the method. ME-LogDet misplaced the divergent sequences, apparently due to its limited ability to model rate heterogeneity. However, a test of this interpretation using bootstrap-sampling or the s gambit, a LogDet method that does model rate

4.2. Comparison of ME-LogDet and Bayesian methods

Ecdysozoan sequences were analyzed here using both ME-LogDet and Bayesian inference based on the likelihood function (GTR + I + Γ model). Both of these methods have acknowledged weaknesses (see Section 2), and initially it was unknown which one would produce the most accurate trees from our data. We conclude that Bayesian inference is preferable because it placed the divergent dipteran and symphylan sequences in more reasonable positions than did ME-LogDet. Thus, for our data set, the strength of Bayesian inference in accounting for rate heterogeneity among sites in a gene seems to overshadow any weakness caused by nucleotide nonstationarity and overconfidence of the method. ME-LogDet misplaced the divergent sequences, apparently due to its limited ability to model rate heterogeneity. However, a test of this interpretation using bootstrap-sampling or the s gambit, a LogDet method that does model rate
heterogeneity (Lake, 1998), yielded mixed results: Gambit correctly put dipterans in insects but still incorrectly joined the ultra-divergent *Hanseniella* with onychophorans (results not shown).

It should be noted that in our earlier studies, which examined much higher-order relationships among animals, likelihood-based methods seemed to provide worse phylogenetic resolution than the ME-LogDet method (Mallatt and Winchell, 2002; Winchell et al., 2002), which is the opposite of what we found here within the Ecdysozoa. This situation may reflect the sparser taxon sampling of the earlier studies (often just one rRNA sequence per phylum), which may have provided too little information to the likelihood algorithm.

4.3. Comparison of 28S and 18S trees

The trees derived from the separate analyses of 28S and 18S sequences have roughly similar topologies (Fig. 4), but we regard the 28S gene as contributing more phylogenetic signal. Specifically, 28S alone puts the collembolan *Podura* with the other hexapods and also puts the divergent *Hanseniella* and dipterans in more reasonable places (compare Figs. 4A and 4B). Also, as explained in Section 3.3, the 28S gene has more influence on the topology of the combined-gene tree than does the 18S gene.

These relative contributions of 28S versus 18S are consistent with the findings of previous studies. Most prior 28S + 18S studies dealt with higher-order clades than used here, namely metazoans and bilaterians at the “phylum”-to-“superphylum” level, and found 18S to contribute more signal than 28S (Mallatt and Winchell, 2002; Medina et al., 2001; Winchell et al., 2002). By contrast, a study of lower-level phylogeny, within the “class” of cartilaginous fishes, found that nearly all the signal came from the 28S gene (Winchell, 2001). The present study covered an intermediate taxonomic level, from “phyla” to “classes” within Ecdysozoa, and found 28S to provide just slightly more signal than 18S. Thus, a trend is evident: In proceeding from high to low taxonomic levels, the 28S gene contributes more signal and the 18S gene less.

4.4. Comparison of clade-specific findings with those of previous studies

4.4.1. Cycloneuralia

The clade Cycloneuralia (Ehlers et al., 1996; Schmidt-Rhaesa et al., 1998) or Introverta (Nielsen, 1995) was proposed to include Scalidophora (Priapulida + Kinorhyncha + Loricifera) and Nematoida (Nematoda + Nematomorpha). The morphological characters used to unite this clade were a circumpharyngeal cerebral ring that is not ganglionated anterodorsally, details of brain histology, and perhaps an introvert organ. Our findings are inconsistent with the Cycloneuralia concept because they place priapulans basal to all other ecdysozoans (Figs. 2 and 3). The other scalidophorans were not sampled here but should be united with priapulans: Indeed, in his analysis of conserved 18S sequences, Garey (2001) recovered kinorhynchs and priapulans as the basal, monophyletic group in Ecdysozoa. If Cycloneuralia are paraphyletic, their shared anatomical characters must either be primitive or convergently evolved.

4.4.2. Nematoida

Schmidt-Rhaesa (1998) identified morphological characters that appear to unite the nematodes and nematomorphs into the taxon Nematoida. These synapomorphies include aflagellate sperm, a cloaca in both sexes, and similarities in the ventral and dorsal epidermal and nerve cords. Most 18S-based studies have not recovered Nematoida (Giribet et al., 2000; Peterson and Eernisse, 2001), but it was supported in Garey’s (2001) analysis of conserved 18S sequences. The present analysis of 28S + 18S sequences also recovered a Nematoida, but this support was weak in that it lost significance when one of the two nearly identical nematomorph sequences was omitted (Fig. 2B). Future studies should include the other nematomorph clade, the presumably primitive marine *Nectonema* (Bleidorn et al., 2002; Schmidt-Rhaesa, 2002).

4.4.3. Relationships among nematodes

We used only five species of nematodes, so we cannot say much about relationships within this large phylum. As seen in Figs. 2 and 3, our trees show two groups, *Xiphinema + Trichinella* versus *Ascaris + Caenorhabditis + Meloidogyne*, with the relationships within the latter group unresolved. These two groups are consistent both with morphology-based systematics of these genera (adenophorans versus secernenteans: Blaxter et al., 1998, their Fig. 2b) and with a tree based on 18S (Blaxter et al., 1998, their Fig. 2a).

4.4.4. Panarthropoda

A panarthropod clade consisting of onychophorans, tardigrades and arthropods is widely accepted, based on such shared characters as a segmental body and ventralateral appendages (Schmidt-Rhaesa et al., 1998). Although this clade was never well supported by 18S, it was strongly supported by 28S + 18S here. However, the relations among the three panarthropodan phyla are uncertain (Dewel and Dewel, 1997; Peterson and Eernisse, 2001), and all possible pairings have been advocated (Eriksson et al., 2003; Giribet, 2002; Ramsköld and Chen, 1998; Schmidt-Rhaesa et al., 1998). Our findings are also inconclusive, but hint at ‘onychophorans + tardigrade’ (Fig. 2). Adding more tardigrade sequences might clarify these relations, but the divergent
4.4.5. Arthropod subphyla

The phylogenetic relationships of the major arthropod groups (Chelicerata, Crustacea, Myriapoda and Hexapoda) are intensely debated. Prior to recent molecular studies, most neontologists favored ‘Mandibulata + Chelicerata’ (Fig. 1A), which unites hexapods, myriapods, and crustaceans to the exclusion of chelicerates (Snodgrass, 1938). By contrast, most paleontologists favor Atelocerata (= Chelicerata + Myriapoda, Cisne, 1974; Colgan et al., 1998; Haas et al., 2003; Wills et al., 1998). Both these hypotheses originally included the monophyly of ‘Hexapoda + Myriapoda,’ but molecular studies now strongly suggest that Hexapoda and Crustacea form a monophyletic group, the Pancrustacea (= Tetraconata: Dohle, 2001) (Giribet et al., 2001; Hwang et al., 2001; Regier and Shultz, 2001; Zrzavy and Stys, 1997; see Giribet and Ribera, 2000, p. 222 for other references). Correspondingly, certain morphological characters are shared by hexapods and crustaceans, especially in the nervous system (Mittmann, 2002; Richter, 2002; Wilson et al., 2000). Some studies suggest that Pancrustacea is the sister group to Myriapoda within Mandibulata (Fig. 1C) (Giribet et al., 2001), but others indicate that Pancrustacea is sister to a clade comprising Chelicerata and Myriapoda (Cook et al., 2001; Hwang et al., 2001). Our results support this ‘chelicerate + myriapod’ hypothesis (Figs. 1D and 2), and reject Mandibulata + Chelicerata and Atelocerata + Schizorormia. Even the Pancrustacea + Myriapoda version of Mandibulata (Fig. 1C), which is the least inconsistent with our trees, was overwhelmingly rejected by parametric bootstrapping (Table 1).

The hypothesis that chelicerates and myriapods form a monophyletic sister group to Pancrustacea was never supported by morphology-based studies and it derives strictly from molecular studies. Friedrich and Tautz (1995) were early supporters, based on their analysis of partial-28S/18S sequences from ten arthropods (also see Turbeville et al., 1991; and Giribet et al., 1996), but the best evidence for ‘chelicerate + myriapods’ has come from Hox- and mitochondrial gene sequences (Cook et al., 2001; Hwang et al., 2001). Given our new rRNA evidence that this clade might actually exist, we name it ‘Paradoxopoda,’ which reflects the conflict between the molecular data that support it and the morphological data that do not.

Overall, our findings resemble those of Regier and Shultz (2001), who explored arthropod interrelationships with the nuclear genes for RNA polymerase II, elongation factor-1α, and elongation factor 2. The number of taxa they used (19 panarthropods) and the number of nucleotide sites in their combined-gene sequences (~4000) are similar to what we used here (26 panarthropods, ~3800 nucleotides). As indicated by the ML-bootstrap values in their three-gene summary table (their p. 142), they found monophyly of arthropods (>87%); of myriapods including a symphyllan (94–99%); of the chelicerates including pycnogonids (75–86%); of Pancrustacea (95–97%); and less strongly, of hexapods (including collemboians) (66–87%). Their only finding that differed from ours was that the ‘chelicerate + myriapod’ clade was never supported.

4.4.6. Chelicerates

Pycnogonids are usually considered chelicerates, but their position is controversial because they have many autapomorphic structures (Blaxter, 2001; Brusca and Brusca, 1990). However, pycnogonids primitively have chelicerae, a key synapomorphy of chelicerates, and Waloszek and Dunlop (2002) argued in detail that their body plan fits that of chelicerates. By contrast, Giribet et al. (2001) concluded that pycnogonids are at the base of all extant arthropods, based on a synthesis of characters from 18S, other genes, and morphology (this placement is also consistent with our 18S tree in Fig. 4B). Our combined 28S + 18S data indicate that pycnogonids are not basal arthropods but are either basal chelicerates or a distinct lineage in the ‘chelicerate + myriapod’ clade (Figs. 2 and 3; Table 1).

4.4.7. Myriapoda

Our findings favor the traditional view of myriapod monophyly (Edgecombe and Giribet, 2002; Giribet et al., 2001; Regier and Shultz, 2001; Wills et al., 1998) over myriapod paraphyly (Kraus, 2001; Shear, 1997) because our trees separate myriapods from hexapods (Fig. 2). Myriapods are the group that would benefit most from more rRNA sequences.

4.4.8. Pancrustacea

The monophyly of Pancrustacea was strongly supported by our 28S and 18S sequences (Figs. 2–4). Among the traditional crustaceans, our rRNA trees indicated that branchiopods (Artemia, Daphnia, and Triops) and eumalacostracans (Squilla and Panulirus) are each monophyletic, a result also indicated by 18S and other genes (Shultz and Regier, 2000; Spears and Abele, 1997). Maxillopoda, by contrast, is problematical. This clade, said to consist of copepods, ostracods,branchiurans, cirripeds, and mystacocarids, was united by such morphological characters as body segmentation and mouth parts (Walossek and Müller, 1998; Wills et al., 1997). However, it was never universally accepted by systematists (see Spears and Abele, 1997) and was never supported by molecular studies (Giribet and Ribera, 2000; Regier and Shultz, 2001; Shultz and Regier, 2000), including the present study, in which the cyclopod
copepod, cypridid ostracod, and branchiuran Argulus do not group as a unit (Figs. 2–4).

Our study, like all previous rRNA-based studies, indicates that hexapods nest within a paraphyletic “crustacea,” which raises the question of which crustacean line is the sister group to Hexapoda (Friedrich and Tautz, 2001; Schram and Jenner, 2001). Although other studies favor either branchiopods (Gaunt and Miles, 2002; Regier and Shultz, 1997) or malacostracans (Hwang et al., 2001; Wilson et al., 2000), our Bayesian trees (Fig. 2) show a copepod as the sister of hexapods. However, the copepod sequence is rather divergent and copepods have never been regarded as relevant to understanding the origin of hexapods.

Finally, our 28S+18S data unite the collemobolan Podura with the other hexapods (Fig. 2), supporting the traditional idea of hexapod monophyly (cf. Nardi et al., 2003).

5. Conclusions

Overall, our findings suggest that recent pessimism about the ability of rRNA sequences to resolve ecdysozoan interrelationships can be replaced with optimism. The use of likelihood-based Bayesian inference on nearly complete 28S+18S sequences may prove to be the long-awaited combination of best genes (Giribet, 2002) and best tree-building method for reconstructing ecdysozoan phylogenies with the fewest long-branch-attraction artifacts. This implies that adding more taxa will finally start to improve, rather than diminish, phylogenetic resolution. It is not meant to imply, however, that rRNA alone is sufficient. Instead, many genes must be used (Brown et al., 2001; Takezaki et al., 2003; Sanderson et al., 2003). To this end, the rRNA sequences could be concatenated with sequences of nuclear protein-coding and mitochondrial genes, which should produce even better taxonomic resolution.

6. Summary

In this study, the use of nearly complete 28S+18S rRNA sequences and of Bayesian inference based on a GTR+I+F likelihood model produced a well-resolved tree of 35 arthropod and ecdysozoan sequences. More specifically, our trees support the following findings (Fig. 2).

1. Priapulans (and presumably the related kinorhynchs and loriciferans) represent the basal line of ecdysozoans.
2. The five nematodes used here form two subgroups consistent with accepted clades.
3. Arthropoda and Panarthropoda are each monophyletic, although the interrelationships of the panarthropod phyla (tardigrades, onychophorans, and arthropods) remain uncertain.
5. Euchelicerates, Hexapoda, ectognathic insects, and Myriapoda are each monophyletic.
6. Pycnogonids are either chelicerates, or part of the ‘chelicerate + myriapod’ clade.
7. Mandibulata, Maxillopoda, Atelocerata, Schizoramia, and Cycloneuralia are paraphyletic and invalid taxa.

Other major findings are:
1. In this data set, the 28S gene contains more phylogenetic signal than the 18S gene.
2. In this study, minimum-evolution analysis of LogDet distances and maximum parsimony produced less resolution than Bayesian inference.
3. Arthropod relationships derived here from rRNA genes resemble those calculated from concatenated elongation-factor and RNA polymerase II genes (Regier and Shultz, 2001).

Acknowledgments

Thanks are extended to Gary Thorgaard, Jack Sullivan, Kevin Pullen, and to all who provided specimens.

References


