

Bilaterian Phylogeny: A Broad Sampling of 13 Nuclear Genes Provides a New Lophotrochozoa Phylogeny and Supports a Paraphyletic Basal Acoelomorpha

Jordi Paps, Jaume Baguñà, and Marta Riutort

Departament de Genètica, Universitat de Barcelona, Barcelona, Spain

During the past decade, great progress has been made in clarifying the relationships among bilaterian animals. Studies based on a limited number of markers established new hypotheses such as the existence of three superclades (Deuterostomia, Ecdysozoa, and Lophotrochozoa) but left major questions unresolved. The data sets used to the present either bear few characters for many taxa (i.e., the ribosomal genes) or present many characters but lack many phyla (such as recent phylogenomic approaches) failing to provide definitive answers for all the regions of the bilaterian tree. We performed phylogenetic analyses using a molecular matrix with a high number of characters and bilaterian phyla. This data set is built from 13 genes (8,880 bp) belonging to 90 taxa from 27 bilaterian phyla. Probabilistic analyses robustly support the three superclades, the monophyly of Chordata, a spiralian clade including Brachiozoa, the basal position of a paraphyletic Acoelomorpha, and point to an ecdysozoan affiliation for Chaetognatha. This new phylogeny not only agrees with most classical molecular results but also provides new insights into the relationships between lophotrochozoans and challenges the results obtained using high-throughput strategies, highlighting the problems associated with the current trend to increase gene number rather than taxa.

Introduction

Small ribosomal subunit RNA gene (*18S* rDNA or small subunit [SSU]) sequences were the first and most extensively used source of information to establish the new, widely accepted bilaterian phylogeny, which features three large superclades: the Lophotrochozoa, the Ecdysozoa, and the Deuterostomia (Halanych 2004). However, the relationships within these superclades and the phylogenetic position of some enigmatic phyla still remain elusive to SSU analyses, in part due to long-branch attraction (LBA) artifacts (Felsenstein 1978) and to their recognized limited resolution (Abouheif et al. 1998). To overcome this problem, other markers such as the large ribosomal subunit RNA gene (*28S* or large subunit [LSU]; Mallatt and Giribet 2006; Passamaneck and Halanych 2006) or protein-coding genes (Ruiz-Trillo et al. 2002; Anderson et al. 2004; Peterson et al. 2004) were introduced. Although these approaches were instrumental in resolving some internal nodes of the tree, they are associated with similar problems to those encountered with SSU genes, namely, stochastic errors and artifacts due to LBA (Philippe and Telford 2006). Alternative sources of information such as sequence signatures in the Hox genes (Balavoine et al. 2002), mitogenomics (Boore et al. 2005), or micro-RNAs (Sempere et al. 2006) were subsequently proposed. Unfortunately, the binary nature of these qualitative characters (present/absent, shared/not shared) has only allowed definition of one clade versus another and has not helped to resolve all the internal relationships yet.

Phylogenomics, based mainly on expressed sequence tag (EST) data, is nowadays the leading approach through which to address this problem. Using up to 183 genes, together with the development of new models of protein evolution, recent phylogenomic studies have lent support to the three superclades (Philippe et al. 2005; Bourlat

et al. 2006; Delsuc et al. 2006; Lartillot et al. 2007; Brinkmann and Philippe 2008; Dunn et al. 2008) but have been unable to produce a clear and robust internal phylogeny of these clades. Phylogenomics claims to overcome stochastic errors by incorporating a high number of characters; however, it is also susceptible to “gappy” alignments (Hartmann and Vision 2008), poor taxon sampling, systematic errors, and paralogy problems (Philippe and Telford 2006). Indeed, reduced sampling might explain conflicting results either supporting or rejecting the Ecdysozoa over the Coelomata (Dopazo et al. 2004), as well as uncertainties related to the true position of the Acoela (Philippe et al. 2007) or the tunicates (Delsuc et al. 2006). Even the incorporation of one taxon per phylum does not guarantee a systematic error-free phylogeny (Philippe and Telford 2006), although the incorporation of new EST projects into future analyses will hopefully prevail over those errors as shown by recent studies (Dunn et al. 2008).

As of today, the molecular matrices used are asymmetric: They include either many phyla and few markers or not many phyla and numerous markers, each case bearing its own flaws. This produces dark areas in some regions of the bilaterian tree, for example, the internal relationships within the Lophotrochozoa, the monophyletic status of Chordata, and the position of groups like the Acoelomorpha and the Chaetognatha. To provide a more robust basis on which to analyze the phylogenetic relationships of these problematic regions, we developed and analyzed a more balanced data set. We evaluated 26 genes for their potential phylogenetic information, and 11 were selected. Sequences already present in GenBank were downloaded and 89 new sequences produced. The final matrix contains 90 representatives from 27 phyla and is 8,880 nt long for the 11 protein-coding genes in addition to the two ribosomal RNA genes with a value of 40% missing data. Maximum likelihood (ML) and Bayesian inference (BI) methods were used to obtain a phylogeny of the bilaterians.

Materials and Methods

Sampling

One hundred and twenty-five samples were collected for 96 species belonging to 31 phyla (see supplementary

Key words: Metazoa, Lophotrochozoa, Ecdysozoa, Deuterostomia, SSU, LSU, LBA, Maximum likelihood, Bayesian inference, molecular phylogeny, multigenic.

E-mail: mriutort@ub.edu.

Mol. Biol. Evol. 26(10):2397–2406. 2009

doi:10.1093/molbev/msp150

Advance Access publication July 14, 2009

table 1, Supplementary Material online). Some groups were not sampled due to their already rich representation in GenBank (such as Craniata, Nematoda, and Arthropoda) or because we had no access to them (Acantocephala, Micrognathozoa, Loricifera, Mesozoa, Pogonophora, and Myzostomida).

Molecular Techniques

RNA was extracted from live animals or preserved in RNAlater (Ambion) with TRIzol reagent (Amersham Pharmacia Biotech) and cDNA was obtained by standard reverse transcription with M-MLV reverse transcriptase (Promega). When the yield of RNA or cDNA was low, the SMART protocol (Invitrogen) was used to increase the number of cDNA copies by polymerase chain reaction (PCR) with adapters, according to the manufacturer's instructions (cDNAs obtained by SMART are noted in supplementary table 1, Supplementary Material online). Gene fragments were amplified by PCR: 25 μ l, with 1 unit of Dynazyme polymerase (Fynnzimes), 40 cycles of 45 s at 94 °C, 45 s at the annealing temperature for each primer pair (supplementary table 2, Supplementary Material online), and 55 s at 72 °C. PCR products were purified (Microcon PCR columns, Millipore) and directly cycle-sequenced from both strands (Big Dye Terminator V.2.0, Applied Biosystems). Sequence products were ethanol precipitated and run on an ABI Prism 3700 (Applied Biosystems) automated sequencer. Contigs were assembled with SeqEd VER. 1.0.3 (Applied Biosystems).

Gene Selection

The phylogenetic potential for 26 genes was evaluated because they had proven to be useful in previous phylogenetic studies, showed good phyla sampling in the genetic databases, or had interesting qualities regarding their rates of evolution (Ehlers et al. 1996; Mushegian et al. 1998; Anderson et al. 2004; Peterson et al. 2004). The genes assessed were as follows: 14-3-3, sodium–potassium ATPase alpha-subunit (*ATPase alpha*), cathepsin, cell division cycle 42, cAMP response element-binding, elongation factor alpha 1, elongation factor alpha 2, eukaryotic translation initiation factor 4E, forkhead, glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*), histone H3 (*H3*), intermediate filaments (*IFs*), laminin-binding protein, myosin heavy chain type II (Myosin), ribosomal protein L13, ribosomal Protein L22, tropomyosin, tubulin, aldolase, ATP synthase beta-chain, methionine adenosyltransferase (*MAT*), DNA helicase, kinesin, phosphofructokinase (*PFK*), catalase, and actin.

Gene sequences were downloaded from GenBank and each gene was aligned independently based on the amino acid sequence using ClustalX 1.81, and the resulting alignments were checked with Bioedit. For organisms which full genome is available (i.e., *Drosophila melanogaster*, *Homo sapiens*, etc.), only those markers with clear orthology were used. Regions of ambiguous alignment were removed using Gblocks (Castresana 2000) with the default options except Allowed gap positions (set to “With half”). For each gene,

taxa lacking representatives were amplified and sequenced, and a Blast search was performed with the new sequences to confirm their identity. They were added to their respective alignments and their orthology was assessed with single-gene phylogenies. Genes that produced poorly resolved phylogenies (comb trees) contained poor taxon sampling or produced trees that were highly inconsistent with previous phylogenetic studies (e.g., placing molluscs inside chordates) were discarded. The final selected genes were *ATPase alpha*, *GAPDH*, *H3*, *IFs*, *myosin*, *tropomyosin*, *aldolase*, *ATP synthase beta*, *MAT*, *PFK*, and *catalase*. Independent alignments for SSU and LSU sequences from a previous study (Paps et al. 2009) were also used. In order to have the same number of operational taxonomic unit (OTUs) for all the genes, the missing representatives for each gene were classified as missing data (filled with Ns).

Data Set

The independent alignments were concatenated into a data set containing 90 OTUs representing 27 phyla and 8,880 positions for 13 genes and a 40% missing data. A summary of the sequences included is provided in supplementary table 3, Supplementary Material online, and a more detailed description for each OTU (species, classification, number of genes available, and accession numbers) is shown in supplementary table 4, Supplementary Material online. To reduce the “gappyness” in the matrix (see Hartmann and Vision 2008), 20 of the 90 OTUs were produced by merging sequences of different species, an approach already used in other studies (Giribet et al. 2001; Bourlat et al. 2008). The merged sequences were from species as related as possible and always belonging to the same class. Only the Echiura representative is constituted by organisms from different classes. These amalgamated OTUs were named consequently (e.g., Oligochaeta, Opiliones, and Teleostei).

Phylogenetic Analyses

TreePuzzle was also employed to carry out the likelihood mapping analyses, with the options estimation accurate (slow), Tamura–Nei 93 model, 4 Gamma categories and 1 invariable and 10,000 quartets. Modeltest (Posada and Crandall 1998) was used to determine the evolutionary model that showed the best fit for each gene, following the Akaike information criterion. The specified model (general time-reversible [GTR] + Γ + I) was used in all the algorithms.

BI trees were inferred with a parallelized version of MrBayes software (Ronquist and Huelsenbeck 2003), using a partitioned data set (one partition for each gene, unlinking for each partition the estimation of Statefreq, revmat, Pinvar, and shape) and running 1,000,000 generations in 2 independent analyses with a sample frequency of 100. To obtain the consensus tree and BI supports, 500,000 generations were removed to discard trees sampled before likelihood values had reached a plateau. PhyloBayes analyses were performed with the CAT mixture model, which accounts for across-site heterogeneities in the amino acid replacement process (Lartillot and Philippe 2004). Two independent runs were performed with a total length of 17,000 cycles and the first 2,500 points were discarded

Table 1
Comparison of Topologies Using the Approximately Unbiased Test

Topology	AU
1. Original ML tree (fig. 1)	0.927
2. Acoelomorpha sister group to Ambulacraria (Philippe et al. 2007)	<0,001 ^a
3. Acoelomorpha sister group to Platyhelminthes (Rieger et al. 1991)	<0,001 ^a
4. <i>Xenoturbella</i> sister group to Nephrozoa ^b	0.294
5. Nematoda + Nematomorpha, sister group to Priapulida + Kinorhyncha (Dunn et al. 2008)	0.091
6. Nematoda + Nematomorpha, sister group to <i>Arthropoda</i> (Mallatt and Giribet 2006)	0.087
7. Chaetognatha sister group to Ecdysozoa (Zrzavy et al. 1998)	0.003 ^a
8. Chaetognatha sister group to Lophotrochozoa (Matus et al. 2006)	0.014 ^a
9. Chaetognatha sister group to Protostomia (Marletaz et al. 2006)	0.004 ^a
10. Bryozoa + Entoprocta, sister group to Rotifera + Acanthocephala ^b	0.001 ^a
11. Bryozoa + Entoprocta, sister group to spiralian clade (Hausdorf et al. 2007)	0.109

^a Hypothesis rejected when *P* values < 0.05 for the AU test. The studies in which the hypothesis is found or referenced are indicated.

^b Variations based on our trees. See text for discussion.

as burn-in, and the posterior consensus was computed on the remaining trees.

ML trees were inferred with RaxML (Stamatakis 2006), ran using the model GTR + Γ + *I* (4 gamma categories + 1 invariable) and using a partitioned data set (one partition for each gene). A random topology was used as starting tree and 1,000 bootstrap replicates were obtained with the “RapidBootstrap” algorithm. To test the results using an alternative heuristic search, PhyML 3.0.1 (Guindon and Gascuel 2003) was run using the model GTR + Γ + *I* (4 gamma categories + 1 invariable), 1,000 replicates obtained and subtree pruning regrafting (SPR) heuristic search was used.

Competing topologies were evaluated. The alternative trees were constructed using Treeview (Page 1996) using the original ML inference tree as a template. The alternative topologies tested were based on previous studies or were considered interesting to be tested (table 1). PAUP (Swofford 2000) was used to obtain the site likelihood for the ML tree and CONSEL (Shimodaira and Hasegawa 2001) was run to perform the approximately unbiased (AU) test. The analyses were run on four different computers: 1) two PCs running Windows XP and SUSE Linux 10.0, 2) a supercomputer located at CESCA (Centre de Supercomputació de Catalunya, <http://www.cesca.es>), and 3) the Marenostrom supercomputer located at the Barcelona Supercomputing Center (<http://www.BPPc.es>).

Results and Discussion

Methodological Problems and Data Set Information

The experimental work endured two bottlenecks causing the missing data in our matrix. The first were the

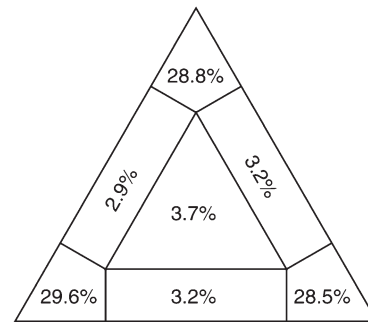


Fig. 1.—The Likelihood mapping analyses are represented as a triangle whose corner values indicate percentage of well-resolved phylogenies for all possible quartets, whereas central and lateral values are percentages of unresolved phylogenies. The cumulatively high percentage (86.9%) from the corner values indicates that the data set is phylogenetically informative.

unsuccessful RNA extractions, mainly from some marine tiny animals (Porifera, Placozoa, Myxozoa, Gnathostomulida, Cycliophora, or Gastrotricha), where a seemingly sufficient amount of tissue was available but the extraction did not yield enough quantity or quality for the following procedures. This explains the lack of some key phyla in our matrix, though these groups were collected several times. The second bottleneck was the PCR amplification, where some primer pairs worked successfully for some samples but not for others, although these very same samples amplified effectively for other pairs. In the end, in this study, 89 new sequences were produced for the 11 selected genes. A complete list of the species and sequences used in this study can be seen in supplementary table 4, Supplementary Material> online. The Likelihood mapping analysis (fig. 1) shows a high proportion of well-resolved quartets (87%), indicating that the data set is phylogenetically informative.

Resolution of the Bilaterian Phylogeny

ML trees (fig. 2) and BI (data not shown, supports in fig. 2) result in a bilaterian phylogeny that sheds light on some current uncertainties. The topology from PhyML using SPR heuristics (data not shown) completely agrees with the RaxML tree with few minor exceptions (see clades discussion). The trees obtained from Phylobayes (supplementary fig. 1, Supplementary Material online) lack convergence, but mostly agree with our original MrBayes and ML topologies; unfortunately, the two Phylobayes trees show some anomalies. The first tree places the Bryozoa (including the rotifer *Philodina*) as sister group to Priapulida within Ecdysozoa. The second run is not able to recover the monophyly of Deuterostomia and places Onychophora within a highly unresolved Lophotrochozoa. Those two anomalies are likely due to trapping in local optima or to the GTR + Γ + *I* model fitting better than CAT-GTR.

The other analyses agree in the general topology, reproducing the three main bilaterian superclades with some minor disagreements in a few internal nodes. The general lack of high support is also seen in other multigene studies including many taxa and specially rogue species (Dunn et al. 2008), probably reflecting these type of data

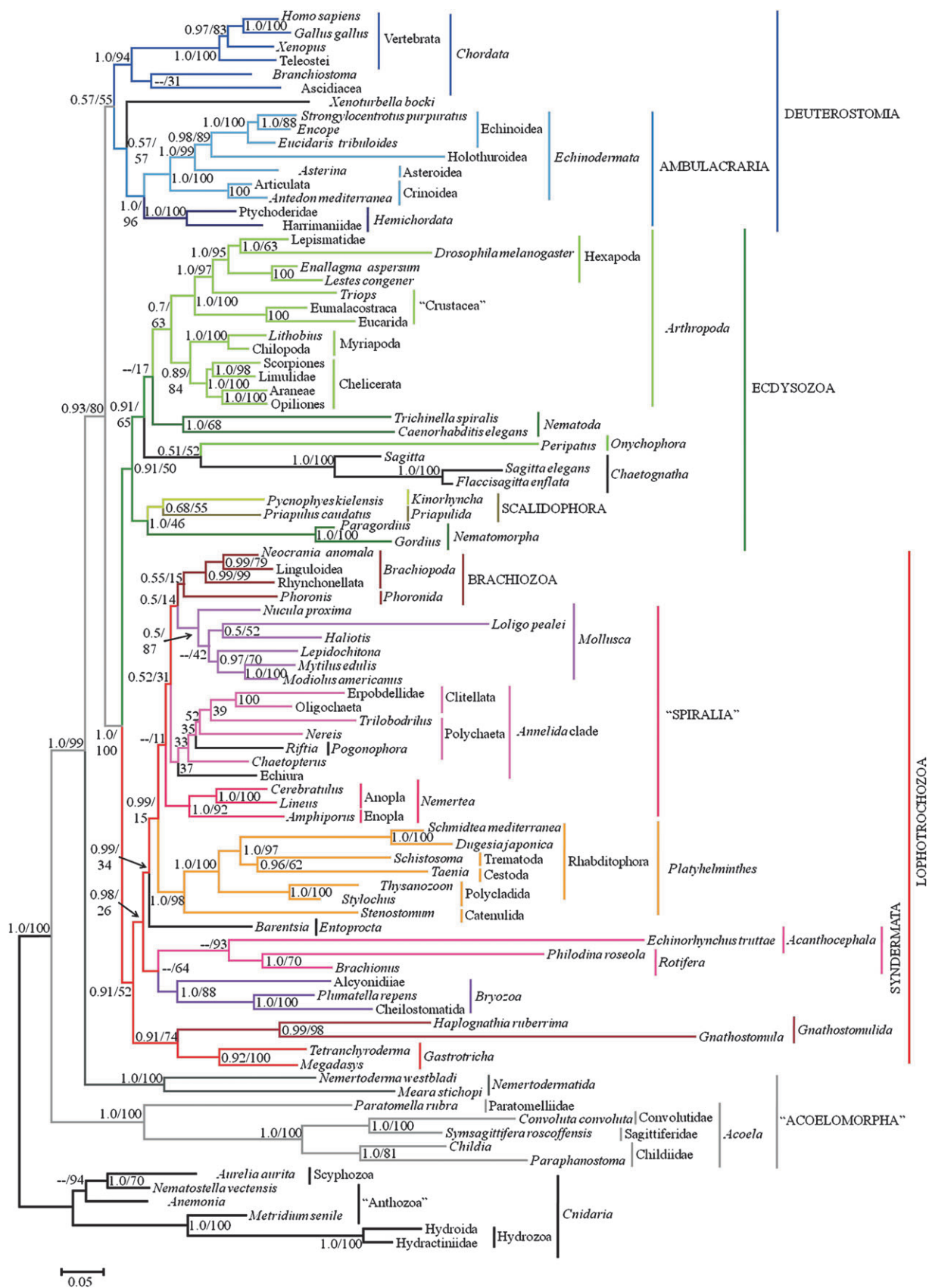


FIG. 2.—Phylogenetic tree for the ML method. Bayesian Posterior Probabilities and Bootstrap Support values are indicated in the nodes. Dashes specify nodes not recovered for BI. The scale bar indicates the number of changes per site. Paraphyletic clades are indicated within quotation marks. For species names corresponding to each terminal, see supplementary table 4, Supplementary Material online.

being at their limit of resolution with present day models and methods of analyses. Nonetheless, the general agreement in the topology that begins to arise in the analyses done with different markers (as an example Dunn et al. and this work) can be seen as an indicator of the topologies getting closer to the real tree, or said in another way, the new markers effectively add information although do not reach high support for complex data matrices.

Acoelomorpha as Paraphyletic Basal Bilaterians

Both inference methods robustly show a paraphyletic Acoelomorpha as sister group to the other bilaterians. Their relationships to Platyhelminths are rejected by the comparison of topologies with AU test (table 1, hypothesis 3) (table 1). This result confirms earlier studies showing them as a paraphyletic assemblage at the base of the bilaterians (Ruiz-Trillo et al. 1999, 2002, 2004; Sempere et al. 2007; Wallberg et al. 2007; Paps et al. 2009). This is contrary to two recent phylogenomic reports, the first placing a single acoel species as a sister group to the deuterostomes (Philippe et al. 2007), an alternative also rejected here by comparison of topologies (table 1, hypotheses 2), and the second showing acoels within a clade that is sister group to the spiralian lophotrochozoans (although with no support, Dunn et al. 2008). The large number of Acoelomorpha included here (five acoels and two nemertodermatids) probably provided greater stability to this clade. Moreover, we recover a partial internal phylogeny of the acoels that is in concordance to a recent systematic proposal based on molecular and sperm structure data (Hooge and Tyler 2006).

Deuterostomia and *Xenoturbella*

The Deuterostomia appear as a monophyletic clade but with low support in both inference methods (fig. 2). Deuterostomes split into a robust Chordata and a weak clade including *Xenoturbella* and a strong Ambulacraria (Hemichordata + Echinodermata). PhyML using SPR heuristics shows Cephalochordata + (Urochordata + Vertebrata), in agreement with recent molecular studies (Bourlat et al. 2006; Philippe et al. 2007; Dunn et al. 2008). The low support for the deuterostomes holds in many recent molecular studies (Bourlat et al. 2006; Delsuc et al. 2006, 2008; Mallatt and Giribet 2006; Philippe et al. 2007) and can be explained here by the unstable nature of *Xenoturbella*. When *Xenoturbella* is removed from the data set, the deuterostomes support increases to 83% Bootstrap Support (BS, results not shown). The best BI tree (data not shown) positions *Xenoturbella* splitting after Acoela and Nemertodermatida, as sister group to the rest of bilaterians. Moreover, this latter position is not rejected by the AU test (table 1, hypothesis 4), though further data are needed to corroborate their position. It is noteworthy that, although we have no representatives of the Class Ophiuroidea, the other echinoderm classes robustly conform to previous studies (Littlewood et al. 1997).

Ecdysozoa and Chaetognatha

The Ecdysozoa phylogeny shows two main clades: 1) Scalidophora (Priapulida + Kinorhyncha) plus Nematoda

and 2) a clade including Nematoda, Onychophora, Chaetognatha, and Arthropoda. The controversial phylogenetic position of Chaetognatha is one of the biggest conundrums in animal phylogeny. Recent data on Hox cluster genes suggest a new position close to the base of the Bilateria (Papillon et al. 2003), whereas phylogeny based on ribosomal genes (Mallatt and Winchell 2002), mitochondrial DNA (Helfenbein et al. 2004; Papillon et al. 2004) and multigenic approaches (Marletaz et al. 2006; Matus et al. 2006; Philippe et al. 2007) place them as sister group to the protostomes or within that group. Placement of chaetognaths inside the ecdysozoans has recently been suggested (Helmkampf et al. 2007), a position also shown in our analyses and supported by the comparison of topologies that clearly rejects chaetognaths as sister group to all the protostomes, to Lophotrochozoa or to Ecdysozoa (table 1, hypotheses 7 to 9).

In our trees, chaetognaths group in a clade together with nematodes, onychophorans, and arthropods. This group, however, is likely a consequence of an LBA artifact. To test it, we ran two ML analyses (data not shown). The first, excluding chaetognaths, placed onychophorans with arthropods with 94% support, whereas the second, excluding onychophorans, positions chaetognaths with nematodes (7%). These results point to an internal LBA effect between nematodes, onychophorans, and the chaetognaths, which can also explain the lack of resolution for the relationships among the main ecdysozoan clades. Despite these problems, arthropods show a reliable internal phylogeny grouping Myriapoda with Chelicerata and Hexapoda with Crustacea, both groups also recovered in other studies (Hwang et al. 2001; Peterson and Eernisse 2001; Dunn et al. 2008; Paps et al. 2009).

Lophotrochozoa

Lophotrochozoa were first defined as the last common ancestor of annelids, molluscs, the lophophorate phyla (Brachiopoda, Phoronida, and Bryozoa), and all the descendants of that ancestor (Halanych et al. 1995). Hox gene residues (Balavoine et al. 2002) and other molecular markers (Ruiz-Trillo et al. 2002; Anderson et al. 2004; Peterson et al. 2005) support this superclade. Their internal relationships, however, are far from settled, as highlighted by the wide variety of different proposals (Zrzavy et al. 1998; Giribet et al. 2000; Peterson and Eernisse 2001; Mallatt and Winchell 2002; Passamanek and Halanych 2006).

A well-supported clade made by Gnathostomulida and Gastrotricha appear as the first splitting lophotrochozoan lineage. A close relationship among gnathostomulans and gastrotrichans has been suggested on basis of morphology (Rieger 1976; Sterrer et al. 1985; Zrzavy et al. 1998) and molecules (Zrzavy et al. 1998; Giribet et al. 2000; Todaro et al. 2006). These two phyla were also recovered as basal lophotrochozoan clades in recent SSU + LSU analyses (Paps et al. 2009) though a recent ESTs study was not able to sort out their relationships (Dunn et al. 2008). Our trees suggest for them a new position as the most basal lophotrochozoans, in contradiction with other proposals such as Gnathifera (Rieger and Tyler 1995; Ahlrichs 1997) or

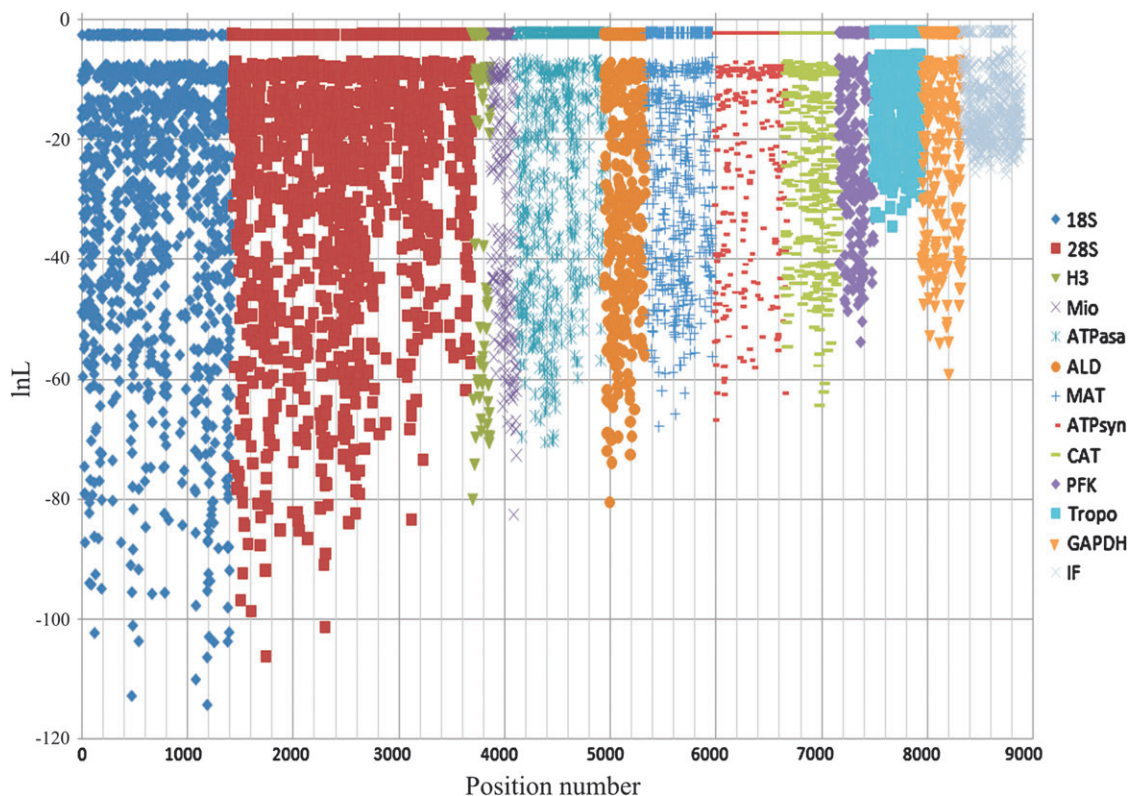


FIG. 3.—Plot of the likelihood values of each site for the ML tree obtained (Y axis) in relation to which gene they belong (X axis).

Cycloneuralia (sensu lato, Gastrotricha + Nematoda + Scalidophora; Nielsen 2001).

Our results relate Bryozoa with Rotifera and Acanthocephala. A close relationship between the last two clades has been suggested by morphology (see review in Garey and Schmidt-Rhaesa 1998) and SSU (Winnepenninckx et al. 1995; named Syndermata in Zrzavy et al. 1998). As regards the Bryozoa, early SSU studies showed that bryozoans do not belong to lophophorates (Cohen 2000), and recent ESTs analyses place them close to spiralian (Hausdorf et al. 2007; Dunn et al. 2008). Our AU test rejects the monophyly of Bryozoa + Entoprocta if the entoproctan is forced inside the Syndermata + Bryozoa but does not reject it if the bryozoans are removed from syndermatans and placed with entoproctans as sister group to spiralian ("Spiralia" + Brachiozoa + Platyhelminthes in fig. 2). Therefore, the position of entoproctans as sister group to spiralian is highly supported by our analyses, whereas the position of bryozoans rests unresolved.

Platyhelminthes and Nemertea have traditionally been clustered together, either because they were supposed to share an acoelomate condition (Hyman 1951; and Parenchymia of Nielsen 1995), or because they share some larval features (Nielsen 2001). However, SSU molecular studies and a reassessment of the morphological features of nemertines convincingly showed them to be coelomate animals more related molluscs and annelids than to Platyhelminthes (Turbeville et al. 1992). In turn, molecular studies placed the Platyhelminthes either as basal lophotrochozoans (Ruiz-Trillo et al. 1999) or within the Platyzoa (Giribet et al. 2000). Our results show instead Platyhelminthes

and Nemertea to branch paraphyletically with a clade of spiralian animals. Moreover, the internal phylogeny for the sampled flatworm classes is robust and highly congruent with the modern systematics of the group (see a review in Bagnù and Riutort 2004).

Monophyly of the Spiralia (animals bearing a spiral-quartet cleavage, indicated in fig. 2) has been recovered in many molecular studies. The Spiralia appear paraphyletic due to the inclusion of Phoronida + Brachiopoda within this group. A clade of phoronids and brachiopods, often named Brachiozoa or Phoronozoa, is increasingly recovered in recent molecular phylogenies (Zrzavy et al. 1998; Cohen 2000; Peterson and Eernisse 2001). The relationship between Brachiozoa and Spiralia has been hinted by recent EST studies (albeit relating them with nemertines rather than molluscs, Dunn et al. 2008) and recent paleontological studies have also pointed out a likely affiliation to molluscs (Morris and Peel 1995). The other spiralian group relates Echiura and Pogonophora with Annelida in agreement with recent studies (Hessling and Westheide 2002; Bleidorn et al. 2003; Struck et al. 2007).

Gene Contribution

It is often suggested that to obtain better resolution, more data are needed, implying more taxa, but most of the time meaning longer sequences (more markers). As is pointed out in the Introduction, the latter is not enough if it falls short regarding the taxon sampling. We wanted to explore the gene contribution to our tree, in order to elucidate if each marker helps to improve or not the phylogeny

Table 2
Likelihood Contribution for Each Gene

Gene	Gene Length (in Percentage)	Contribution to $-\ln L$ (in Percentage)
<i>28S</i>	25.48	25.78
<i>18S</i>	16.06	22.22
<i>ATPase</i>	9.37	8.97
<i>ATPsyn</i>	7.68	4.64
<i>MAT</i>	7.09	7.3
<i>IF</i>	6.26	4.93
<i>CAT</i>	5.74	4.88
<i>Tropo</i>	5.09	3.6
<i>ALD</i>	4.39	5.77
<i>GAPDH</i>	4.27	3.16
<i>PFK</i>	3.76	3.37
<i>Mio</i>	2.76	3.83
<i>H3</i>	2.04	1.54

resolution. In order to do that, two approaches were used: The first uses the site likelihood values for the ML tree (fig. 3 and table 2), whereas the second removes one gene at a time to infer a new tree (table 3).

Figure 3 plots the site likelihood score of each position into the ML tree; at this point, it is important to remember that the $\ln L$ is a negative value and that the closer it is to zero the higher is the likelihood. Curiously there is a “band” of sites around an $\ln L$ value of -2 for all genes that correspond to the conserved sites in the alignment. The rest of the sites have variable $-\ln L$ values, with some genes having nearly all the positions with values close to zero (*IF*, *Tropo*), and others presenting a greater dispersion (the two ribosomal genes). Table 2 shows the relative length of the gene in the alignment and its contribution to the $-\ln L$ of the tree (both in percentage). If all sites of all genes were contributing exactly the same, the two values would be expected to be correlated. We can observe how some genes have higher contribution to the $-\ln L$ value than assumed for their length (hence holding a higher proportion of low probability sites) and the other way around. *Tropo* and *IF*, as expected from the observations on the graph, have a greater proportion of high probability sites than expected by their length, whereas genes like *ALD*, myosin, or *18S* have a lower proportion. In conclusion, in our data set, more characters do not necessarily mean more information, and this shows that long genes can have many sites with very low probabilities, whereas short genes can hold many high-probability sites.

The site likelihood approach used, though informative, has two flaws: First, the gene contribution is measured on

the phylogenetic hypothesis inferred, despite the latter being true or not. Second, the site likelihood calculation enforces all the positions to contribute somewhat to the obtained tree; therefore, any gene adding more noise than information cannot be detected. To work around these limitations, we used another method based on removing one gene at a time from the matrix and evaluating its impact on the BS for different nodes (table 3). We have selected the clades that are widely accepted and at the same time have an intermediate BS value in the original tree, so the values can go either up or down in our experiment. Theoretically, the BS increase in a clade when a gene is removed will be interpreted as a higher proportion of sites in the alignment giving support to the node, whereas a BS decrease means an increase in the proportion of sites not supporting that node. Hence, the observation of a general fall in BS values when removing a marker means that this gene held sites supporting the node, whereas a general BS upraise could be interpreted as the gene not contributing or even having contradictory information.

The removal of the ribosomal genes results in a BS decrease for all the nodes analyzed, and the same holds for *ALD*. The rest of the genes show a variable proportion of nodes going up or down, with the exception of *H3* for which all nodes but one increase. This last result is expected taking into account the profile shown by this molecule in figure 3, where it shows a group of highly conserved sites and presents a gap (between 20 and 40 $-\ln L$), whereas the rest of the sites have low probabilities likely not contributing to the tree. The rest of the genes all seem to be contributing information in some nodes, whereas their removal increases the BS value for others. These groupings are still a simplification, as they do not take into account at which levels the nodes are affected. For instance, when comparing all the tree nodes (data not shown), some genes affect nodes close to the outgroup, whereas others affect values only at the tips.

We can conclude that the addition of genes is highly positive, most of them adding some information to the tree, but the first approach shows that the quantity of information is not rigorously correlated to the length of sequence and the second method shows that not all the genes improve all the nodes of the phylogeny. Therefore, the indiscriminate addition of huge quantities of information does not grant a phylogeny of higher resolution. The idea of removing markers from a matrix is not new and has been already used in previous studies (Philippe et al. 2005). The consequences of these results into the new phylogenomic field warn about

Table 3
Effect on the Bootstrap Values When a Gene Is Removed

	Original Tree	<i>18S</i>	<i>28S</i>	<i>H3</i>	<i>Mio</i>	<i>ATPasa</i>	<i>ALD</i>	<i>MAT</i>	<i>ATPsyn</i>	<i>CAT</i>	<i>PFK</i>	<i>Tropo</i>	<i>GAPDH</i>	<i>IF</i>
Basal “Acoelomorpha”	80	49 ^a	72	87	76	85	62	88	75	77	80	77	66	74
Deuterostomia	55	NR	49	61	66	68	49	58	53	58	61	27	52	63
Ecdysozoa	50	NR	27	58	NR	NR	44	37	48	5	45	35	55	NR
Lophotrochozoa	52	NR	33	52	48	NR	44	31	51	18	48	35	58	NR
Annelida + Echiura + Pogonophora	37	16	NR	60	45	65	32	40	42	55	42	51	32	38

NR stands for nonrecovered clades. Italic numbers indicate bootstrap values five units under the original tree score and bold numbers indicate bootstrap values five units above.

^a Indicates that in this case, *Xenoturbella* groups with Acoela (83% BS).

the arbitrary addition of markers to animal molecular matrices. Although when using fast molecular methods adding much information is an added bonus, here we would like to encourage a further filtering of the data before final analyses are carried out. Thus, the simple removal of one gene (H3) results in a BS increase of many nodes. However, that raises a main new question: How to devise objective parameters to evaluate which genes are mostly informative and which ones mostly noisy?

Concluding Remarks

To obtain a robust phylogeny of the bilaterians, a necessary prerequisite seems to get a fair balance between number of taxa and number of characters sampled. In addition, a deep taxon sampling, the use of probabilistic methods and adequate models has helped us to recover a novel bilaterian phylogeny. Our results show, as in most previous studies, the monophyly of Deuterostomia, Protostomia, Lophotrochozoa, and Ecdysozoa. However, it sheds new light on some dark, conflicting areas of the bilaterian tree that appear better resolved than those derived from current ESTs analyses. We mainly refer to the basal relationships of the Lophotrochozoa and to the position of a paraphyletic Acoelomorpha as earliest extant branching bilaterians. Moreover, we also suggest to include the Chaetognatha within ecdysozoans, though further information is needed to settle their position, and new internal relationships for many phyla that match studies focused to solve their internal relationships. On the other side of the coin, our results did not find high support for some regions of the tree such as the status of *Xenoturbella*, the internal relationships of Ecdysozoa, and some intermediate branches of the Lophotrochozoa. Finally, our analysis on gene contribution points out that more data do not necessarily mean a better resolved phylogeny. Therefore, the markers produced by high-throughput methods must be carefully evaluated before being unsystematically added to the final matrix. We predict that a similarly balanced approach, incorporating better filtered EST collections, and better taxon sampling, will substantially improve our understanding of bilaterian evolution.

Supplementary Material

Supplementary tables 1–4 and supplementary figure 1 are available at *Molecular Biology and Evolution* online (<http://www.mbe.oxfordjournals.org/>).

Acknowledgments

We thank Dr Kevin J. Peterson and Dr Iñaki Ruiz-Trillo for suggestions, inspiring discussions and critical reading of early versions of the manuscript. We also thank the suggestions made by the anonymous referees that helped to improve the manuscript. We are grateful to Dr Kevin J. Peterson for kindly providing the *Childia* and *Brachionus* sequences (for ALD, MAT, ATPsyn, Catalase, and PFK) and to all the experts who provided biological material,

especially to Dr Carles Ribera and Dr Ulf Jondelius. This research was supported by Dirección General de Investigación-Ministerio de Educación y Ciencia grant and CIRIT grant to J.B. We also wish to thank David Vicente (Barcelona Supercomputing Centre) and Alfred Gil (CESCA) for their kind help in installing programs and resolving problems in their use.

Literature Cited

- Abouheif E, Zardoya R, Meyer A. 1998. Limitations of metazoan 18S rRNA sequence data: implications for reconstructing a phylogeny of the animal kingdom and inferring the reality of the Cambrian explosion. *J Mol Evol*. 47:394–405.
- Ahlrichs WH. 1997. Epidermal ultrastructure of *Seison nebaliae* and *Seison annulatus*, and a comparison of epidermal structures within the Gnathifera. *Zoomorphology*. 117:41–48.
- Anderson FE, Cordoba AJ, Thollessen M. 2004. Bilaterian phylogeny based on analyses of a region of the sodium-potassium ATPase beta-subunit gene. *J Mol Evol*. 58:252–268.
- Baguña J, Riutort M. 2004. Molecular phylogeny of the Platyhelminthes. *Can J Zool*. 82:168–193.
- Balavoine G, de Rosa R, Adoutte A. 2002. Hox clusters and bilaterian phylogeny. *Mol Phylogenet Evol*. 24:366–373.
- Bleidorn C, Vogt L, Bartolomaeus T. 2003. New insights into polychaete phylogeny (Annelida) inferred from 18S rDNA sequences. *Mol Phylogenet Evol*. 29:279–288.
- Boore JL, Macey JR, Medina M. 2005. Sequencing and comparing whole mitochondrial genomes of animals. *Methods Enzymol*. 395:311–348.
- Bourlat SJ, Juliusdottir T, Lowe CJ, et al. (13 co-authors). 2006. Deuterostome phylogeny reveals monophyletic chordates and the new phylum Xenoturbellida. *Nature*.
- Bourlat SJ, Nielsen C, Economou AD, Telford MJ. 2008. Testing the new animal phylogeny: a phylum level molecular analysis of the animal kingdom. *Mol Phylogenet Evol*. 49:23–31.
- Brinkmann H, Philippe H. 2008. Animal phylogeny and large-scale sequencing: progress and pitfalls. *J Syst Evol*. 46:274–286.
- Castresana J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol*. 17:540–552.
- Cohen BL. 2000. Monophyly of brachiopods and phoronids: reconciliation of molecular evidence with Linnaean classification (the subphylum *Phoroniformea* nov.). *Proc R Soc B Biol Sci*. 267:225–231.
- Delsuc F, Brinkmann H, Chourrout D, Philippe H. 2006. Tunicates and not cephalochordates are the closest living relatives of vertebrates. *Nature*. 439:965–968.
- Delsuc F, Tsagkogeorga G, Lartillot N, Philippe H. 2008. Additional molecular support for the new chordate phylogeny. *Genesis*. 46:592–604.
- Dopazo H, Santoyo J, Dopazo J. 2004. Phylogenomics and the number of characters required for obtaining an accurate phylogeny of eukaryote model species. *Bioinformatics*. 20:i116–i121.
- Dunn CW, Hejnal A, Matus DQ, et al. (17 co-authors). 2008. Broad phylogenomic sampling improves resolution of the animal tree of life. *Nature*. 452:745–749.
- Ehlers U, Ahlrichs W, Lemburg C, Schmidt-Rhaesa A. 1996. Phylogenetic systematization of the Nematelminthes (Aschelminthes). *Verh Deutsch Zool Ges*. 89:8.
- Felsenstein J. 1978. Cases in which parsimony or compatibility methods will be positively misleading. *Syst Biol*. 27:401–410.

- Garey JR, Schmidt-Rhaesa A. 1998. The Essential role of “minor” phyla in molecular studies of animal evolution. *Am Zool.* 38:907–917.
- Giribet G, Distel DL, Polz M, Sterrer W, Wheeler WC. 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 GD, Wheeler WC. 2001. Arthropod phylogeny based on eight molecular loci and morphology. *Nature.* 413:157–161.
- Guindon S, Gascuel O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol.* 52:696–704.
- Halanych KM. 2004. The new view of animal phylogeny. *Annu Rev Ecol Evol Syst.* 35:229–256.
- Halanych KM, Bacheller JD, Aguinaldo AM, Liva SM, Hillis DM, Lake JA. 1995. Evidence from 18S ribosomal DNA that the lophophorates are protostome animals. *Science.* 267:1641–1643.
- Hartmann S, Vision T. 2008. Using ESTs for phylogenomics: can one accurately infer a phylogenetic tree from a gappy alignment? *BMC Evol Biol.* 8:95.
- Hausdorf B, Helmkamp M, Meyer A, Witek A, Herlyn H, Bruchhaus I, Hankeln T, Struck TH, Lieb B. 2007. Spiralian phylogenomics supports the resurrection of Bryozoa comprising Ectoprocta and Entoprocta. *Mol Biol Evol.* 24:2723–2729.
- Helgenbein KG, Fourcade HM, Vanjani RG, Boore JL. 2004. The mitochondrial genome of *Paraspadella gotoi* is highly reduced and reveals that chaetognaths are a sister group to protostomes. *Proc Natl Acad Sci USA.* 101:10639–10643.
- Helmkamp M, Bruchhaus I, Hausdorf B. 2007. Multigene analysis of lophophorate and chaetognath phylogenetic relationships. *Mol Phylogenet Evol.* 46:206–214.
- Hessling R, Westheide W. 2002. Are Echiura derived from a segmented ancestor? Immunohistochemical analysis of the nervous system in developmental stages of *Bonellia viridis*. *J Morphol.* 252:100–113.
- Hooge MD, Tyler S. 2006. Concordance of molecular and morphological data: the example of the Acoela. *Integr Comp Biol.* 46:118–124.
- Hwang UW, Friedrich M, Tautz D, Park CJ, Kim W. 2001. Mitochondrial protein phylogeny joins myriapods with chelicerates. *Nature.* 413:154–157.
- Hyman LH. 1951. *The invertebrates: Platyhelminthes and Rhynchoela. The acelomate bilateria.* New York: McGraw-Hill. Vol. 2
- Lartillot N, Brinkmann H, Philippe H. 2007. Suppression of long-branch attraction artefacts in the animal phylogeny using a site-heterogeneous model. *BMC Evol Biol.* 7(Suppl 1):S4.
- Lartillot N, Philippe H. 2004. A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. *Mol Biol Evol.* 21:1095–1109.
- Littlewood DTJ, Smith AB, Clough KA, Emson RH. 1997. The interrelationships of the echinoderm classes: morphological and molecular evidence. *Biol J Linn Soc.* 61:409–438.
- Mallat 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.
- Mallat J, Winchell CJ. 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.
- Marletaz F, Martin E, Perez Y, et al. (11 co-authors). 2006. Chaetognath phylogenomics: a protostome with deutero-stome-like development. *Curr Biol.* 16:R577–R578.
- Matus DQ, Copley RR, Dunn CW, Hejnal A, Eccleston H, Halanych KM, Martindale MQ, Telford MJ. 2006. Broad taxon and gene sampling indicate that chaetognaths are protostomes. *Curr Biol.* 16:R575–R576.
- Morris SC, Peel JS. 1995. Articulated Halkieriids from the lower Cambrian of North Greenland and their role in early Protostome evolution. *Phil Trans R Soc Lond Ser B Biol Sci.* 347:305–358.
- Mushegian AR, Garey JR, Martin J, Liu LX. 1998. Large-scale taxonomic profiling of eukaryotic model organisms: a comparison of orthologous proteins encoded by the human, fly, nematode, and yeast genomes. *Genome Res.* 8:590–598.
- Nielsen C. 1995. *Animal evolution: interrelationships of the living phyla.* Oxford: Oxford University Press.
- Nielsen C. 2001. *Animal evolution: interrelationships of the living phyla.* 2nd ed. (31 March 2001), Oxford: Oxford University Press.
- Page RDM. 1996. TREEVIEW: an application to display phylogenetic trees on personal computers. *Compu Appl Biosci.* 12:357–358.
- Papillon D, Perez Y, Caubit X, Le Parco Y. 2004. Identification of Chaetognaths as Protostomes is supported by the analysis of their mitochondrial genome. *Mol Biol Evol.* 21:2122–2129.
- Papillon D, Perez Y, Fasano L, Le Parco Y, Caubit X. 2003. Hox gene survey in the chaetognath *Spadella cephaloptera*: evolutionary implications. *Dev Genes Evol.* 213:142–148.
- Paps J, Bagaña J, Riutort M. 2009. Lophotrochozoa internal phylogeny: new insights from an up-to-date analysis of nuclear ribosomal genes. *Proc R Soc B Biol Sci.* 276:1245–1254.
- Passamanek Y, Halanych KM. 2006. Lophotrochozoan phylogeny assessed with LSU and SSU data: evidence of lophophorate polyphyly. *Mol Phylogenet Evol.* 40:20–28.
- Peterson KJ, Eernisse DJ. 2001. Animal phylogeny and the ancestry of bilaterians: inferences from morphology and 18S rDNA gene sequences. *Evol Dev.* 3:170–205.
- Peterson KJ, Lyons JB, Nowak KS, Takacs CM, Wargo MJ, McPeck MA. 2004. Estimating metazoan divergence times with a molecular clock. *Proc Natl Acad Sci USA.* 101:6536–6541.
- Peterson KJ, McPeck MA, Evans DAD. 2005. Tempo and mode of early animal evolution: inferences from rocks, Hox, and molecular clocks. *Paleobiology.* 31:36–55.
- Philippe H, Brinkmann H, Martinez P, Riutort M, Bagaña J. 2007. Acoel flatworms are not Platyhelminthes: evidence from phylogenomics. *PLoS ONE.* 8:e717.
- Philippe H, Lartillot N, Brinkmann H. 2005. Multigene analyses of bilaterian animals corroborate the monophyly of Ecdysozoa, Lophotrochozoa, and Protostomia. *Mol Biol Evol.* 22:1246–1253.
- Philippe H, Telford MJ. 2006. Large-scale sequencing and the new animal phylogeny. *Trends Ecol Evol.* 21:614–620.
- Posada D, Crandall KA. 1998. MODELTEST: testing the model of DNA substitution. *Bioinformatics.* 14:817–818.
- Rieger RM. 1976. Monociliated epidermal cells in *Gastrotricha*: significance for concepts of early metazoan evolution. *Z Zool Syst Evol.* 14:198–226.
- Rieger RM, Tyler S. 1995. Sister-group relationship of Gnathostomulida and Rotifera–Acanthocephala. *Invertebr Biol.* 114:186–188.
- Rieger RM, Tyler S, Smith JPS III. 1991. Rieger GE. 1991. *Platyhelminthes: Turbellaria.* New York: Wiley-Liss.
- Ronquist F, Huelsenbeck JP. 2003. MrBayes 3: bayesian phylogenetic inference under mixed models. *Bioinformatics.* 19:1572–1574.
- Ruiz-Trillo I, Paps J, Loukota M, Ribera C, Jondelius U, Bagaña 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 HM, Baguna J, Boore JL. 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 DT, Herniou EA, Baguna J. 1999. Acoel flatworms: earliest extant bilaterian Metazoans, not members of Platyhelminthes. *Science*. 283:1919–1923.
- Sempere LF, Cole CN, McPeck MA, Peterson KJ. 2006. The phylogenetic distribution of metazoan microRNAs: insights into evolutionary complexity and constraint. *J Exp Zool (Mol Dev Evol)*. 306:575–588.
- Sempere LF, Martinez P, Cole C, Baguna J, Peterson KJ. 2007. Phylogenetic distribution of microRNAs supports the basal position of acoel flatworms and the polyphyly of Platyhelminthes. *Evol Dev*. 9:409–415.
- Shimodaira H, Hasegawa M. 2001. CONSEL: for assessing the confidence of phylogenetic tree selection. *Bioinformatics*. 17:1246–1247.
- Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics*. 22:2688–2690.
- Sterrer W, Mainitz M, Rieger R. 1985. *Gnathostomulida: enigmatic as ever*. Oxford: Clarendon Press.
- Struck T, Schult N, Kusen T, Hickman E, Bleidorn C, McHugh D, Halanych K. 2007. Annelid phylogeny and the status of Sipuncula and Echiura. *BMC Evol Biol*. 7:57.
- Swofford DL. 2000. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. ed, Sunderland (MA): Sinauer Associates.
- Todaro MA, Telford MJ, Lockyer AE, Littlewood DTJ. 2006. Interrelationships of the Gastrotricha and their place among the Metazoa inferred from 18S rRNA genes. *Zool Scripta*. 35:251–259.
- Turbeville JM, Field KG, Raff RA. 1992. Phylogenetic position of phylum Nemertini, inferred from 18S rRNA sequences: molecular data as a test of morphological character homology. *Mol Biol Evol*. 9:235–249.
- 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.
- Winnepenninckx B, Backeljau T, Mackey LY, Brooks JM, Dewachter R, Kumar S, Garey JR. 1995. 18s Ribosomal-Rna data indicate that Aschelminthes are Polyphyletic in origin and consist of at least 3 distinct clades. *Mol Biol Evol*. 12:1132–1137.
- Zrzavy J, Mihulka S, Kepka P, Bezdek A, Tietz D. 1998. Phylogeny of the Metazoa based on morphological and 18S ribosomal DNA evidence. *Cladistics*. 14:249–285.

Manolo Gouy, Associate Editor

Accepted July 9, 2009