EVOLUTION & DEVELOPMENT 5:4, 346–359 (2003)

Conflicting phylogenetic signals at the base of the metazoan tree

Antonis Rokas,[a,1] Nicole King,[a,1] John Finnerty,[b] and Sean B. Carroll[a,*]

[a]Howard Hughes Medical Institute and Laboratory of Molecular Biology, University of Wisconsin-Madison, 1525 Linden Drive, Madison, WI 53706-1596, USA

[b]Department of Biology, Boston University, 5 Cummington St., Boston, MA 02215, USA

*Author for correspondence (e-mail: sbcarrol@facstaff.wisc.edu)

SUMMARY A phylogenetic framework is essential for understanding the origin and evolution of metazoan development. Despite a number of recent molecular studies and a rich fossil record of sponges and cnidarians, the evolutionary relationships of the early branching metazoan groups to each other and to a putative outgroup, the choanoflagellates, remain uncertain. This situation may be the result of the limited amount of phylogenetic information found in single genes and the small number of relevant taxa surveyed. To alleviate the effect of these analytical factors in the phylogenetic reconstruction of early branching metazoan lineages, we cloned multiple protein-coding genes from two choanoflagellates and diverse sponges, cnidarians, and a ctenophore. Comparisons of sequences for α-tubulin, β-tubulin, elongation factor 2, HSP90, and HSP70 robustly support the hypothesis that choanoflagellates are closely affiliated with animals. However, analyses of single and concatenated amino acid sequences fail to resolve the relationships either between early branching metazoan groups or between Metazoa and choanoflagellates. We demonstrate that variable rates of evolution among lineages, sensitivity of the analyses to taxon selection, and conflicts in the phylogenetic signal contained in different amino acid sequences obscure the phylogenetic associations among the early branching Metazoa. These factors raise concerns about the ability to resolve the phylogenetic history of animals with molecular sequences. A consensus view of animal evolution may require investigations of genome-scale characters.

INTRODUCTION

The origin of multicellular Metazoa from a unicellular ancestor, and the subsequent metazoan radiation, involved the evolution of new modes of development. Comparing modern representatives of the earliest branching metazoan lineages (e.g., sponges, cnidarians, ctenophores) with bilaterians may reveal the role of novel developmental mechanisms in the early evolution of Metazoa. Inferences regarding the polarity of changes in development along evolutionary lineages (i.e., distinguishing between what is ancestral and what is derived) require a robust phylogenetic framework. Therefore, a well-resolved phylogeny of early branching metazoans and candidate outgroups, such as the choanoflagellates, is essential for understanding the beginnings of animal evolution.

The phylogenetic relationships of sponges, cnidarians, and ctenophores have recently received increased attention. From a histological perspective, sponges are perhaps the simplest of all animals, lacking tissue level organization. Higher level sponge taxonomy is based on the composition of small skeletal elements known as spicules (Brusca and Brusca 1990). Members of the sponge classes Demospongiae and Hexactinellida are often treated as sister-groups because both synthesize siliceous spicules, whereas a third class, Calcarea, synthesizes spicules from calcium carbonate. However, despite the differences in the composition of their spicules, Calcarea and Demospongiae are sometimes linked to the exclusion of Hexactinellida because their cellularization and lack of symmetry differ dramatically from the syncytia and partial radial symmetry found in hexactinellids (Brusca and Brusca 1990).

Uncertainty also exists about whether the sponges are monophyletic or paraphyletic. For instance, one study found that 18S rDNA analyses cannot differentiate between paraphyly and monophyly of the sponges (Medina et al. 2001), whereas another analysis supported sponge paraphyly but failed to resolve the position of Hexactinellida (Borchiellini et al. 2001). Furthermore, analyses of single protein-coding genes either fail to reveal the interrelationships between Calcarea, Demospongiae, and Hexactinellida (Borchiellini et al. 1998) or support a hypothesis (the grouping of Calcarea and Demospongiae; Kruse et al. 1998) that is in conflict with 18S rDNA analyses (Medina et al. 2001). Whether sponges are monophyletic or paraphyletic bears directly on interpretations of the early stages in the evolution of metazoan development.

[1]These authors contributed equally to this work.

Similarly, the relationships between ctenophores, cnidarians, and Bilateria have been obscured by conflicts between inferences from morphological and molecular phylogenetic analyses. Ctenophores and cnidarian jellyfish are superficially similar in adult morphology, implying a sister-group relationship. However, histological, developmental, and molecular phylogenetic analyses suggest that they represent independent lineages (Nielsen 2001). Although developmental and morphological data suggest that Cnidaria is a sister-group to Ctenophora and Bilateria (Martindale et al. 2002), 18S rDNA studies repeatedly recover ctenophores as an outgroup of Cnidaria and Bilateria (Collins 1998; Medina et al. 2001; Podar et al. 2001).

In addition to the controversies plaguing the interrelationships between early branching Metazoa, uncertainty surrounds the exact placement of a candidate outgroup of Metazoa, the Choanoflagellata. This group of unicellular flagellated protozoa was first linked to animal evolution due to the striking similarities between their ultrastructure and that of a unique cell type, the choanocyte, found in sponges (James-Clark 1866; Saville-Kent 1880–1882; Hibberd 1975). Both choanoflagellates and sponge choanocytes possess a single apical flagellum surrounded by a ring or "collar" of actin-filled tentacles, allowing the cells to generate water currents and capture bacteria from the water column. Early morphological studies on choanoflagellates supported a close affiliation with Metazoa (James-Clark 1866; Saville-Kent 1880–1882; Hibberd 1975), providing evidence for the idea that choanoflagellates are a close outgroup of Metazoa (Nielsen 2001). However, other studies have suggested that choanoflagellates evolved before the divergence of the fungal and metazoan lineages (Cavalier-Smith 1987) and that the collar structures of choanoflagellates and sponge choanocytes are nonhomologous (Ax 1996). Finally, although the morphological simplicity of choanoflagellates supports a divergence before the origin of multicellular animals, it is also possible that choanoflagellates are metazoans that secondarily lost multicellularity (Rieger and Weyrer 1998). Conflicts in the inferences made from ultrastructure leave the phylogenetic placement of choanoflagellates unresolved.

Uncertainty concerning the placement of choanoflagellates has also arisen from molecular phylogenetic analyses. Although one 18S rDNA study supports a close affiliation between choanoflagellates and Metazoa (Atkins et al. 2000), other studies indicate that 18S rDNA lacks phylogenetic signal regarding the exact placement of choanoflagellates on the eukaryotic tree (Wainright et al. 1993; Kumar and Rzhetsky 1996; Van de Peer and De Wachter 1997). In contrast, single-gene analyses from five nuclear protein-coding genes (King and Carroll 2001; Snell et al. 2001) and analysis of multiple concatenated mitochondrial protein-coding genes (Lang et al. 2002) consistently and robustly support a choanoflagellate/metazoan clade but do not exclude the possibility that choanoflagellates evolved from sponges. The importance of robustly placing choanoflagellates on the eukaryotic tree, either as an outgroup of Metazoa or as a member of Metazoa, bears on their relevance as reference organisms for studies on the origin of metazoan multicellularity.

18S rDNA studies have offered valuable insights into our current view about metazoan and, more generally, about eukaryotic phylogenies (Wainright et al. 1993; Sogin and Silberman 1998). However, they have not been able to robustly resolve important branches in the metazoan tree, particularly ones addressed in this work, such as the exact placement of choanoflagellates and the relative branchings of sponges, cnidarians, ctenophores, and bilaterians (Kumar and Rzhetsky 1996; Van de Peer and De Wachter 1997). The failure of 18S rDNA to address these particular questions may arise from mutational saturation due to multiple substitutions, poorly aligned regions, or base frequency heterogeneity among taxa, each of which can create serious obstacles for use of DNA sequence in phylogenetic analyses of ancient divergences (Hasegawa and Hashimoto 1993; Galtier and Gouy 1995; Yang and Roberts 1995).

In cases where 18S rDNA provides conflicting or ambiguous results, phylogenetic analyses of amino acid sequences may be advantageous (Hasegawa and Hashimoto 1993; Yang et al. 1998), because they are known to be less affected by mutational saturation (although by no means immune). In recent years, amino acid sequences from well-conserved genes have shown to be effective in resolving "deep" branches of the eukaryotic tree (Gupta 1995; Hashimoto et al. 1995; Baldauf et al. 2000; Edgcomb et al. 2001; Fast et al. 2002); therefore, evaluation of evolutionary relationships between early branching metazoan phyla may benefit from the analysis of protein-coding genes. However, in contrast with the extensively sampled 18S rDNA data set, sampling of protein-coding loci from diverse Metazoa (and in particular from choanoflagellates, sponges, cnidarians, and ctenophores) has been sparse.

For "deep" phylogenies, single gene studies from different genes have been shown to contradict each other. This holds true even for well-supported and widely accepted phylogenetic hypotheses such as the sister-group relationship between Fungi and Metazoa (Baldauf and Palmer 1993; Baldauf et al. 2000), where data sets that robustly support alternative scenarios are available (Gupta 1995; Loytynoja and Milinkovitch 2001). It is therefore essential to compare the findings from 18S rDNA studies of early branching Metazoa (Wainright et al. 1993; Kumar and Rzhetsky 1996; Van de Peer and De Wachter 1997; Collins 1998; Medina et al. 2001) with those from protein-coding genes so that we can confidently extrapolate from the gene trees to the organismal trees.

To establish a phylogenetic framework within which to study the origin and evolution of multicellular Metazoa, we sought to reexamine the interrelationships between choanoflagellates, early branching Metazoa, and Bilateria using

multiple protein-coding loci. We cloned sequences for four highly conserved genes (α-tubulin, β-tubulin, elongation factor-2 [EF-2], and HSP90) from two choanoflagellates, seven sponges, two cnidarians, and a ctenophore. Previous phylogenetic analyses suggest that these genes have a slow rate of evolution and exist in low copy numbers in most eukaryotic genomes studied; furthermore, they have been reasonably well sampled from Bilateria (Gupta 1995; Baldauf et al. 2000; Keeling et al. 2000; Moreira et al. 2000). Moreover, after addition of data from a choanoflagellate, we reanalyzed the three HSP70 paralogs (cytoplasmic, endoplasmic reticulum [ER], and mitochondrial) found in metazoans, with special emphasis on cytoplasmic HSP70 (Boorstein et al. 1994; Borchiellini et al. 1998; Snell et al. 2001). These data sets represent an unprecedented sampling of multiple amino acid sequences from diverse choanoflagellates and early branching Metazoa. We analyzed these data, singly and in combination, under a statistical phylogenetic framework (maximum likelihood) to address the following questions: Are choanoflagellates an outgroup of Metazoa? What are the exact phylogenetic relationships among sponges, cnidarians, ctenophores, and Bilateria? What are the potential limitations in the interpretation of early branching metazoan relationships?

MATERIALS AND METHODS

Taxon and data set selection

Choanoflagellata and early branching metazoans

To avoid potential biases stemming from taxon sampling (Zwickl and Hillis 2002), we chose to analyze (where possible) multiple representative species from each phylum. In particular, we included two species from the phylum Choanoflagellata (*Monosiga brevicollis*, American Type Culture Collection [ATCC] 50154 and a *Proterospongia*-like species, ATCC 50818), seven species representing three classes in the phylum Porifera (Demospongiae: *Clypeatula cooperensis*, *Halichondria* sp., *Haliclona rubens* and *Suberites fuscus*; Hexactinellida: *Aphrocallistes vastus*; Calcarea: *Leucosolenia* sp. and *Scypha* sp.), two species representing two different classes in the phylum Cnidaria (Anthozoa: *Nematostella vectensis*; Scyphozoa: *Aurelia aurita*) and one species from the phylum Ctenophora (*Mnemiopsis leidyi*). Source material and classification of species are shown in Tables 1 and 2.

Ingroup taxa

In analyses of data sets constructed from concatenated amino acid sequences, inclusion of species from which large amounts of data are missing may cause phylogenetic artifacts. To avoid such artifacts, as well as biases due to uneven taxon sampling in the combined data analysis (see below), we selected diverse bilaterian species (*Mus musculus*, *Danio rerio*, *Caenorhabditis elegans*, *Drosophila melanogaster*, and *Schistosoma mansoni*) for which at least three of four analyzed genes were available in GenBank (classification shown in Table 2). For analyses of single genes we also included sequences of one species each from Echinodermata, Mollusca, Cnidaria, Ctenophora, and Choanoflagellata, when available.

Outgroup taxa

Fungi represent a well-established outgroup of Metazoa and choanoflagellates (Baldauf et al. 2000; King and Carroll 2001; Snell et al. 2001). *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*, two distantly related ascomycetes, were chosen as outgroup taxa. To increase our sampling of fungal diversity, for single-gene analyses we included (wherever sequences are available) an early branching ascomycete (*Pneumocystis carinii*) and a basidiomycete (*Schizophyllum commune*).

For brevity, in all subsequent text, figures, and tables, taxa will be mentioned by their generic name.

Gene selection

On the basis of previously published promising results (Gupta 1995; King and Carroll 2001), sequence length, broad taxonomic sampling across the Metazoa, and ease of amplification by degenerate polymerase chain reaction (PCR), we selected four protein-coding genes, α-tubulin, β-tubulin, EF-2, and HSP90, for analysis. We also assembled a data set for the HSP70 multigene family. Members of three HSP70 paralog groups are encoded by the nuclear genome and are distinguished by their intracellular localization in major cellular compartments; the organelles (mitochondria or chloroplasts), the ER, and the cytoplasm (Boorstein et al. 1994; Germot and Philippe 1999). The cytoplasmic and ER genes arose from a duplication that occurred in the ancestor of extant eukaryotes, whereas organelle HSP70s represent earlier lateral transfers from proteobacteria/cyanobacteria after the endosymbiotic events occurring in the eukaryotic progenitor (Boorstein et al. 1994). Previously published work sampled a choanoflagellate (*Monosiga ovata*) and multiple early branching Metazoa for the cytoplasmic HSP70 gene (Borchiellini et al. 1998; Snell et al. 2001). We supplemented this data set with a more even taxonomic sampling, with previously published data for the ER- and mitochondrial-HSP70s and with sequences for three HSP70 paralogs from a second choanoflagellate (*Monosiga brevicollis*).

Synthesis of cDNA templates for PCR

Plasmid or phage DNA prepared from cDNA libraries was used as template for *Monosiga*, ATCC 50818, *Halichondria*, *Nematostella*, and *Mnemiopsis* (Table 1). Purified cDNA from *Clypeatula* was provided by Kevin Peterson (personal communication). For *Suberites*, *Leucosolenia*, *Scypha*, and *Aurelia*, fresh tissue was acquired from the Marine Biological Laboratories (www.mbl.edu/animals/index.html), cleared of visible contaminants, and washed thoroughly in sterile seawater before use. These tissues, as well as frozen tissue from *Haliclona* (April Hill, personal communication) and *Aphrocallistes* (Sally Leys, personal communication) were homogenized for isolation of total RNA using TRIZOL (Invitrogen, Carlsbad, CA). Purified RNA was resuspended in depc-dH_2O and used as template for cDNA synthesis with Superscript II reverse transcriptase (Invitrogen).

Amplification, cloning, and sequencing

For each primer set, buffer conditions were determined using the OptiPrime PCR Optimization Kit (Stratagene, La Jolla, CA;

Table 1. Source of the cDNA used as PCR template for species sequenced in this study

Species	PCR template	Reference/source
Monosiga brevicollis	Plasmid DNA from cDNA library	King and Carroll 2001
ATCC 50818	Plasmid DNA from cDNA library	NK and SBC, unpublished
Clypeatula cooperensis	cDNA	Kevin Peterson[1]
Halichondria sp.	Excision product from cDNA library	April Hill[1]
Haliclona rubens	cDNA[2]	This study
Suberites fuscus	cDNA	This study
Aphrocallistes vastus	cDNA[3]	This study
Leucosolenia sp.	cDNA	This study
Scypha sp.	cDNA	This study
Nematostella vectensis	Plasmid DNA from cDNA library	J.F., unpublished data
Aurelia aurita	cDNA	This study
Mnemiopsis leidyi	Excision product from cDNA library	Mark Martindale[1]

[1]Personal communication.
[2]Source tissue provided by April Hill.
[3]Source tissue provided by Sally Leys.

Table 2. The combined amino acid data set and classification of taxa analyzed in this study

Classification	Code	Genus	α-tubulin	β-tubulin	EF-2	HSP90
A						
Choanoflagellata	1.0	*Monosiga*	◆	◆	◆	◆
		ATCC 50818	◆	◆	◆	◈
Porifera, Hexactinellida	2.1	*Aphrocallistes*	◆	◆	◆	◆
Porifera, Demospongiae	2.2	*Clypeatula*	◆	◆	◇	◈
		Halichondria	◆	◆	◆	◆
		Haliclona	◆	◆	◇	◈
		Suberites	◆	◆	◆	◆
Porifera, Calcarea	2.3	*Leucosolenia*	◆	◆	◆	◆
		Scypha	◆	◆	◆	◆
Cnidaria, Anthozoa	3.1	*Nematostella*	◆	◆	◆	◆
Cnidaria, Scyphozoa	3.2	*Aurelia*	◆	◆	◆	◈
Ctenophora, Lobata	4.0	*Mnemiopsis*	◆	◆	◇	◈
B						
Protostomia, Platyhelminthes	5.0	*Schistosoma*	◆	◆	◇	◈
Protostomia, Nematoda	5.0	*Caenorhabditis*	◆	◆	◆	◆
Protostomia, Arthropoda	5.0	*Drosophila*	◆	◆	◆	◆
Deuterstomia, Vertebrata	6.0	*Danio*	◆	◆	◇	◆
Deuterstomia, Vertebrata	7.0	*Mus*	◆	◆	◆	◆
Fungi, Ascomycota	7.0	*Saccharomyces*	◆	◆	◆	◆
Fungi, Ascomycota	7.0	*Schizosaccharomyces*	◆	◆	◆	◆

◆, Full length; ◈, partial; ◇ N/A. Taxa are numbered according to their classification. This coding scheme also applied to Figs. 1–5. (A) Fragments of the α-tubulin, β-tubulin, EF-2, and HSP90 genes were sequenced from diverse choanoflagellates and early branching Metazoa. For each species–gene combination, a full-length sequence is indicated with a filled diamond and failure to clone the gene is indicated with a white diamond. HSP90 was sequenced through cloning of two overlapping 5′ and 3′ gragments. Where cloning of only one portion of HSP90 was successful, the clone is indicated with a gray diamond and the portion of the gene (5′ and 3′) used for analysis. (B) Amino acid sequences for α-tubulin, β-tubulin, EF-2, and HSP90 (indicated with filled diamonds) were retrived form GenBank for five Bilateria (*Schistosoma*, *Caenorhabditis*, *Drosophila*, *Danio*, and *Mus*) and two fungi (*Saccharomyces* and *Schizosaccharomyces*). The absence of *Schistosoma* and *Danio* EF2 sequences in GenBank is indicated with white diamonds and the partial sequence of *Schistosoma* HSP90 is indicated in gray.

Table 3). Amplification reactions of 50 μl were assembled with 1 × OptiPrime buffer (Stratagene), 0.2 mM dNTPs, 0.28 μM each of forward and reverse primer, and 2.5 U *Taq* polymerase (Promega, Madison, WI). Reactions were first denatured at 94°C and then subjected to 35 cycles at 94°C for 30 sec, 52°C for 2 min, and 72°C for 1–1.5 min followed by a 10-min extension at 72°C. Primer sequences and optimal buffer conditions are shown in Table 3. PCR products were electrophoresed, excised from the gel, and

Table 3. PCR primers and optimal buffer conditions

Gene	Forward primer (5′–3′)	Reverse primer (5′–3′)	Approximate product size (base pairs)	Optimal buffer	Reference
α-tubulin	RGTNGGNAAYGCN TGYTGGGA	CCATNCCYTCNCCNA CRTACCA	1150	2	Edgcomb et al. (2001)
β-tubulin	GCAGGNCARTGYG GNAAYCA	AGTRAAYTCCATYTCRTCCAT	1160	6	Edgcomb et al. (2001)
HSP90 (5′)	GARACNTTYGCNTT YCARGC	ARRTGRTCYTCCCARTCRTT	880	4	This report
HSP90 (3′)	GARGARTAYGGNGARTTYTA	TCCATRTTNGCNGTC CANCCRTA	870	2	This report
EF-2 (3′)	GGNGCNGGNGARYT NCAYYTNGA	CCARTGRTCRAANACRCA YTGNGGRAA	760	1.5 mM $MgCl_2$	Hashimoto et al. (1995)

purified using QiaexII (Qiagen, Valencia, CA). Gel-purified DNA fragments with 3′-deoxyadenosine overhangs were inserted into pCRII-TOPO (Invitrogen), pCR4 (Invitrogen), or pGEM-T (Promega). Sequencing was carried out using BigDye Terminator chemistry (Perkin-Elmer, Boston, MA) on an ABI 377 sequencer. For each species, three to seven independent clones from each amplification reaction were sequenced from primers specific to the cloning vector. The full-length sequences were then used to build contigs using Sequencher 4.1 (Gene Codes, Ann Arbor, MI). For some species, a given gene was represented by multiple isoforms or alternative splicing products. In these cases, the predominant sequence was used for phylogenetic analysis. *Monosiga brevicollis* HSP70 sequences were isolated by random sequencing of clones from a cDNA library (King and Carroll 2001). All sequences are deposited in GenBank (AY226048–AY226092).

Phylogenetic analysis

Amino acid sequences were aligned using ClustalX (Thompson et al. 1997) with a GONNET weight matrix, a gap opening penalty value of 10, and a gap extension penalty of 0.2. Alignments for all data sets were robust to changes in weight matrix and gap opening/extension penalties and have very few insertions/deletions (indels). All amino acid sequences in all data sets were tested for significant amino acid composition deviations from the frequency distribution assumed by the maximum likelihood model using the chi-square test implemented in TREE-PUZZLE, version 5.0 (Strimmer et al. 1996) (data not shown). Sequences showing significant deviation were excluded.

Phylogenies were estimated using maximum likelihood as implemented in two different packages, TREE-PUZZLE (Strimmer et al. 1996) and PHYLIP (Felsenstein 1993). Maximum likelihood algorithms take into account important aspects of amino acid sequence evolution, such as rate heterogeneity and amino acid composition bias, which are known to affect phylogenetic reconstructions based on amino acid data (Thorne 2000) and have been shown to outperform parsimony and distance method algorithms (Huelsenbeck 1995a,b; Swofford et al. 2001). In TREE-PUZZLE, phylogenies were estimated using the quartet puzzling algorithm (10,000 puzzling steps) of Strimmer and von Haeseler (1996) and the JTT model (Jones et al. 1992) as implemented in the program TREE-PUZZLE 5.0 (Strimmer and von Haeseler 1996). Amino acid composition and rate heterogeneity among sites (discretely approximated using a Γ distribution with four rate categories) were estimated for each data set. In PHYLIP, phylogenies were estimated using a maximum likelihood algorithm and the JTT model of amino acid substitution (Jones et al. 1992) as implemented in the PROML program of the PHYLIP package, version 3.6a3 (Felsenstein 1993) on 100 bootstrap replicates. Rate heterogeneity among sites was discretely approximated using a Γ distribution with five rate categories. All the alignments (and their corresponding phylogenies) are available electronically from TreeBASE (www.herbaria.harvard.edu/treebase/, study accession number S852).

Bayesian inference holds great promise as a new statistical method for phylogenetic reconstruction (Huelsenbeck et al. 2001). However, in contrast with studies using maximum likelihood, its application to amino acid data sets has been limited. We analyzed our data sets using Bayesian inference as implemented in the program MrBayes, version 3.0b (Huelsenbeck and Ronquist 2001). Analyses used the JTT model of amino acid substitution; rate heterogeneity among sites was discretely approximated using a Γ distribution with four rate categories. Each Bayesian analysis was performed with four chains (one cold and three heated) of length 2×10^6 steps, with sampling occurring every 10^3 steps, yielding a total of 2000 observations (topologies). The first 10^5 steps (100 observations) were considered as the burn-in period and were excluded from subsequent analyses. When compared with the results obtained from the maximum likelihood analyses (see below), the results of the Bayesian inference analyses yield no additional insights into the evolutionary relationships between early branching Metazoa and are not shown.

RESULTS

Amino acid variation in the five protein data sets

Molecular markers for phylogenetics must be of sufficient length and variability to provide signal for use in phylogenetic analysis. If the rate of evolution for a particular gene is too

Table 4. Proportion (%) of variable and parsimony-informative sites in the five gene data sets according to four different taxon groupings

	Variable sites (%)				Parsimony-informative sites (%)			
Data set (no. of amino acids)	A	A-F	A-FH	CPCC	A	A-F	A-FH	CPCC
α-tubulin (394)	68.3	61.4	44.7	43.1	46.4	36.0	25.9	25.1
β-tubulin (392)	50.8	39.3	26.0	13.0	37.0	31.1	11.7	5.9
EF-2 (253)	58.5	52.2	49.4	45.5	41.1	34.0	31.2	29.6
HSP90 (530)	62.1	54.9	53.4	45.7	48.3	35.8	34.9	28.3
HSP70 (484)	51.7	46.9	45.7	43.6	31.6	26.4	25.2	23.3

A, All taxa; A-F, all taxa without Fungi; A-FH, all taxa without Fungi and Hexactinellida; CPCC, Choanoflagellata, Porifera (without *Aphrocallistes* for α- and β-tubulin), Cnidaria, and Ctenophora. Variable sites contain more than one type of amino acid. Parsimony-informative sites contain at least two types of amino acids that appear more than once in the data set.

slow for the question at hand, sequences from sampled taxa will not be sufficiently divergent for use in phylogenetic analysis. To test whether the five genes we selected for this study harbor ample variation, we measured the percentage of variable and parsimony-informative sites in data sets for α-tubulin, β-tubulin, EF-2, HSP90, and cytoplasmic HSP70. For each of the genes, 51–68% of the sites are variable and 32–48% are parsimony-informative (Table 4, column A). Because some of the variation derives from outgroup sequences, we also analyzed the data sets after removal of sequences from Fungi. Although exclusion of fungal sequences reduces variation somewhat, 39–61% of sites remain variable and 26–36% of sites parsimony-informative (Table 4, column A-F). Similar results were obtained if only early branching metazoans and choanoflagellates were considered (Table 4, column CPCC). The levels of variation observed in the α-tubulin, β-tubulin, EF-2, HSP90, and cytoplasmic HSP70 data sets validate their potential for phylogenetic analysis of early branching metazoans.

Independent analyses of five protein data sets fail to resolve the phylogenetic relationships of early branching Metazoa and choanoflagellates

The most striking results from independent phylogenetic analyses of four protein coding genes (α-tubulin, β-tubulin, EF-2, and HSP90) are the lack of topological resolution among early branching Metazoa and Choanoflagellata and the failure to reconstruct well-supported groups such as the Bilateria, Protostomia, and, on occasion, Deuterostomia (Fig. 1). Despite their inability to resolve the relationships between early branching metazoan phyla, each of the genes contributes some level of phylogenetic resolution for several groups currently assumed to be monophyletic (Fig. 1). For example, the two calcarean species (*Leucosolenia* and *Scypha*) group together consistently. Furthermore, the two choanoflagellate species used in this study are shown to be monophyletic and three (*Halichondria*, *Haliclona*, and *Suberites*) of the four Demospongiae species group together in most analyses. However, the fourth species, *Clypeatula*, does not group with the other demosponges, although a previous analysis has placed it firmly within Demospongiae (Peterson and Addis 2000).

One potential concern stems from the unusually long branch length exhibited by the Hexactinellid *Aphrocallistes* in the α- and β-tubulin data sets (Fig. 1A and B). Removal of *Aphrocallistes* from the α-tubulin data set decreases the percentage of parsimony-informative sites by a net total of 16.7% (Table 4, column A-FH), suggesting that the substitution rate within this lineage may be much higher than in others. Less severe examples of fast evolving sequences are the α-tubulins from two calcareans (*Scypha* and *Leucosolenia*, Fig. 1A). It is noteworthy that α-tubulin is the only single-gene data set that gives robust support for a sister-group relationship of Hexactinellida and Calcarea as an outgroup of the rest of Metazoa and choanoflagellates (Fig. 1A). However, unusually long branches tend to group together with robust support irrespective of the "true" underlying phylogeny, leading to long branch attraction (Felsenstein 1978; Philippe and Adoutte 1998; Philippe and Laurent 1998). In this study, because the outgroup fungal taxa form a long branch, any member of the ingroup with a higher substitution rate relative to the other taxa (as is the case for *Aphrocallistes* and Calcarea) may emerge near the base of the tree, artifactually attracted by the very long branch of the outgroup.

The surprising lack of resolution from analyses of α-tubulin, β-tubulin, EF-2, and HSP90 and the aberrant be-havior of *Aphrocallistes* sequences in phylogenetic analysis raise the possibility, however unlikely, that the combination of species selected for study (Table 2A) may prove unusually resistant to phylogenetic analysis. Fortunately, an entirely different selection of early branching Metazoa and a choanoflagellate (*Monosiga ovata*) have been sampled for a fifth protein, cytoplasmic HSP70 (Borchiellini et al. 1998; Snell et al. 2001). In addition, we recovered sequences of three different HSP70

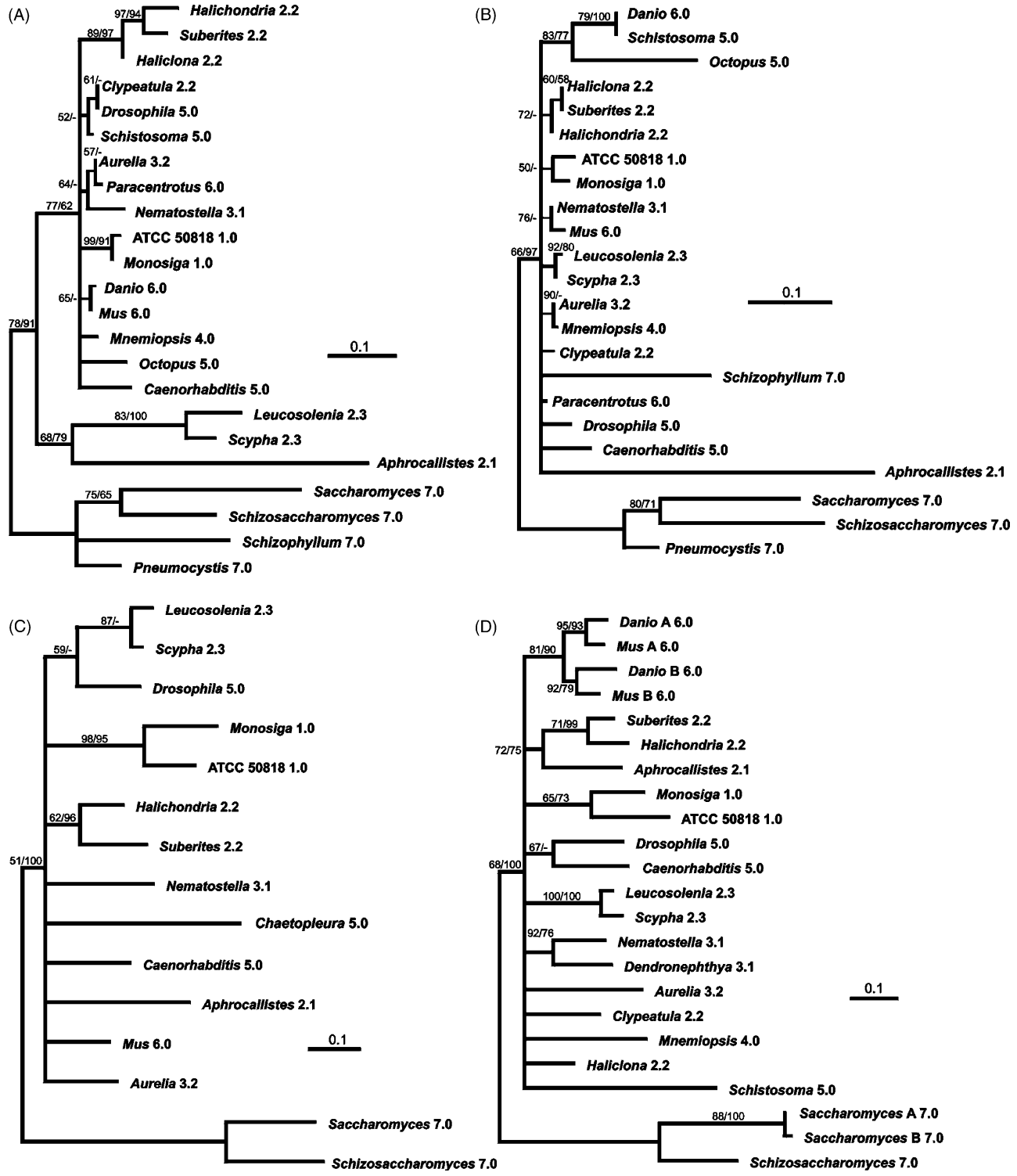

Fig. 1. Lack of resolution and conflicting tree topologies from single gene analyses. Phylogenetic analyses of α-tubulin (A), β-tubulin (B), EF-2 (C), and HSP90 (D) from choanoflagellates, diverse Metazoa, and Fungi produced conflicting results and left the relationships between early branching phyla unresolved. Phylogenetic reconstruction: phylograms estimated using the quartet puzzling search algorithm used by TREE-PUZZLE. Branches with 50% quartet puzzling support are collapsed. The quartet puzzling support values (estimated using TREE-PUZZLE) are indicated on the left of the slash and bootstrap support values > 50% (estimated using PROML) are indicated on the right. Branches not supported by PROML are indicated with a dash. Classifications of taxa are indicated using the coding scheme from Table 2. For example, members of Protostomia (*Schistosoma*, *Drosophila*, and *Caenorhabditis*) are indicated by the number 5.0.

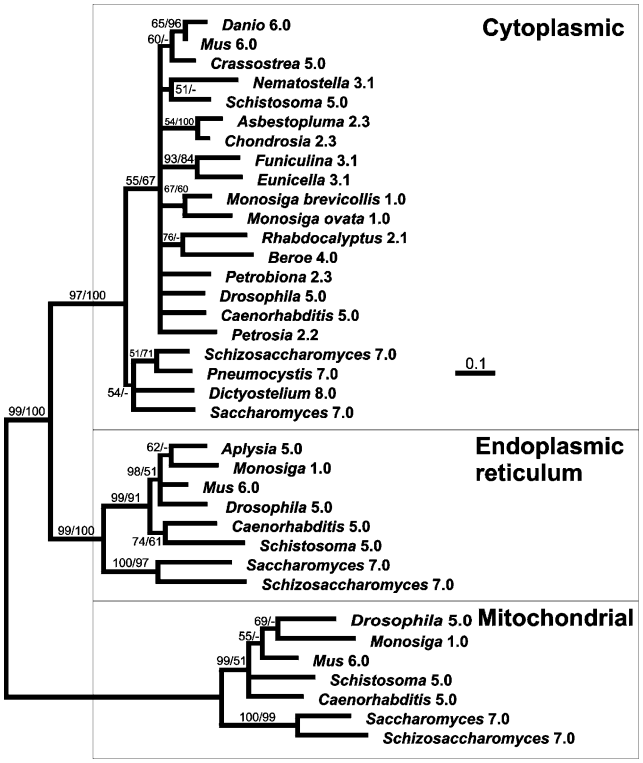

Fig. 2. Analysis of the three HSP70 paralogs (cytoplasmic, ER, and mitochondrial) failed to resolve the phylogenetic relationships of choanoflagellates, early branching Metazoa, and Bilateria. Note the robust support for a choanoflagellate/metazoan clade inferred from all three HSP70 paralogs and the lack of resolution within that clade inferred from the densely sampled cytoplasmic HSP70 paralog. Phylogram construction and coding scheme as described in legend to Fig. 1. (For example, the hexactinellid of this data set, i.e., *Rhabdocalyptus*, is indicated by the number 2.1.)

paralogs from a second choanoflagellate, *Monosiga brevicollis*. Phylogenetic analysis of a complete HSP70 data set containing metazoan, choanoflagellate, and fungal sequences from the three paralog groups recovers three subtrees representing the cytoplasmic, ER, and mitochondrial HSP70 families (Fig. 2; see Materials and Methods for data set construction). The topological relationships within each of the three paralog subtrees suggest a close affinity between Metazoa and choanoflagellates to the exclusion of Fungi, agreeing with our analyses of α-tubulin, β-tubulin, EF-2, and HSP90 and those reported in recently published analyses (King and Carroll 2001; Snell et al. 2001). Importantly, this analysis also fails to resolve the relationships between early branching metazoan phyla and choanoflagellates, in support of findings from previous reports (Borchiellini et al. 1998; Snell et al. 2001) (Fig. 2). Therefore, independent analyses of five different protein sequences from two different well-sampled sets of taxa suggest that single-gene analyses lack the power to uncover the relationships between choanoflagellates, sponges, cnidaria, and ctenophores.

Analyses of alternative combined data sets reveal the impact of long branch attraction within an otherwise unresolved metazoan tree

Phylogenetic analysis of single genes is often plagued by the effects of taxon sampling, unequal substitution rates among lineages, and low signal-to-noise ratio; it has been suggested that these problems may be overcome by combining individual gene data sets (Baldauf et al. 2000). Therefore, we generated a larger data set by concatenating sequences from α-tubulin, β-tubulin, EF-2, and HSP90, including only those taxa for which at least three of the four genes were available (shown in Table 2). Sequences from the cytoplasmic HSP70 gene were not included due to differences in taxon sampling.

The resulting phylogenetic analysis of the combined data set is shown in Figure 3A and, in general agreement with the single gene studies, is characterized by lack of resolution. However, the topology of Figure 3A suggests that the earliest branching taxon is the hexactinellid *Aphrocallistes* with the class Calcarea as the second earliest branch. If true, such a topology implies that sponges are of paraphyletic (or polyphyletic) origin and Choanoflagellata are single-celled metazoans. However, Hexactinellida and Calcarea are the two groups that show unusually long branches in the α- and β-tubulin data sets (Fig. 2A and B), and as discussed above, this grouping may be due to long branch attraction.

To test whether long branch attraction contributes to the topology inferred from the combined data set, we removed sequences from the rapidly evolving *Aphrocallistes*. Exclusion of *Aphrocallistes* from the data set restores the polytomy observed in the single gene analyses (Fig. 3B). We suggest that the early branching position of *Aphrocallistes* and Calcarea in the combined data set is an artifact of long-branch attraction.

The effect of taxon selection in phylogenetic analysis of early branching Metazoa

In phylogenetic analyses, the inclusion or exclusion of certain taxa may influence topology inference. To test the effect of taxon selection on the phylogenetic placement of choanoflagellates, we assembled concatenated sequences from α-tubulin, β-tubulin, EF-2, and HSP90 for four sets of taxa in which only one representative of Ctenophora (*Mnemiopsis*; Fig. 4A), Cnidaria (*Nematostella*; Fig. 4B), Demospongiae (*Suberites*; Fig. 4C), or Calcarea (*Leucosolenia*; Fig. 4D) was included. Each analysis included a core set of taxa composed of one choanoflagellate (*Monosiga*), five Bilaterians (Table 2B), and two Fungi (Table 2B) as the outgroup. Although resolution is not dramatically improved compared with the complete data set (Fig. 3A and B), and groups such as the Bilateria and Protostomia are not recovered, inclusion of a representative ctenophore, cnidarian, or demosponge (Fig. 4 A–C) gives support (in some cases weak) for *Monosiga* as the outgroup to Metazoa. In contrast, inclusion of a calcareous

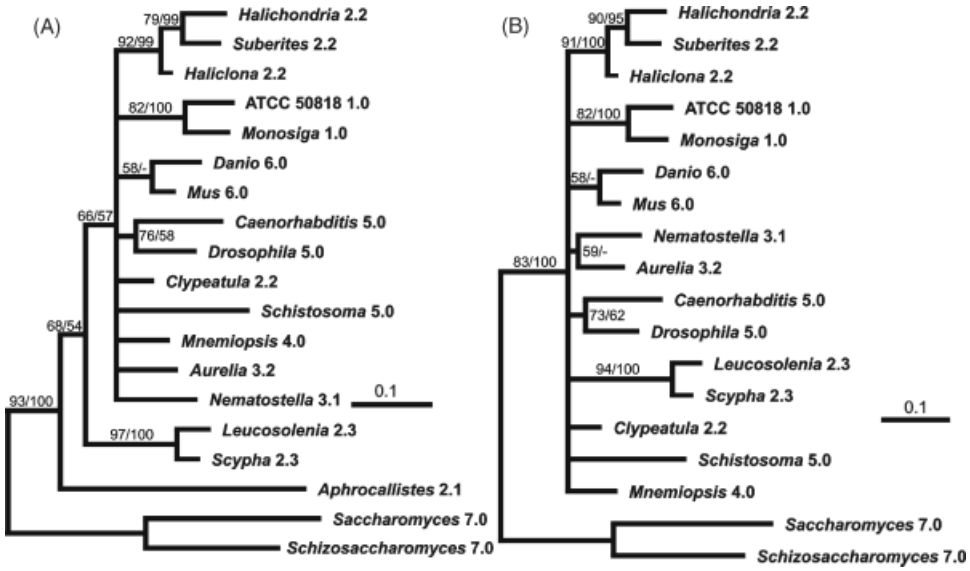

Fig. 3. Concatenation of α-tubulin, β-tubulin, EF-2, and HSP90 fails to resolve the phylogenetic relationships between choanoflagellates and early branching Metazoa. Inclusion (A) or exclusion (B) of sequences from *Aphrocallistes* reveals the bias imposed by long branch attraction. Phylogram construction and coding scheme as described in legend to Fig. 1.

sponge (Fig. 4D) supports a different branch topology in which Calcarea forms the outgroup of a metazoan/choanoflagellate clade. The alternative topologies obtained from this exercise suggest that the phylogenetic placement of choanoflagellates is sensitive to the sampling of different metazoan taxa. These results could have been interpreted as evidence that choanoflagellates are an outgroup to Metazoa (Fig. 4A–C) or that Calcarea diverged earlier than Choanoflagellata (Fig. 4D). However, further analyses (Figs. 1, 3, and 5) indicate that neither phylogenetic hypothesis is robustly supported.

Lack of resolution in combined data analyses stems from conflicting phylogenetic signal between individual gene trees

Wide taxonomic sampling of multiple protein-coding genes has proven useful for resolving the relationships between other "deeply" branching lineages (Gupta 1995; Baldauf et al. 2000; Fast et al. 2002). It is therefore somewhat surprising that our results from both single-gene and combined data set analyses show a lack of resolution (Fig. 3A and B). Three factors may contribute to the unresolved topology of the metazoan tree. First, the lack of resolution may be due to conflicts in the phylogenetic signals obtained from different genes. Alternatively, a scenario in which sponges, ctenophores, cnidarians, and Bilateria evolved simultaneously, or near simultaneously, would result in a "hard polytomy" thought to be refractory to phylogenetic analysis (Maddison 1989). Finally, the inclusion of taxa from many early branching phyla may create many short internal branches that are difficult to resolve by existing analytical methods and the data at hand, leading to a "soft polytomy." A way to distinguish between soft and hard polytomies involves sampling less densely from closely related taxa; the inferred topology would be unresolved for a hard polytomy but resolved for a soft polytomy (Jackman et al. 1999).

To test between these alternatives, we analyzed concatenated sequences of all four genes from a taxon subset composed of five Bilateria (Table 2B), a choanoflagellate, and single representatives from Ctenophora, Cnidaria, Demospongiae, and Calcarea (*Mnemiopsis*, *Nematostella*, *Suberites*, and *Leucosolenia*). Analysis of this reduced data set fails to resolve interrelationships among metazoan taxa (Fig. 5A) and suggests either a hard polytomy or conflicting phylogenetic signal. To determine whether the latter might be influencing the resulting topology (Fig. 5A), we analyzed sequences for single genes (α-tubulin, Fig. 5B; and HSP90, Fig. 5C) from the same set of taxa. Although the topologies of the α-tubulin and HSP90 trees are fairly well resolved, they are in conflict with each other. Therefore, conflicts in phylogenetic signal contribute to the lack of resolution in the combined data set analyses and interfere with elucidation of the nature of the polytomy in our analyses.

DISCUSSION

This study of the base of the metazoan tree analyzes the largest number of protein-coding genes from the broadest sampling of early branching Metazoa to date. Phylogenetic

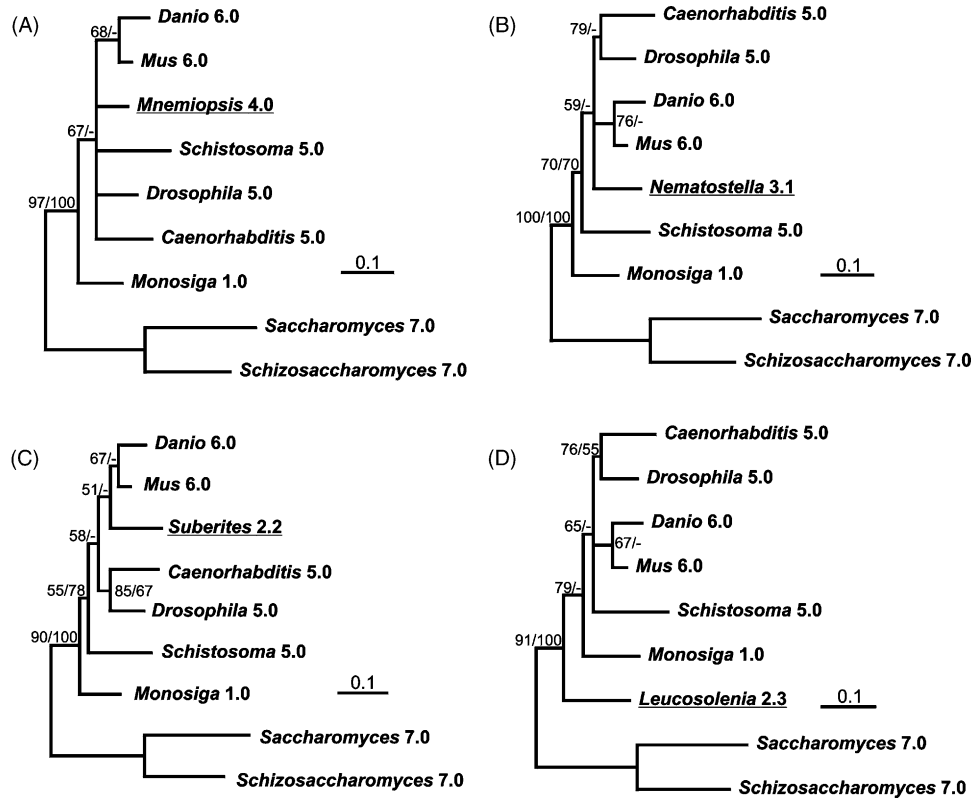

Fig. 4. Taxon selection influences the topology inferred from data sets with reduced taxonomic diversity. Maximum likelihood analysis was performed on concatenated sequences (α-tubulin, β-tubulin, EF-2, and HSP90) from a choanoflagellate (*Monosiga*), five bilaterians, and a representative (underlined) from either Ctenophora (*Mnemiopsis*, A), Cnidaria (*Nematostella*, B), Demospongiae (*Suberites*, C) or Calcarea (*Leucosolenia*, D). Analyses including a ctenophore, cnidarian, or demosponge support the placement of a choanoflagellate as an outgroup of Metazoa (A–C). In contrast, a data set containing sequences from a calcareous sponge suggests that the Calcarea diverged earlier than the choanoflagellates (D). Phylogram construction and coding scheme as described in legend to Fig. 1.

analyses of five protein-coding genes fail to reveal either the exact placement of Choanoflagellata or the evolutionary relationships between early branching metazoan taxa. Multiple alternative data sets suggest that not one of the genes studied contains sufficient phylogenetic signal to resolve the topology of metazoans; furthermore, analyses of different genes support conflicting phylogenetic groupings. We show that variable rates of evolution between lineages and sensitivity of the analyses to taxon selection can cause serious topological artifacts (e.g., the artifactual early branching placement of *Aphrocallistes* and Calcarea in analyses of α-tubulin and the combined data set due to long branch attraction). The conflict in phylogenetic signal derived from different genes, and the sensitivity of phylogenetic analyses to taxon selection, may reveal obstacles to topological inference of deep phylogenies using molecular sequence data.

Inference of phylogenetic relationships between distantly related taxa may be compromised by both biological and analytical factors. Analytical factors that influence phylogenetic analysis stem from shortcomings of the techniques and methods used in current phylogenetic practice (e.g., limitations of models of sequence evolution). Biological factors include forces that act on species, both at the molecular (e.g., mutation and recombination) and the population level (e.g., genetic drift, natural selection, speciation, and extinction). Accounting for the effect of biological factors on the evolution of molecular sequences may require innovative approaches to phylogeny reconstruction.

Biological factors affecting phylogenetic reconstruction

Phylogenetic reconstruction of deeply diverged taxa, such as the early branching metazoans, rests on the presence and availability of highly conserved genes that can be accurately identified as orthologs. Such strong conservation of amino acid sequences over long time periods can only be explained by invoking the action of strong purifying selection; in the absence of selection, genetic drift is expected to lead to wide

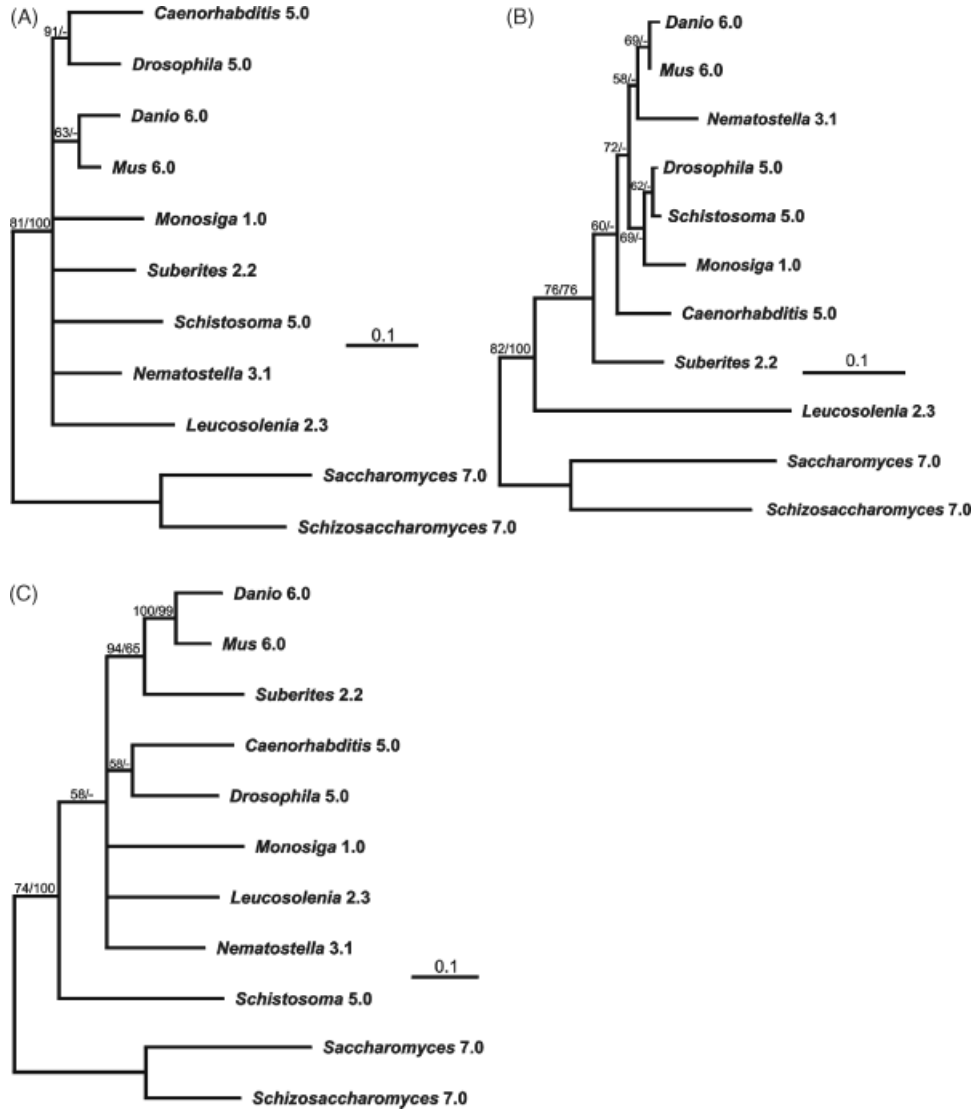

Fig. 5. The lack of resolution in the combined data set stems, in part, from conflicting phylogenetic signal in the individual gene data sets. Maximum likelihood analysis was performed on concatenated sequences from a choanoflagellate (*Monosiga*), a demosponge (*Suberites*), a calcarean (*Leucosolenia*), a cnidarian (*Nematostella*), and five bilaterians (A); maximum likelihood analyses performed for the same set of taxa with α-tubulin (B) and HSP90 (C) data sets support conflicting topologies. Phylogram construction and coding scheme as described in legend to Fig. 1.

divergence between orthologs, making them indiscernible after a few million generations (Gillespie 1991). Despite this realization, the assumption behind every phylogenetic analysis is that variation between sequences is neutral. Indeed, models of sequence evolution and phylogenetic reconstruction algorithms do not specifically take natural selection into account (Swofford et al. 1996). However, if only sites of lesser functional importance are allowed to vary (i.e., purifying selection on these sites is not too strong), the expectation is that over a large amount of time, the phylogenetic signal of these sites may be lost due to substitutional saturation. As a consequence, the differences between sequences that are used for phylogenetic inference may not accurately reflect the evolutionary history of the taxa under consideration.

The discriminatory power of the data at hand in relation to the time window of divergence between the taxa of interest may additionally limit the phylogenetic reconstruction of evolutionary relationships (Fig. 6). Indeed, work on 18S rDNA has shown that branchings in the metazoan tree separated by less than 40 million years cannot be confidently resolved (Philippe et al. 1994). For example, if the divergence of choanoflagellates, sponges, cnidarians, ctenophores, and

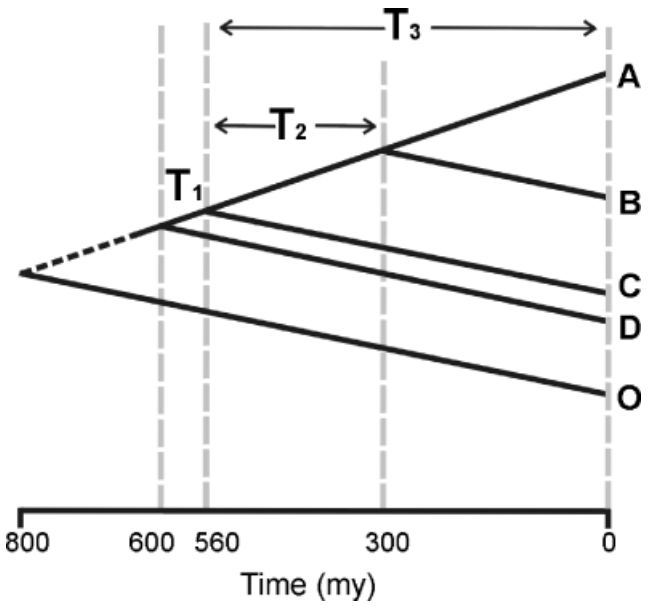

Fig. 6. The evolutionary histories of early branching taxa may limit the discriminatory power of molecular sequence data for phylogenetic analysis. When attempting to infer the evolutionary relationships between four taxa of interest (A, B, C, and D with taxon O as the outgroup), the strength of the phylogenetic signal for each phylogenetic association will be strongly influenced by the time between successive speciation events. For instance, resolution of (A, B) clade will be easier than resolution of the (A, B, C) clade from D, because T2 $\gg$ T1, allowing for accumulation of a much larger amount of phylogenetic information. Resolution is further influenced by the ratio of the time *before* divergence (T1) over time *since* divergence (T3). If the latter (T3) is much larger than T1, the effects of homoplasy and substitutional saturation may obscure the real associations.

bilaterians occurred over a 50-million-year window some 600 million years ago, robust resolution of their evolutionary relationships may not be feasible, at least with currently available data.

Discordance between individual gene phylogenies can also affect the effort to reconstruct deep phylogenies. Although this situation may arise due to analytical factors (e.g., not accounting for rate heterogeneity among sites within a gene; Yang 1996), a number of examples of conflict between different gene trees may arise from biological factors, such as through adaptive selection-driven convergence (Nei and Hughes 1991) or from lateral gene transfer (Keeling and Palmer 2001). In cases where the interval between successive speciation events is short, lineage sorting of ancestral polymorphisms may also prove important (Hey and Kliman 1993; Rokas et al. 2003). Furthermore, recent work has revealed high rates of gene duplications in populations (Fryxell 1996; Lynch and Conery 2000). The presence of unrecognized paralogs stemming from such gene duplication events can have a pervasive influence in phylogenetic analyses (Telford and Holland 1997; Martin and Burg 2002), and paralogous genes have been shown for both rDNA and protein-coding genes (Telford and Holland 1997; Israel et al. 2002; Martin and Burg 2002). In summary, if different regions of a genome represent different evolutionary histories, no single gene or character may capture all the nuances of an organism's phylogenetic relationships (Maddison 1997).

Analytical factors affecting phylogenetic reconstruction

Analytical factors such as limitations of models of amino acid sequence evolution, taxon sampling, and selection bias are known to influence phylogenetic reconstruction. Models of amino acid evolution are designed to capture the most essential aspects of processes shaping amino acid evolution. However, these models rely on assumptions that render them sensitive to a number of influences. For example, extreme variability in rates of evolution among lineages has been shown to lead to long branch attraction (Aguinaldo et al. 1997; Philippe and Adoutte 1998; Philippe and Laurent 1998; Philippe 2000), an artifact we observe in our analyses.

A considerable body of work has addressed the effect of taxon sampling on phylogenetic reconstruction, especially in relation to density of sampling (Hillis 1996; Zwickl and Hillis 2002). Whereas randomly chosen taxa are sometimes pruned from analyses to reduce the computation time required, it has also been common practice to specifically exclude taxa showing a high rate of evolution (Aguinaldo et al. 1997; Kim et al. 1999). Whether intentional or not, such practices may alter the outcomes of phylogenetic analyses; indeed, in the case of taxon selection, altering the outcome to "improve" phylogenetic resolution is the motivation. If we had chosen to analyze a data set with a demosponge and a choanoflagellate, we would have concluded that choanoflagellates are an outgroup of Metazoa. Denser sampling of sponges shows that any conclusion regarding the exact placement of choanoflagellates is unwarranted. Therefore, although taxon selection may aid the phylogenetic resolution of certain phyla by allowing exclusion of fast-evolving species from a data set, it should be practiced with caution.

CONCLUSIONS

The robust reconstruction of metazoan history has proven to be a difficult task (Philippe et al. 1994; Kumar and Rzhetsky 1996; Van de Peer and De Wachter 1997). However, given that a large number of data sources and strategies are unexplored, we remain optimistic about the potential success of this endeavor. Three factors will strongly influence the effort to reconstruct the evolutionary relationships between metazoan phyla: a vast increase in data availability, better and

faster phylogenetic algorithms, and a realistic assessment of the task at hand. Although this study suggests that sequence data from individual, or even concatenated, gene sequences may prove phylogenetically uninformative, complete genome sequences from diverse metazoans may allow the exploration of different types of characters with promise for phylogenetics, such as gene order, signature sequences, intron insertions, and differences in chromosome structure (reviewed in Rokas and Holland 2000). Additionally, particular emphasis should be given to improving the representation from less well-studied metazoan phyla. This flood of genomic data will need to be accompanied by the development of new algorithms for efficient analysis of large data sets, a process that has already started (Bininda-Emonds et al. 2001; Lemmon and Milinkovitch 2002). The final, and perhaps most difficult, aspect of this effort is the need for caution in interpreting and reporting phylogenetic results. To expect that information contained in any single gene will resolve without ambiguity branching events in a taxonomic group with a more than 600 million year history is simply unrealistic. However, by shifting the focus to lessons learned from the failures of single gene studies, we may uncover more successful strategies for reconstructing the evolutionary history of deeply diverging taxa.

Acknowledgments

We are tremendously grateful to April Hill, David Jacobs, Sally Leys, Mark Martindale, Kevin Peterson, and Babcock Hall, each of whom provided unpublished reagents used in this study. In addition, we acknowledge the operators of ROPOS, a remotely operated vehicle used for the collection of *Aphrocallistes* tissue by Sally Leys. April Hill was instrumental in guiding the selection of sponge taxa for study and John Holt in providing computing support. We also thank Jayne Selenue for technical help, Leanne Olds for help with the graphics, Nipam Patel for providing potential cover images, and David Baum and John Huelsenbeck for discussion on phylogenetics and Bayesian inference, respectively. A. R. is supported by a Human Frontier Science Program long-term fellowship. N. K. is supported by an National Institutes of Health postdoctoral fellowship. This work was supported by the Howard Hughes Medical Institute.

REFERENCES

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.

Atkins, M. S., McArthur, A. G., and Teske, A. P. 2000. Ancyromonadida: a new phylogenetic lineage among the protozoa closely related to the common ancestor of metazoans, fungi, and choanoflagellates (Opisthokonta). *J. Mol. Evol.* 51: 278–285.

Ax, P. 1996. *Multicellular Animals: A New Approach to the Phylogenetic Order in Nature*. Springer-Verlag, Berlin.

Baldauf, S. L., and Palmer, J. D. 1993. Animals and fungi are each other's closest relatives: congruent evidence from multiple proteins. *Proc. Natl. Acad. Sci. USA* 90: 11558–11562.

Baldauf, S. L., Roger, A. J., Wenk-Siefert, I., and Doolittle, W. F. 2000. A kingdom-level phylogeny of eukaryotes based on combined protein data. *Science* 290: 972–977.

Bininda-Emonds, O. R., Brady, S. G., Kim, J., and Sanderson, M. J. 2001. Scaling of accuracy in extremely large phylogenetic trees *Pac. Symp. Biocomput* 547–558.

Boorstein, W. R., Ziegelhoffer, T., and Craig, E. A. 1994. Molecular evolution of the HSP70 multigene family. *J. Mol. Evol.* 38: 1–17.

Borchiellini, C., Boury-Esnault, N., Vacelet, J., and Le Parco, Y. 1998. Phylogenetic analysis of the Hsp70 sequences reveals the monophyly of Metazoa and specific phylogenetic relationships between animals and fungi. *Mol. Biol. Evol.* 15: 647–655.

Borchiellini, C., Manuel, M., Alivon, E., Boury-Esnault, N., Vacelet, J., and Le Parco, Y. 2001. Sponge paraphyly and the origin of Metazoa. *J. Evol. Biol.* 14: 171–179.

Brusca, R. C., and Brusca, G. J. 1990. *Invertebrates*. Sinauer, Sunderland, MA.

Cavalier-Smith, T. 1987. The origin of Fungi and pseudofungi. In A. D. M. Rayner, C. M. Brasier, and D. Moore, (eds.). *Evolutionary Biology of the Fungi*. Cambridge University Press, Cambridge, pp. 339–354.

Collins, A. G. 1998. Evaluating multiple alternative hypotheses for the origin of Bilateria: an analysis of 18S rRNA molecular evidence. *Proc. Natl. Acad. Sci. USA* 95: 15458–15463.

Edgcomb, V. P., Roger, A. J., Simpson, A. G., Kysela, D. T., and Sogin, M. L. 2001. Evolutionary relationships among "jakobid" flagellates as indicated by alpha- and beta-tubulin phylogenies. *Mol. Biol. Evol.* 18: 514–522.

Fast, N. M., Xue, L., Bingham, S., and Keeling, P. J. 2002. Re-examining alveolate evolution using multiple protein molecular phylogenies. *J. Eukaryot. Microbiol.* 49: 30–37.

Felsenstein, J. 1978. Cases in which parsimony and compatibility methods will be positively misleading. *Syst. Zool.* 27: 401–410.

Felsenstein, J. 1993. *PHYLIP (Phylogeny Inference Package)*. Distributed by the Author, Department of Genetics, University of Washington, Seattle.

Fryxell, K. J. 1996. The coevolution of gene family trees. *Trends Genet.* 12: 364–369.

Galtier, N., and Gouy, M. 1995. Inferring phylogenies from DNA sequences of unequal base compositions. *Proc. Natl. Acad. Sci. USA.* 92: 11317–11321.

Germot, A., and Philippe, H. 1999. Critical analysis of eukaryotic phylogeny: a case study based on the HSP70 family. *J. Eukaryot. Microbiol.* 46: 116–124.

Gillespie, J. H. 1991. *The Causes of Molecular Evolution*. Oxford University Press, Oxford.

Gupta, R. S. 1995. Phylogenetic analysis of the 90 kD heat shock family of protein sequences and an examination of the relationship among animals, plants, and fungi species. *Mol. Biol. Evol* 12: 1063–1073.

Hasegawa, M., and Hashimoto, T. 1993. Ribosomal RNA trees misleading. *Nature* 361: 23.

Hashimoto, T., Nakamura, Y., Kamaishi, T., Nakamura, F., Adachi, J., Okamoto, K., and Hasegawa, M. 1995. Phylogenetic place of mitochondrion-lacking protozoan, *Giardia lamblia,* inferred from amino acid sequences of elongation factor 2. *Mol. Biol. Evol.* 12: 782–793.

Hey, J., and Kliman, R. M. 1993. Population genetics and phylogenetics of DNA sequence variation at multiple loci within the *Drosophila melanogaster* species complex. *Mol. Biol. Evol.* 10: 804–822.

Hibberd, D. J. 1975. Observations on the ultrastructure of the choanoflagellate *Codosiga botrytis* (Ehr.) Saville-Kent with special reference to the flagellar apparatus. *J. Cell Sci.* 17: 191–219.

Hillis, D. M. 1996. Inferring complex phylogenies. *Nature* 383: 130–131.

Huelsenbeck, J. P. 1995a. Performance of phylogenetic methods in simulation. *Syst. Biol.* 44: 17–48.

Huelsenbeck, J. P. 1995b. The robustness of two phylogenetic methods: four-taxon simulations reveal a slight superiority of maximum likelihood over neighbor joining. *Mol. Biol. Evol.* 12: 843–849.

Huelsenbeck, J. P., and Ronquist, F. 2001. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* 17: 754–755.

Huelsenbeck, J. P., Ronquist, F., Nielsen, R., and Bollback, J. P. 2001. Bayesian inference of phylogeny and its impact on evolutionary biology. *Science* 294: 2310–2314.

Israel, R. L., Kosakovsky Pond, S. L., Muse, S. V., and Katz, L. A. 2002. Evolution of duplicated alpha-tubulin genes in ciliates. *Evolution* 56: 1110–1122.

Jackman, T. R., Larson, A., de Queiroz, A., and Losos, J. B. 1999. Phylogenetic relationships and tempo of early diversification in *Anolis* lizards. *Syst. Biol.* 48: 254–285.

James-Clark, H. 1866. Note on the infusoria flagellata and the spongiae ciliatae. *Am. J. Sci.* 1: 113–114.

Jones, D. T., Taylor, W. R., and Thornton, J. M. 1992. The rapid generation of mutation data matrices from protein sequences. *Bioinformatics* 8: 275–282.

Keeling, P. J., Luker, M. A., and Palmer, J. D. 2000. Evidence from beta-tubulin phylogeny that microsporidia evolved from within the fungi. *Mol. Biol. Evol.* 17: 23–31.

Keeling, P. J., and Palmer, J. D. 2001. Lateral transfer at the gene and subgenic levels in the evolution of eukaryotic enolase. *Proc. Natl. Acad. Sci. USA* 98: 10745–10750.

Kim, J., Kim, W., and Cunningham, C. W. 1999. A new perspective on lower metazoan relationships from 18S rDNA sequences. *Mol. Biol. Evol.* 16: 423–427.

King, N., and Carroll, S. B. 2001. A receptor tyrosine kinase from choanoflagellates: molecular insights into early animal evolution. *Proc. Natl. Acad. Sci. USA* 98: 15032–15037.

Kruse, M., Leys, S. P., Muller, I. M., and Muller, W. E. 1998. Phylogenetic position of the Hexactinellida within the phylum Porifera based on the amino acid sequence of the protein kinase C from *Rhabdocalyptus dawsoni*. *J. Mol. Evol.* 46: 721–728.

Kumar, S., and Rzhetsky, A. 1996. Evolutionary relationships of eukaryotic kingdoms. *J. Mol. Evol.* 42: 183–193.

Lang, B. F., O'Kelly, C., Nerad, T., Gray, M. W., and Burger, G. 2002. The closest unicellular relatives of animals. *Curr. Biol.* 12: 1773–1778.

Lemmon, A. R., and Milinkovitch, M. C. 2002. The metapopulation genetic algorithm: an efficient solution for the problem of large phylogeny estimation. *Proc. Natl. Acad. Sci. USA* 99: 10516–10521.

Loytynoja, A., and Milinkovitch, M. C. 2001. Molecular phylogenetic analyses of the mitochondrial ADP-ATP carriers: the Plantae/Fungi/Metazoa trichotomy revisited. *Proc. Natl. Acad. Sci. USA* 98: 10202–10207.

Lynch, M., and Conery, J. S. 2000. The evolutionary fate and consequences of duplicate genes. *Science* 290: 1151–1155.

Maddison, W. P. 1989. Reconstructing character evolution on polytomous cladograms. *Cladistics* 5: 365–377.

Maddison, W. P. 1997. Gene trees in species trees. *Syst. Biol.* 46: 523–536.

Martin, A. P., and Burg, T. M. 2002. Perils of paralogy: using HSP70 genes for inferring organismal phylogenies. *Syst. Biol.* 51: 570–587.

Martindale, M., Finnerty, J., and Henry, J. 2002. The Radiata and the evolutionary origins of the bilaterian body plan *Mol. Phylog. Evol.* 24: 358–365.

Medina, M., Collins, A. G., Silberman, J. D., and Sogin, M. L. 2001. Evaluating hypotheses of basal animal phylogeny using complete sequences of large and small subunit rRNA. *Proc. Natl. Acad. Sci. USA* 98: 9707–9712.

Moreira, D., Le Guyader, H., and Philippe, H. 2000. The origin of red algae and the evolution of chloroplasts. *Nature* 405: 69–72.

Nei, M., and Hughes, A. L. 1991. Polymorphism and evolution of the major histocompatibility complex loci in mammals. In R. K. Selander, A. G. Clark, and T. S. Whittam (eds.). *Evolution at the Molecular Level*. Sinauer, Sunderland, MA, pp. 222–247.

Nielsen, C. 2001. *Animal Evolution*. Oxford University Press, Oxford.

Peterson, K. J., and Addis, J. S. 2000. *Clypeatula cooperensis* gen. n., sp. n., a new freshwater sponge (Porifera, Spongillidae) from the Rocky Mountains of Montana, USA. *Zool. Scr.* 29: 265–274.

Philippe, H. 2000. Long branch attraction and protist phylogeny. *Protist* 151: 307–316.

Philippe, H., and Adoutte, A. 1998. The molecular phylogeny of Eukaryota: solid facts and uncertainties. In G. H. Coombs, K. Vickerman, M. A. Sleigh, and A. Warren (eds.). *Evolutionary Relationships among Protozoa*. Chapman & Hall, London, pp. 25–56.

Philippe, H., Chenuil, A., and Adoutte, A. 1994. Can the Cambrian explosion be inferred through molecular phylogeny? *Development* (Suppl): 15–25.

Philippe, H., and Laurent, J. 1998. How good are deep phylogenetic trees. *Curr. Opin. Genet. Dev.* 8: 616–623.

Podar, M., Haddock, S. H., Sogin, M. L., and Harbison, G. R. 2001. A molecular phylogenetic framework for the phylum Ctenophora using 18S rRNA genes. *Mol. Phylog. Evol.* 21: 218–230.

Rieger, R., and Weyrer, S. 1998. The evolution of the lower metazoa: evidence from the phenotype. In W. E. G. Muller (ed.). *Progress in Molecular and Subcellular Biology*. Springer-Verlag, Berlin, pp. 21–43.

Rokas, A., and Holland, P. W. H. 2000. Rare genomic changes as a tool for phylogenetics. *Trends Ecol. Evol.* 15: 454–459.

Rokas, A., Melika, G., Abe, Y., Nieves-Aldrey, J.-L., Cook, J. M., and Stone, G. N. 2003. Lifecycle closure, lineage sorting and hybridization revealed in a phylogenetic analysis of European oak gallwasps (Hymenoptera: Cynipidae Cynipini) using mitochondrial sequence data. *Mol. Phylog. Evol.* 26: 36–45.

Saville-Kent, W. 1880–1882. *A Manual of the Infusoria*. David Bogue, London.

Snell, E. A., Furlong, R. F., and Holland, P. W. 2001. Hsp70 sequences indicate that choanoflagellates are closely related to animals. *Curr. Biol.* 11: 967–970.

Sogin, M. L., and Silberman, J. D. 1998. Evolution of the protists and protistan parasites from the perspective of molecular systematics. *Int. J. Parasitol.* 28: 11–20.

Strimmer, K., Goldman, N., and vonHaeseler, A. 1996. Bayesian probabilities and quartet puzzling. *Mol. Biol. Evol.* 14: 210–211.

Strimmer, K., and von Haeseler, A. 1996. Quartet puzzling: a quartet maximum-likelihood method for reconstructing tree topologies. *Mol. Biol. Evol.* 13: 964–969.

Swofford, D. L., Olsen, G. J., Waddell, P. J., and Hillis, D. M. 1996. Phylogenetic inference. In D. M. Hillis, C. Moritz, and B. K. Mable (eds.). *Molecular Systematics*. Sinauer, Sunderland, MA, pp. 407–514.

Swofford, D. L., Waddell, P. J., Huelsenbeck, J. P., Foster, P. G., Lewis, P. O., and Rogers, J. S. 2001. Bias in phylogenetic estimation and its relevance to the choice between parsimony and likelihood methods. *Syst. Biol.* 50: 525–539.

Telford, M. J., and Holland, P. W. H. 1997. Evolution of 28S ribosomal DNA in chaetognaths: duplicate genes and molecular phylogeny. *J. Mol. Evol.* 44: 135–144.

Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F., and Higgins, D. G. 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* 25: 4876–4882.

Thorne, J. L. 2000. Models of protein sequence evolution and their applications. *Curr. Opin. Genet. Dev.* 10: 602–605.

Van de Peer, Y., and De Wachter, R. 1997. Evolutionary relationships among the eukaryotic crown taxa taking into account site-to-site rate variation in 18S rRNA. *J. Mol. Evol.* 45: 619–630.

Wainright, P. O., Hinkle, G., Sogin, M. L., and Stickel, S. K. 1993. Monophyletic origins of the metazoa: an evolutionary link with fungi. *Science* 260: 340–342.

Yang, Z. 1996. Among-site rate variation and its impact on phylogenetic analyses. *Trends Ecol. Evol.* 11: 367–372.

Yang, Z., Nielsen, R., and Hasegawa, M. 1998. Models of amino acid substitution and applications to mitochondrial protein evolution. *Mol. Biol. Evol.* 15: 1600–1611.

Yang, Z., and Roberts, D. 1995. On the use of nucleic acid sequences to infer early branchings in the tree of life. *Mol. Biol. Evol.* 12: 451–458.

Zwickl, D. J., and Hillis, D. M. 2002. Increased taxon sampling greatly reduces phylogenetic error. *Syst. Biol.* 51: 588–598.