Current Biology, Vol. 14, 1644–1649, September 21, 2004, ©2004 Elsevier Ltd. All rights reserved. DOI 10.1016/j.cub.2004.09.027

Bayesian Inference of the Metazoan Phylogeny: A Combined Molecular and Morphological Approach

Henrik Glenner,[1] Anders J. Hansen,[1] Martin V. Sørensen,[2] Frederik Ronquist,[3] John P. Huelsenbeck,[4] and Eske Willerslev[1,5,*]

[1]Department of Evolutionary Biology
Biological Institute
University of Copenhagen
Universitetsparken 15
DK-2100 Copenhagen
Denmark

[2]Department of Organismic and Evolutionary Biology
Harvard University
16 Divinity Avenue
Cambridge, Massachusetts 02138

[3]Evolutionary Biology Centre
Uppsala University
Nordbyvägen 18D
SE-752 36 Uppsala
Sweden

[4]Section of Ecology, Behavior, and Evolution
Division of Biological Sciences
University of California, San Diego
La Jolla, California 92093-0116

[5]Department of Zoology
University of Oxford
South Parks Road, Oxford OX1 3PS
United Kingdom

Summary

Metazoan phylogeny remains one of evolutionary biology's major unsolved problems. Molecular and morphological data, as well as different analytical approaches, have produced highly conflicting results due to homoplasy resulting from more than 570 million years of evolution [1–4]. To date, parsimony has been the only feasible combined approach but is highly sensitive to long-branch attraction [5–7]. Recent development of stochastic models for discrete morphological characters and computationally efficient methods for Bayesian inference has enabled combined molecular and morphological data analysis with rigorous statistical approaches less prone to such inconsistencies [8–10]. We present the first statistically founded analysis of a metazoan data set based on a combination of morphological and molecular data and compare the results with a traditional parsimony analysis. Interestingly, the Bayesian analyses demonstrate a high degree of congruence between morphological and molecular data, and both data sets contribute to the result of the combined analysis. Additionally, they resolve several irregularities obtained in previous studies and show high credibility values for controversial groups such as the ecdysozoans and lophotrochozoans. Parsimony, on the contrary, shows conflicting results, with morphology being congruent to the Bayesian results and the molecular data set producing peculiarities that are largely reflected in the combined analysis.

*Correspondence: eske.willerslev@zoology.oxford.ac.uk

Results and Discussion

Combining Morphological and Molecular Data

The metazoans (all multicellular animals) are believed to have diversified around the beginning of the Cambrian period (~543 million years ago); this diversification makes morphological and molecular homoplasy (convergence, parallelism, and reversals) a serious concern in the study of their evolutionary relationships [1–4]. Combining both sets of information in a single phylogenetic analysis makes it easier to recognize true homologies from homoplasy, the rationale being that characters representing very different levels of evolutionary change (nucleotide substitutions in a single gene and complex morphological changes possibly encoded by multiple gene complexes) are unlikely to point falsely in the same direction [4–6]. Parsimony programs for combined morphological and molecular data have been available for more than a decade [11]. However, compared to maximum likelihood, Bayesian inference, and other statistical methods of phylogenetic inference, parsimony is known to perform poorly when a data matrix includes distant taxa with different evolutionary rates and when long branches tend to draw them together by homoplasy (long-branch attraction) [12]. These are expected properties of a data set as diverse as the metazoans. Unfortunately, statistical methods for Bayesian inference of phylogeny have long been restricted to molecular data. Only recently has the theoretical basis for a realistic evolutionary model for morphological data been developed [8].

The aim of this study was to estimate the first metazoan phylogeny with combined molecular and morphological data in a purely statistical framework. The same morphological and molecular data sets, both separate and combined, were run with Bayesian and parsimony algorithms, allowing comparison of results.

The Bayesian analysis of phylogeny was performed on the combined data with the program MrBayes [9]. The morphological and molecular data could not be analyzed with the same model of evolution. Instead, we modeled them separately. For the sequence data, we assumed the GTR + I + G model of evolution. This model allows different rates of substitution types among the nucleotides as well as unequal nucleotide frequencies. We accounted for rate variation across sites by assuming that a proportion of the sites are invariable (cannot change) but that the remaining variable sites have a rate that is drawn from a γ distribution. The parameters of the model of DNA sequence evolution are all treated as random variables in the Bayesian analysis and are estimated during the course of the analysis. For the morphological characters we followed the approach described in [10]. In brief, this approach assumes that the evolution of the morphological characters follows a continuous-time Markov process with two states.

The likelihoods are calculated on the assumption that no invariable character patterns have been sampled. However, this approach does assume that all of the morphological characters share both a common phylogeny and a common set of branch lengths.

In the parsimony analyses, the most parsimonious trees were found with PAUP*. In all analyses, character states were unordered and weighted equally, and alignment gaps were coded as missing data. We evaluated the most parsimonious topology by bootstrap resampling with 500 replications.

The morphological data were primarily compiled from [13], which represents the most up-to-date morphological character matrix on metazoans, but the scoring was revised and additional characters were added according to new information (see the Supplemental Data available with this article online). A total of 94 morphological characters were included in the analysis. All characters were unordered, and only two-state characters were used.

To ensure consistency of a model-based approach and reduce the risk of saturation due to long-term separation of taxa, we downloaded 57 complete 18S rDNA sequences (likely to be the most slowly evolving molecular marker known for eukaryotes) from the small subunit (SSU) rRNA database, in which sequences are prealigned according to a model of secondary structure [14]. Although sequences from the 18S rDNA gene are available for representatives of nearly all metazoan phyla in GenBank, the number of sequences aligned according to secondary structure is still restricted. Consequently, sequences from newer and enigmatic taxa such as Loricifera and Micrognathozoa were not included in the analyses. However, employing a data set with an incomplete taxon sampling proved useful for comparison of robustness of results obtained by the Bayesian and parsimony methods.

Traditionally, the bilaterian metazoans are divided into three major groups according to the organization of body cavity: the acoelomate (platyhelminths and nemerteans), pseudocoelomate (nematodes, gastrotrichs, and rotifers), and coelomate (molluscs, chordates, annelids, and arthropods) animals. The largest and morphologically most diverse group, the coelomates, is further subdivided into two distinct lineages based on embryonic cell cleavage patterns: Protostomia, consisting of annelids, arthropods, and molluscs, and Deuterostomia, containing chordates and lophophorates (bryozoans, brachiopods, and phoronids). However, this hypothesis on the metazoan relationship was challenged by molecular analyses that found the lophophorates to be more closely related to the annelids and the molluscs than to the arthropods, creating the group Lophotrochozoa, which includes the last common ancestor of all annelids, molluscs, bryozoans, brachiopods, and phoronids [15]. Additionally, molecular analyses repeatedly find support for a group of molting but otherwise morphologically very dissimilar animals termed Ecdysozoa, which includes the arthropods and the nematodes [3, 6, 16–19].

The validity of these two new assemblages of metazoans has been greatly debated: that of Lophotrochozoa because the clade is difficult to identify from morphological characters [20], and that of Ecdysozoa because it questions the morphologically well-established com-

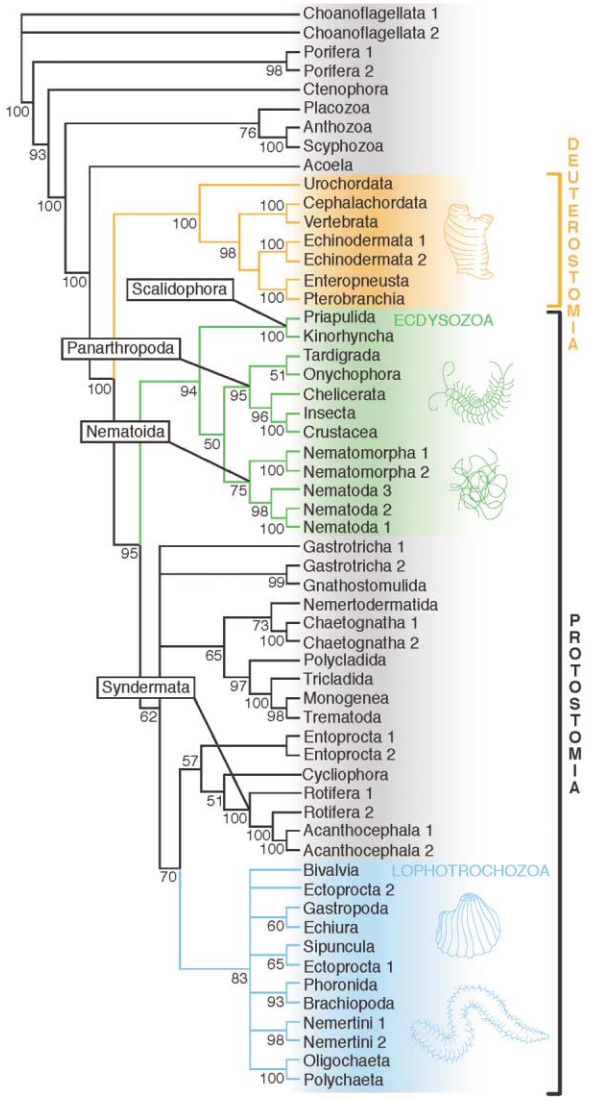

Figure 1. Combined Morphological and Molecular Data-Based Bayesian Inference of Metazoan Phylogeny

Groups discussed in the text are demarcated. Bilaterian animals are: Deuterostomia (orange), Ecdysozoa (green), and Lophotrochozoa (blue). Note that the platyhelminth taxon, Acoela, branches off the bilaterian clade before the major Deuterostomia/Protostomia split. Only clades with posterior probabilities greater than 50% are shown.

mon ancestry of arthropods and annelids [21]. However, the debate has mainly been due to morphological objections, and molecular studies provide compelling evidence for a lophotrochozoan and an ecdysozoan assemblage of animals. In particular, Ecdysozoa appears to be well defined, although the internal relationship of ecdysozoans remains unsettled. Some authors support a close relationship between the Panarthropoda (tardigrades, onychophorans, and arthropods) and Scalidophora (priapulids, kinorhynchs, and loriciferans) with Nematoida (Nematoda and Nematomorpha) as their sister group [20]. Alternatively, the Scalidophora and Nematoida are considered sister groups in Ecdysozoa and are united in the clade Cycloneuralia [4, 6, 17, 18].

The result of the combined Bayesian analysis is shown

in Figure 1 and supports a basal division of the bilaterian metazoans into two major lineages, the Deuterostomia and Protostomia, as defined in contemporary textbooks [22]. The protostomes form one large, weakly supported, monophyletic clade consisting of the gnathostomulids, chaetognaths, gastrotrichs, lophotrochozoans, rotifers, acanthocephalans, entoprocts, cycliophorans, and platyhelminths, with the exclusion of the acoels, branching off as the first bilaterian taxon. The paraphyly of platyhelminths and the basal position of Acoela are in agreement with the results of previous molecular studies [23, 24]. The remaining platyhelminths are grouped together in a monophyletic clade with the surprising but weakly supported inclusion of chaetognaths as a sister group to the platyhelminth group Nemertodermatida. The analysis places Acanthocephala within Rotifera. This relationship has been supported by several analyses based both on morphological and molecular data, and the Acanthocephala/Rotifera group is often referred to as Syndermata. Syndermata is a monophyletic group with Entoprocta and Cycliophora as basal branches. This grouping has earlier been found in morphological analyses [13] and has recently been confirmed in analyses with combined data [4, 17]. However, the molecular data set alone did not support this clade in the present study.

The combined Bayesian analysis further supports a clade of Lophotrochozoa; this clade would include the molluscs, annelids, brachiopods, bryozoans, echiurids, sipunculids, and nemerteans. The inclusion of the bryozoans and the exclusion of the entoprocts from the Lophotrochozoa are contrary to previous combined analyses, all of which support an inclusion of Entoprocta but not bryozoans in Lophotrochozoa [4, 25]. The combined analysis also supports an Ecdysozoa group with Panarthropoda and Scalidophora as sister groups and not the alternative cycloneuralian monophyly previously suggested [19]. Thus, the result of the combined Bayesian analysis fully supports Ecdysozoa and Lophotrochozoa as they were originally proposed [15, 26].

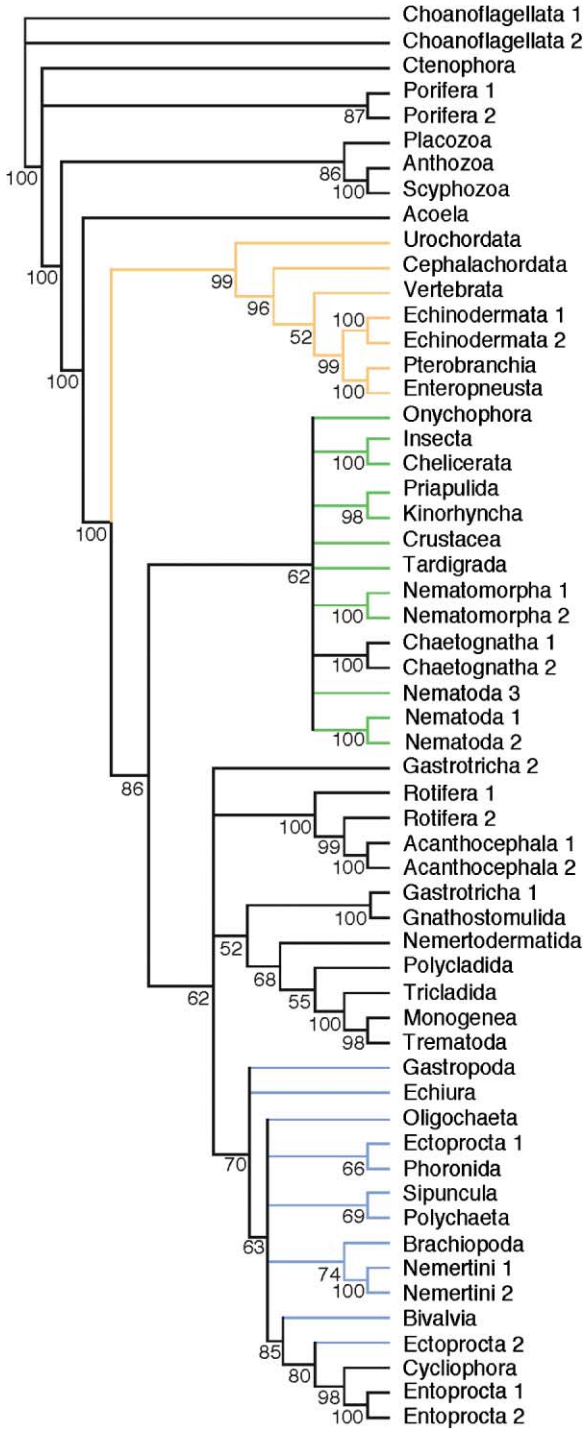

Figure 2. Bayesian Analysis of the Molecular Data

The Deuterostomia branch is marked with orange, and the Lophotrochozoa (including Entoprocta and Cycliophora) branch is blue. The Ecdysozoa branch (green) includes Chaetognatha, in contrast to the combined and morphological analyses.

Comparison of the Separate Bayesian Data Sets

The topology of the tree resulting from the Bayesian analysis of the molecular data (Figure 2) is more similar to that of the combined analysis (Figure 1) than is that of the morphology-based tree (Figure 3). The morphological analysis places Acoela among the other platyhelminths in the morphological data set, and the phoronids, brachiopods, and chaetognaths are affiliated with the Deuterostomia. Interestingly, neither the morphological nor the molecular data set supports a close relationship between Panarthropoda and Annelida. Instead, both analyses support a clade of ecdysozoans.

Generally, the deeper bilaterian nodes remain unresolved in the morphological tree (Figure 3) and are, therefore, largely affected by the molecular data set in the combined analysis (Figures 1 and 2). However, both data sets contribute to resolving the terminal branches in the combined analysis as seen in Ecdysozoa; this clade is supported both in the combined analysis and in the separate analyses of the two data sets. The separate analyses additionally include the gastrotrichs (morphology, Figure 3) or the chaetognaths (DNA, Figure 2). Combining the data results in the exclusion of both of these taxa from the ecdysozoan clade, demonstrating that morphological characters, even if they are massively outnumbered by molecular data, still bear a phyloge-

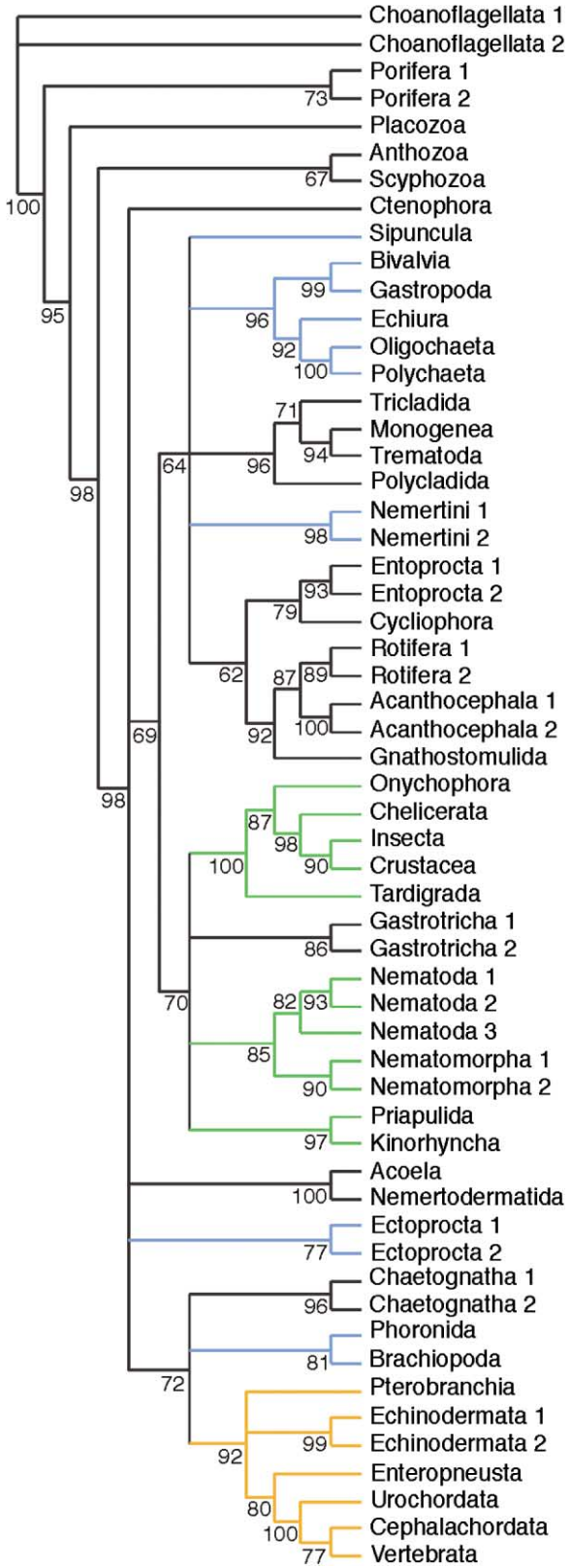

Figure 3. Bayesian Analysis of the Morphological Data Set

The bilaterian metazoans form an unresolved fork consisting of (1) the traditional assembly of Deuterostomids (orange), which, together with the phoronids, the Branchiopods, and the chaetognaths, form a monophyletic clade, (2) the Ectoprocts, (3) the two flatworms Acoela and Nemertodermatida, (4) the Ctenophora, and (5) the remaining protostomid metazoans. The ecdysozoans (green), including the gastrotrichs, form a monophyletic group, but only a group of core lophotrochozoans (blue) can be identified (Molluscs, Annelids, and Echiura).

netic signal strong enough to profoundly influence the topology of the tree.

Parsimony Analyses

Interestingly, the Bayesian and parsimony analyses of the morphological data set generate identical trees that, although not fully resolved, show congruency with the phylogenies from both the molecular and combined data sets analyzed with a Bayesian approach (Figures 1 and 2; see also Supplemental Data). In contrast, the parsimony analysis on the molecular data conjugates in a very distinct tree (see Supplemental Data), and it is the substantial impact of this signal that is reflected in the combined parsimony analysis (Figure 4). Although the combined parsimony analysis strongly supports the bilaterian animals as a monophyletic clade, the traditional and well-established division of the coelomate animals into Deuterostomia and Protostomia is not supported. A clade of lophotrochozoan animals, a modified ecdysozoan group including Panarthropoda, Priapulida, Kinorhyncha, and Chaetognatha, and a paraphyletic group of nematodes are identified, but none of them is supported by bootstrap values above 50 (Figure 4). The molecular and combined phylogenies are most likely resulting from the well-known statistical inconsistencies of the parsimony method [7].

Concluding Remarks

This study represents the first statistically based phylogenetic analysis on a combined molecular and morphological data set, but it will undoubtedly not be the last. The application of stochastic evolutionary models to discrete morphological characters has yet to be broadly accepted. However, it makes it possible to use powerful statistical methods to infer molecular phylogenies without disregarding the information supplied by traditional morphological characters. The results obtained in this pilot study on metazoan phylogeny are indeed encouraging and suggest that the combined statistical approach will become a valuable alternative to the traditional parsimony method.

Experimental Procedures

Bayesian Analyses

All Bayesian sequence analyses were initiated from random starting trees. In both the combined and the molecular analyses, the applied model for the molecular data, GTR + I + G, was favored by Akaike Information Criterion and hierarchical likelihood ratio testing as they are implemented in MrModeltest version 1.1b.

The morphological data were run under the datatype = standard option that activates the M2 model in MrBayes. The M2 model [9] was employed as implemented in MrBayes version 3.0b4 (http://morphbank.ebc.uu.se/mrbayes3), and likelihood was corrected for the scoring bias (only parsimony-informative characters were scored). All parameters were associated with diffuse priors.

Both the combined and the molecular data sets were run for 3,000,000 generations, and that for morphology was run for only

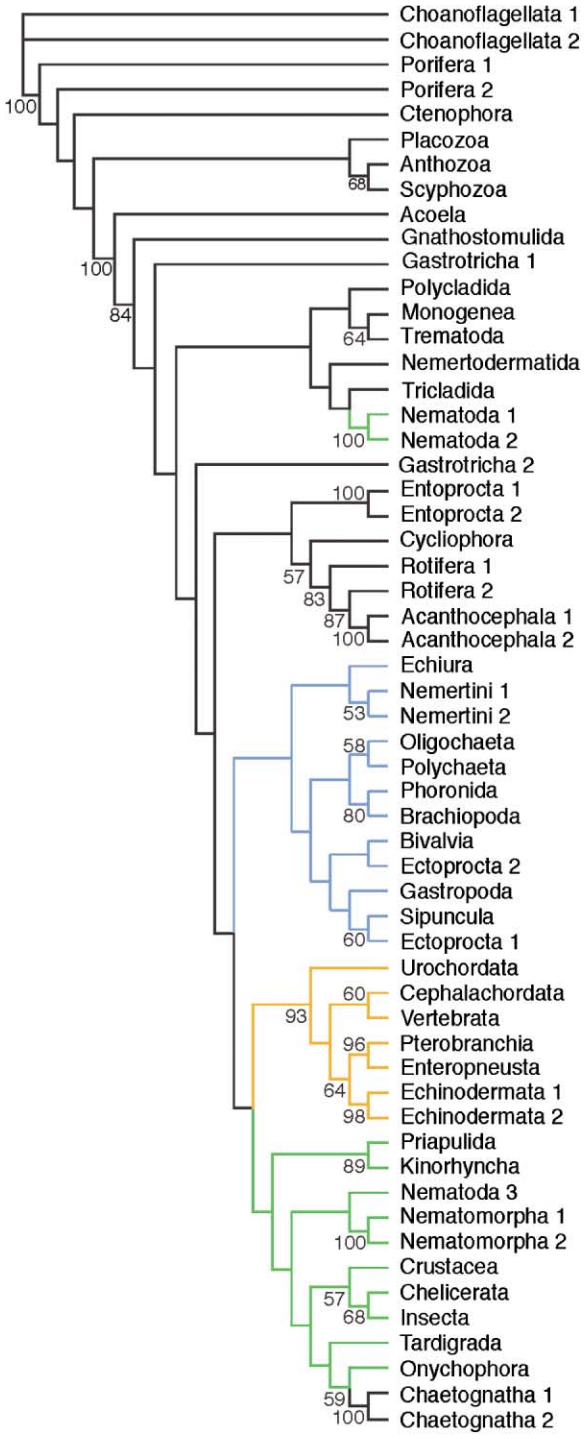

Figure 4. Parsimony Analysis of the Combined Data Set

The single most parsimonious tree based on the same combined data set that produced the Bayesian tree in Figure 1 is shown. The Deuterostomia branch is marked with orange, the Ecdysozoa branch including Chaetognatha in green, and the Lophotrochozoa branch in blue. Also note that the combined parsimony analyses place Acoela as the sister group to the remaining bilaterian animals. Bootstrap branch support is shown below the branches. Of 2860 characters, 1222 were parsimony informative. All characters were of type "unordered" and equally weighted. Gaps were treated as "missing." The tree length was 10,883 steps. The consistency index (CI) = 0.2579, and the homoplasy index (HI) = 0.7421 (only informative sites were included).

2,000,000 generations. A tree was sampled every 300 generations (morphology: 200), resulting in 10,000 trees. Chain stationarity was achieved after 1,200,000 generations (morphology: 200,000) (burn-in) and, therefore, 4,000 (morphology: 1,000) trees were subsequently discarded. To calculate the posterior probability of each analysis, we constructed a 50% majority rule consensus tree from these remaining trees, and the percentage of times a clade occurred among this sampling of trees was interpreted as its posterior probability. Three independently repeated analyses resulted in similar tree topologies and comparable clade probabilities and substitution model parameters, suggesting that reasonable estimates of the posterior probability distributions were obtained (see Supplemental Data). The combined Bayesian analysis also employed a rate multiplier (prset ratepr = variable;), allowing molecular and morphological evolution to proceed at different rates.

Parsimony Analyses

Maximum-parsimony (MP) analyses of aligned sequences were performed with PAUP* test version 4.0b10. Searches for most parsimonious trees were conducted with the heuristic maximum-parsimony algorithms with random stepwise sequence addition. For each search, 100 repetitions were performed. The most parsimonous topology was evaluated by bootstrap resampling [27] for 500 replicates, each with 100 random-addition heuristic searches.

Supplemental Data

A taxon list, a list of GenBank accession numbers, a list of characters used in the data matrix, and several supplemental figures are available at www.current-biology.com/cgi/content/full/14/18/1644/DC1/.

Acknowledgments

We are grateful to S.Y.W. Ho, L. Menez, J. Haile, and M.B. Hebsgaard for help and discussion. H.G. was supported by Danish National Science Research Council grant number 51-00-0457. E.W. was supported by the Henry Wellcome Trust Foundation, United Kingdom, the National Science Foundation, Denmark, and the Carlsberg Foundation, Denmark.

Received: April 1, 2004
Revised: July 28, 2004
Accepted: July 29, 2004
Published: September 21, 2004

References

1. Winnepenninckx, B., Backeljau, T., Mackey, L.Y., Brooks, J.M., Dewachter, R., Kumar, S., and Garey, J.R. (1995). Phylogeny of protostome worms derived from 18S ribosomal-RNA sequences. Mol. Biol. Evol. *12*, 641–649.
2. Littlewood, D.T.J., Telford, M.J., Clough, K.A., and Rohde, K. (1998). Gnathostomulida – An enigmatic metazoan phylum from both morphological and molecular perspectives. Mol. Phylogenet. Evol. *9*, 72–79.
3. Edgecombe, G.D., Wilson, G.D.F., Colgan, D.J., Gray, M.R., and Cassis, G. (2000). Arthropod cladistics: Combined analysis of histone H3 and U2 snRNA sequences and morphology. Cladistics *16*, 155–203.
4. Giribet, G., Distel, D.L., Polz, M., Sterrer, W., and Wheeler, W.C. (2000). Triploblastic relationships with emphasis on the acoelomates and the position of Gnathostomulida, Cycliophora, Platyhelminthes, and Chaetognatha: A combined approach of 18S rDNA sequences and morphology. Syst. Biol. *49*, 539–562.
5. Giribet, G. (2002). Current advances in the phylogenetic reconstruction of metazoan evolution. A new paradigm for the Cambrian explosion? Mol. Phylogenet. Evol. *24*, 345–357.
6. Peterson, K.J., and Eernisse, D.J. (2001). Animal phylogeny and the ancestry of bilaterians: Inferences from morphology and 18S rDNA gene sequences. Evol. Dev. *3*, 170–205.
7. Felsenstein, J. (1978). Cases where parsimony or compatibility methods will be positively misleading. Syst. Zool. *27*, 401–410.
8. Lewis, P.O. (2001). A likelihood approach to estimating phylog-

eny from discrete morphological character data. Syst. Biol. *50*, 913–925.

9. Huelsenbeck, J.P., and Ronquist, F. (2001). MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics *17*, 754–755.
10. Nylander, J.A.A., Ronquist, F., Huelsenbeck, J.P., and Nieves-Aldrey, J.L. (2004). Bayesian phylogenetic analysis of combined data. Syst. Biol. *53*, 47–67.
11. Felsenstein, J. (1989). PHYLIP–Phylogeny Inference Package (Version 3.2). Cladistics *5*, 164–166.
12. Pollock, D.D., and Bruno, W.J. (2000). Assessing an unknown evolutionary process: Effect of increasing site-specific knowledge through taxon addition. Mol. Biol. Evol. *17*, 1854–1858.
13. Sorensen, M.V., Funch, P., Willerslev, E., Hansen, A.J., and Olesen, J. (2000). On the phylogeny of the metazoa in the light of Cycliophora and Micrognathozoa. Zool. Anz. *239*, 297–318.
14. Van de Peer, Y., Caers, A., De Rijk, P., and De Wachter, R. (1998). Database on the structure of small ribosomal subunit RNA. Nucleic Acids Res. *26*, 179–182.
15. Halanych, K.M., Bacheller, J.D., Aguinaldo, A.M.A., Liva, S.M., Hillis, D.M., and Lake, J.A. (1995). Evidence from 18S ribosomal DNA that the lophoporates are protostome animals. Science *267*, 1641–1643.
16. Aguinaldo, A.M.A., Turbeville, J.M., Linford, L.S., Rivera, M.C., Garey, J.R., Raff, R.A., and Lake, J.A. (1997). Evidence for a clade of nematodes, arthropods and other moulting animals. Nature *387*, 489–493.
17. Schmidt-Rhaesa, A., Bartolomaeus, T., Lemburg, C., Ehlers, U., and Garey, J.R. (1998). The position of the Arthropoda in the phylogenetic system. J. Morphol. *238*, 263–285.
18. Zrzavy, J. (2003). Gastrotricha and metazoan phylogeny. Zool. Scr. *32*, 61–81.
19. Garey, J.R. (2001). Ecdysozoa: The relationship between Cycloneuralia and Panarthropoda. Zool. Anz. *240*, 321–330.
20. Nielsen, C. (2001). Animal Evolution: Interrelationships of the Living Phyla (New York: Oxford University Press).
21. Anderson, D.T. (1973). Embryology and Phylogeny in Annelids and Arthropods (Oxford: Pergamon Press).
22. Ruppert, E.E., Fox, R.S., and Barnes, R.D. (2004). Invertebrate Zoology: A Functional Evolutionary Approach, 7th Edition (Pacific Grove, CA: Brooks Cole Publishing Company).
23. Ruiz-Trillo, I., Riutort, M., Littlewood, D.T.J., Herniou, E.A., and Baguna, J. (1999). Acoel flatworms: Earliest extant bilaterian metazoans, not members of Platyhelminthes. Science *283*, 1919–1923.
24. Jenner, R.A. (2004). Towards a phylogeny of the Metazoa: Evaluating alternative phylogenetic positions of Platyhelminthes, Nemertea, and Gnathostomulida, with a critical reappraisal of cladistic characters. Contrib. Zool. *73*, 3–163.
25. Giribet, G., and Ribera, C. (2000). A review of arthropod phylogeny. New data based on ribosomal DNA sequences and direct character optimization. Cladistics *16*, 204–231.
26. Aguinaldo, A.M.A., and Lake, J.A. (1998). Evolution of the multicellular animals. Am. Zool. *38*, 878–887.
27. Swofford, D.L. (2003). PAUP 4.0 User's Manual: Phylogenic Analysis Using Parsimony (Sunderland, Massachusetts: Sinauer Associates Inc.).