J Mol Evol (1998) 47:394–405

JOURNAL OF MOLECULAR EVOLUTION

Limitations of Metazoan 18S rRNA Sequence Data: Implications for Reconstructing a Phylogeny of the Animal Kingdom and Inferring the Reality of the Cambrian Explosion

Ehab Abouheif, Rafael Zardoya,* Axel Meyer†

Department of Ecology and Evolution, State University of New York, Stony Brook, New York 11794-5245, USA

Received: 9 December 1997 / Accepted: 23 March 1998

Abstract. We document the phylogenetic behavior of the 18S rRNA molecule in 67 taxa from 28 metazoan phyla and assess the effects of among-site rate variation on reconstructing phylogenies of the animal kingdom. This empirical assessment was undertaken to clarify further the limits of resolution of the 18S rRNA gene as a phylogenetic marker and to address the question of whether 18S rRNA phylogenies can be used as a source of evidence to infer the reality of a Cambrian explosion. A notable degree of among-site rate variation exists between different regions of the 18S rRNA molecule, as well as within all classes of secondary structure. There is a significant negative correlation between inferred number of nucleotide substitutions and phylogenetic information, as well as with the degree of substitutional saturation within the molecule. Base compositional differences both within and between taxa exist and, in certain lineages, may be associated with long branches and phylogenetic position. Importantly, excluding sites with different degrees of nucleotide substitution significantly influences the topology and degree of resolution of maximum-parsimony phylogenies as well as neighbor-joining phylogenies (corrected and uncorrected for among-site rate variation) reconstructed at the metazoan scale. Together, these data indicate that the 18S rRNA molecule is an unsuitable candidate for reconstructing the evolutionary history of all metazoan phyla, and that the polytomies, i.e., unresolved nodes within 18S rRNA phylogenies, cannot be used as a single or reliable source of evidence to support the hypothesis of a Cambrian explosion.

Key words: 18S rRNA — Molecular phylogeny — Metazoan phylogeny — Cambrian explosion — Among-site rate variation

Introduction

The evolutionary origin and relationships among all metazoan phyla remain a subject of great controversy (Field et al. 1988; Erwin 1991; Raff et al. 1994; Wray et al. 1996). Acquiring a robust phylogeny of the animal kingdom would provide a window through which we can examine the evolutionary patterns and processes involved in the origination of all animal phyla (Valentine et al. 1996). Here we examine the phylogenetic behavior of the 18S rRNA molecule at the metazoan scale and address the issue of whether reconstructions of metazoan phylogenies based on 18S rRNA data can be used as a source of evidence to infer the existence of the Cambrian explosion.

The Cambrian explosion refers to the sudden appearance of all skeletonized animal phyla in the fossil record within a relatively short period of about 25 million years at the base of the Cambrian approximately 545 million

**Present address:* Museo Nacional de Ciencias Naturales, Jose Gutierrez Abascal 2, 28006 Madrid, Spain
†*Present address:* Department of Biology, University of Konstanz, D-78457 Konstanz, Germany
Correspondence to: A. Meyer; *e-mail:* axel.meyer@uni-konstanz.de

Table 1. Taxon names and accession numbers

Taxon	Accession No.
Acanthocephala	
1. *Neoechinorhynchus pseudemydis*	U41400
2. *Centrorhynchus conspectus*	U41399
Annelida	
3. *Eisenia foetida*	X79872
4. *Lanice conchilega*	X79873
5. *Glycera americana*	U19519
6. *Nereis limbata*	U36270
Echiura	
7. *Ochetostoma erythrogrammon*	X79875
Pogonophora	
8. *Ridgeia piscesae*	X79877
Arthropoda	
Chelicerata	
9. *Ixodes affinis*	L76350
10. *Amblyomma tuberculatum*	L76345
11. *Haemaphysalis inermis*	L76338
Crustacea	
12. *Branchinecta packardi*	L26512
13. *Artemia salina*	X01723
14. *Astacus asatcus*	U33181
Insecta	
15. *Aeschna cyanea*	X89481
16. *Acheta domesticus*	X95741
17. *Prokelisia marginata*	U09207
Brachiopoda	
18. *Terebratalia transversa*	U12650
19. *Glottidia pyramidata*	U12647
20. *Lingula lingua*	X81631
Ectoprocta	
21. *Plumatella repens*	U12649
Entoprocta	
22. *Pedicellina cernua*	U36273
23. *Barentsia benedeni*	U36272
Cephalochordata	
24. *Branchiostoma floridae*	M97571
Chaetognatha	
25. *Sagitta elegans*	Z19551
26. *Paraspadella gotoi*	D14362
Cnidaria	
27. *Hydra littoralis*	U32392
28. *Bellonella rigida*	Z49195
29. *Anemonia sulcata*	X53498
30. *Tripedalia cystophora*	L10829
Ctenophora	
31. *Beroe cucumis*	D15068
32. *Mnemiopsis leidyi*	L10826
Echinodermata	
33. *Strongylocentrotus intermedius*	D14365
34. *Colobocentrotus atratus*	Z37121
35. *Echinodiscus bisperforatus*	Z37124
36. *Mespilia globulus*	Z37130
Hemichordata	
37. *Balanoglossus carnosus*	D14359
Gastrotricha	
38. *Lepidodermella squammata*	U29198
Mollusca	
Gastropoda	
39. *Limicolaria kambeul*	X66374
40. *Onchidella celtica*	X70211
Bivalvia	
41. *Crassostrea virginica*	X60315
42. *Mytilus galloprovincialis*	L33452
Polyplacophora	
43. *Liolophura japonica*	X70210
Nematoda	
44. *Strongyloides stercoralis*	M84229
45. *Caenorhabditis remanei*	U13930
Nematomorpha	
46. *Gordius aquaticus*	X80233
Nemertini	
47. *Prostoma eilhardi*	U29494
48. *Lineus* sp.	X79878
Phoronida	
49. *Phoronis vancouverensis*	U12648
50. *Phoronis architecta*	U36271
Placozoa	
51. *Trichoplax* sp.	Z22783
52. *Trichoplax adhaerens*	L10828
Platyhelminthes	
53. *Opisthorchis viverrini*	X55357
Porifera	
54. *Scypha ciliata*	L10827
Priapulida	
55. *Priapulus caudatus*	Z38009
Rotifera	
56. *Brachionus plicatilis*	U29235
57. *Philodina acuticornis*	U41281
Sipuncula	
58. *Phascolosoma granulatum*	X79874
Tardigrada	
59. *Macrobiotus* sp.	U32393
Urochordata	
60. *Herdmania momus*	X53538
61. *Thalia democratica*	D14366
62. *Styela plicata*	M97577
Vertebrata	
Agnatha	
63. *Petromyzon marinus*	M97575
Pisces	
64. *Fundulus heteroclitus*	M91180
65. *Latimeria chalumnae*	L11288
Amphibia	
66. *Xenopus laevis*	X04025
Mammalia	
67. *Homo sapiens*	X03205

years ago (mya) (Bowring et al. 1993; reviewed by Conway-Morris 1993; 1994). Interpreting this sudden appearance of the major animal phyla in the fossil record as an unparalleled explosion and rapid divergence of animal phyla requires a literal reading of the fossil record (Wray et al. 1996). Since the fossil record may in certain cases be a biased recording of the history of life (Signor 1990), investigators have searched for other, independent, lines of evidence to document the Cambrian radiation. In this context, the lack of resolution in metazoan phylogenies (i.e., polytomies) derived from 18S rRNA data has been interpreted as evidence for a Cambrian explosion (e.g., Erwin 1991; Phillipe et al. 1994). Although this observation is consistent with the interpretations of the fossil record, it is in conflict with recent molecular analyses

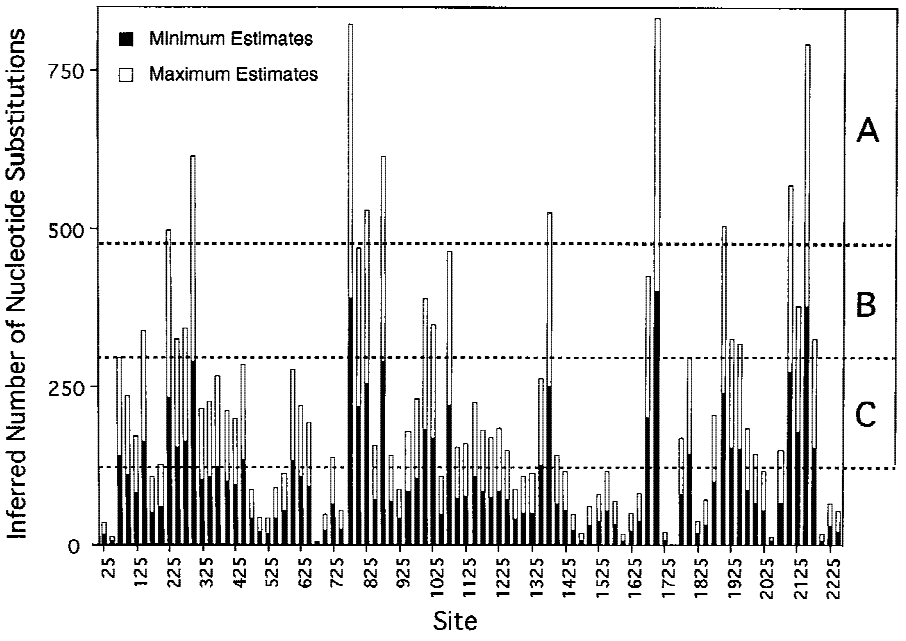

Fig. 1. The inferred number of nucleotide changes per site over the entire alignment of the 18S rRNA molecule calculated for 25-bp windows. Estimates of the number of nucleotide changes were obtained from a parsimony based algorithm, in which the uncertainty of the estimates is due to the uncertainties in the phylogeny used to obtain these estimates (Maddison and Maddison 1992). The *black bars* indicate the minimum number of estimated nucleotide changes, while the *open bars* indicate the maximum number. The *dashed lines* roughly demarcate the three main levels of variability in the molecule, i.e., exclusion sets A, B, and C, which were progressively removed from the alignment prior to phylogenetic analyses.

based on calibrated rates of sequence divergence (Runegar 1982; Doolittle et al. 1996; Wray et al. 1996; Nikoh et al. 1997), as well as phylogenetic analyses of extant and Cambrian arthropods (Fortey et al. 1997), both of which suggest deep Precambrian divergences. The recent discovery of microscopic Precambrian metazoans (Li et al. 1998; Xiao et al. 1998) further supports this inference.

Interpreting the polytomies among phyla in 18S rRNA phylogenies as a true reflection of biological events during the Cambrian fails to consider seriously the alternative, that they are an artifact of the phylogenetic behavior of this molecule. It has long been known that rate heterogeneity among sites within 18S rRNA sequences exist and that it creates a major methodological bias (Dixon and Hillis 1990; Olsen and Woese 1993; Van de Peer et al. 1993; Yang 1996; Van de Peer and De Wachter 1997). Yet the effects of rate heterogeneity among sites on phylogeny reconstruction has not been thoroughly investigated at this phylogenetic scale. The existence of among-site rate variation within 18S rRNA allows the same molecule to be used for both higher and lower relationships, but as a consequence, it dilutes the amount of phylogenetic information available at any one level (Olsen and Woese 1993). Highly conserved sites within the molecule are good markers for recovering deep divergences but contribute little phylogenetic information for recovering those which are recent, while highly variable sites can recover recent divergences but contribute noise in our attempts to recover those which are deep. Furthermore, there are other potential sources of bias: high levels of homoplasy, substitutional saturation, sequence alignment ambiguity, compositional disparities, and compensatory changes among nucleotide bases, as well as rate-heterogeneity between lineages, some of which may be correlated with among-site rate variation, and have been previously considered (Phillipe et al. 1994).

To determine whether 18S rRNA sequences are suitable candidates to reconstruct the evolutionary patterns involved in the origination of all metazoan phyla, we assemble a data set comprised of 67 taxa from 28 metazoan phyla (Table 1). First, we characterize the phylogenetic behavior of this molecule, and second, we test the effects of among-site rate variation on reconstructing phylogenies of the animal kingdom. Together, these analyses aim to further clarify the phylogenetic utility and limits of resolution of 18S rRNA sequence data (Smith 1989; Dixon and Hillis 1990; Marshall 1992; Olsen and Woese 1993; Maley and Marshall 1998).

Methods

Analyses of Among-Site Rate Variation

The 18S rRNA nucleotide sequences analyzed here represent 67 taxa and 28 phyla and were all obtained from GenBank (see Table 1 for taxon names and accession numbers). These sequences were aligned using CLUSTAL W (default settings) followed by refinement by eye (alignments are available for inspection on our World Wide Web page: http://life.bio.sunysb.edu/ee/ehab). The alignment was then mapped onto the 18S rRNA secondary structure model of *Xenopus laevis* (Gutell 1994) to document among-site rate variation within each of the structural classes [i.e., stems, loops, bulges, and other, which were assigned according to the classification of Vawter and Brown (1992)].

The total number of nucleotide substitutions, transitions, transversions, and the consistency index at each site was calculated using a parsimony based algorithm from the phylogenetic tree in Fig. 7 (see below). This was done using the CHART STATE CHANGES AND STASIS option in MacClade (Maddison and Maddison 1992) with 1000 random resolutions of the polytomies contained within the tree.

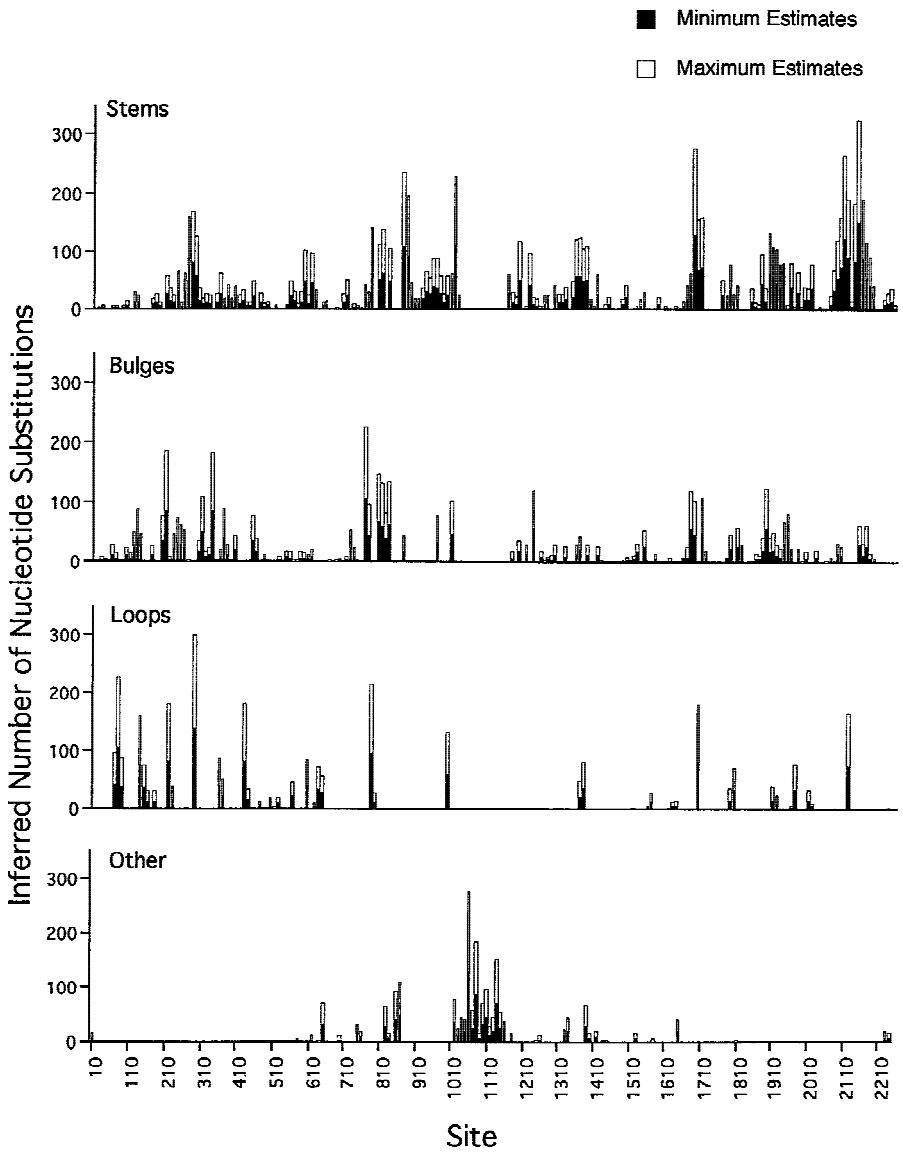

Fig. 2. The inferred number of nucleotide changes within each of the secondary structural classes, i.e., stems, bulges, loops, and other, of the 18S rRNA molecule calculated for 10-bp windows. Estimates of the number of nucleotide changes were obtained from a parsimony-based algorithm, in which the uncertainty of the estimates is due to the uncertainties in the phylogeny used to obtain these estimates (Maddison and Maddison 1992). The *black bars* indicate the minimum number of estimated nucleotide changes, while the *open bars* indicate the maximum number.

These analyses were performed over the full length of the molecule and within each of the secondary structural classes in 25- and 10-base pair (bp) windows, respectively. Base compositional frequencies for this data set were obtained from PAUP* (d55) (Swofford 1993).

To investigate the correlation among levels of homoplasy (estimated by the consistency index), substitutional saturation (estimated by the transition/transversion ratio), and inferred number of nucleotide changes, we calculated the Pearson product–moment correlation coefficient and least-squares regression slope between these variables (Sokal and Rohlf 1995).

Molecular Phylogenetic Analysis

To assess the effects of among-site rate variation on reconstructing metazoan phylogenies, we specified three alignment exclusion sets (A, B, C) (Fig. 1), which were specifically designed to remove successively regions with differing degrees of variability from the alignment. The boundary of exclusion set A marks the removal of all regions in the alignment that were ambiguous and difficult to align; they are also the most variable regions in the alignment. We then arbitrarily specified exclusion sets B and C (Fig. 1) based on their different levels of variability. These exclusion sets were successively removed from the alignment, i.e., exclusion set A was removed first, then exclusion sets A and B, and finally, exclusion sets A, B and C. A phylogeny was reconstructed after the successive removal of each of these exclusion sets.

Phylogenetic trees were reconstructed with the method of maximum parsimony (MP) (heuristic search, TBR branch swapping, MULPARS option in effect) and neighbor joining (NJ) (based on the Kimura two-parameter model, one-category of substitution rates) using the PHYLIP 3.5.5 package (Felsenstein 1989). NJ analyses were also performed with corrections for among-site rate variation by applying a gamma distribution derived from MP analyses of the same data sets (Yang and Kumar 1996). These had a shape parameter of $\alpha = 0.43$ for the alignment after the removal of exclusion set A, $\alpha = 0.42$ for the alignment after the removal of exclusion set A and B, and $\alpha = 0.35$ for the alignment after the removal of exclusion sets A, B, and C using PAUP* (Version d55) (Swofford 1997). Confidence in both the MP and the NJ trees was determined by analyzing 100 bootstrap replicates (Felsenstein 1985) in PHYLIP 3.5.5 and PAUP*. The phylum Porifera (sponges) was used as the outgroup for all phylogenetic analyses, as several independent molecular and morphological studies have placed the Porifera as the most basal of all metazoan phyla (Brusca and Brusca 1990; Raff et al. 1994; Nielsen 1995).

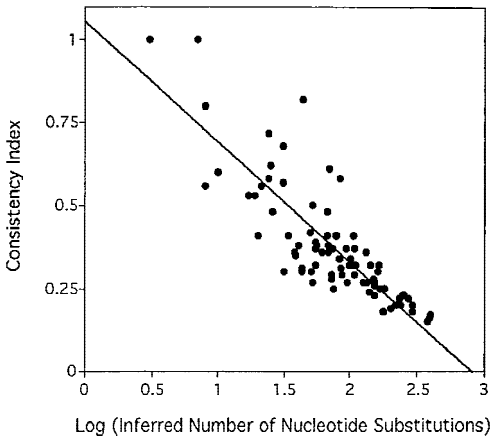

Fig. 3. The relationship between homoplasy and substitution rate as measured by a regression between the log of the inferred number of nucleotide changes and the consistency index across the entire 18S rRNA alignment calculated for 25-bp windows. The *solid black line* is the least-squares regression line, $r = -0.845$.

Results and Discussion

Analyses of Among-Site Rate Variation

Examining among-site rate variation along the entire length of the 18S rRNA alignment in 25-bp windows reveals that there are at least 10-fold differences in magnitude in the observed number of nucleotide substitutions between different regions of the molecule (Fig. 1). Surprisingly, the same 10-fold difference in magnitude is observed when among-site rate variation is mapped within each of the secondary structural classes in 10-bp windows (Fig. 2). Thus, there is notable among-site rate variation in all classes of secondary structure at the metazoan scale, which could potentially introduce inaccuracy and bias in the results of any phylogenetic reconstructions of the animal kingdom.

The wide range of estimates between the minimum and the maximum values of the number of nucleotide changes within each of the 25-bp windows in Fig. 1 is due to the large number of unresolved nodes contained within the phylogenetic tree in Fig. 7 (see below). These parsimony-based estimates, however, do not take multiple substitutions on the same branch into account. Therefore, this estimation procedure will tend strongly to underestimate the real substitution rate (Yang and Kumar 1996; Nielsen 1997), making our estimates conservative in illustrating notable among-site rate variation.

A significant negative correlation exists between the consistency index and the log of the inferred number of nucleotide substitutions for the entire alignment (derived from 25-bp windows), in which the centroid of points falls between a consistency index of 0.25 and 0.45 (Fig. 3). This indicates that sites which evolve more rapidly are more homoplasious and that the levels of homoplasty in this data set are quite high compared to other data sets

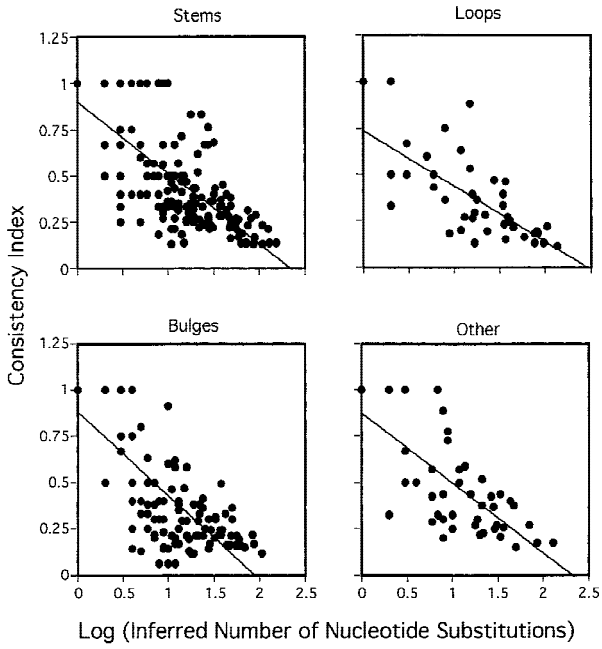

Fig. 4. The relationship between homoplasy and substitution rate as measured by the regression between the log of the inferred number of nucleotide changes and the consistency index within each of the secondary structural classes, i.e., stems, loops, bulges, and other, calculated for 10-bp windows. The *solid black line* is the least-squares regression line. The Pearson product–moment correlation coefficient within stems = −0.750, loops = −0.732, bulges = −0.790, and other = −0.733.

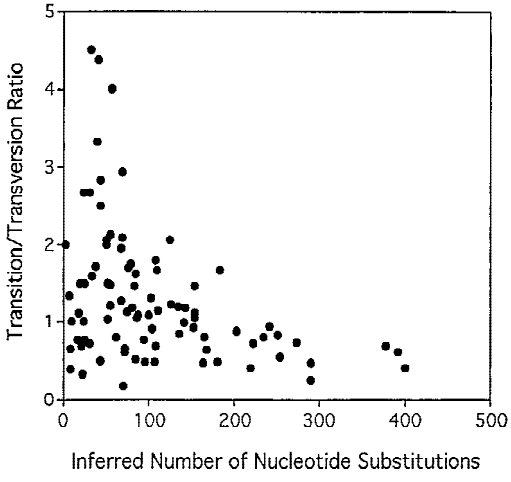

Fig. 5. The correlation between the transition:transversion ratio as a rough estimate of substitutional saturation and the inferred number of nucleotide changes along the entire 18S rRNA alignment calculated for 25-bp windows.

with similar numbers of taxa and characters (Sanderson and Donoghue 1996). Once again, similar slopes, correlations, and centroids are found within each of the structural classes of the 18S rRNA molecule (Fig. 4).

The negative correlation between inferred number of nucleotide changes and CI is interesting, as it holds important implications for the manner in which confidence in phylogenetic trees is assessed. This result demon-

Table 2. Mean base frequencies combined for all taxonomic groups as well as individual subtaxa within the data set

Taxon	Mean base frequency				χ^2 test		Range of branch lengths from most recent common ancestor to tip within each taxon[a]
	A	C	G	T	df	*P*	
All taxa	0.26834	0.21011	0.27205	0.2495	198	0.007	0.0013–0.74
Annelida	0.26775	0.21066	0.27392	0.24767	15	0.999	0.023–0.13
Arthropoda	0.26424	0.22198	0.2707	0.24307	24	1	0.0011–0.12
Brachiopoda	0.26812	0.20552	0.2726	0.25376	6	0.99	0.020–0.043
Entoprocta	0.26507	0.21174	0.27269	0.2505	6	0.99	0.013–0.086
Chordata	0.25938	0.22139	0.27789	0.24133	27	0.99	0.0024–0.11
Nemertini	0.2681	0.2095	0.27417	0.24823	3	0.98	0.066–0.082
Phoronida	0.26541	0.20882	0.27585	0.24992	3	0.95	0.029–0.030
Echinodermata	0.2646	0.21741	0.28072	0.23727	9	0.99	0.0020–0.018
Mollusca	0.26646	0.21126	0.27535	0.24694	12	1	0.010–0.084
Chaetognatha	0.24465	0.24941	0.29562	0.21633	3	0.97	0.019–0.49
Acanthocephala	0.2757	0.19912	0.27297	0.25221	3	0.204	0.059–0.74
Rotifera	0.28692	0.18971	0.25563	0.26774	3	0.31	0.048–0.74
Nematoda	0.28894	0.18247	25402	0.27456	3	0.02	0.11–0.53
Placozoa	0.28296	0.18526	0.25828	0.2735	3	1	0.0097–0.023
Cnidaria	0.28066	0.18632	0.26146	0.27156	9	0.99	0.032–0.11
Ctenophora	0.27873	0.19699	0.25889	0.26539	3	1	0.014–0.030

[a] Branch lengths represent the total difference between characters and were obtained from a NJ phylogram (Kimura two-parameter model, exclusion set A, γ corrected for among-site rate variation).

strates that the commonly used ''mean Cl'' as a measure of phylogenetic information and confidence may not capture the entire pattern of homoplasy contained within a given cladogram. This may explain the results of a recent comparative study by Sanderson and Donoghue (1996), in which they find that Cl is generally decoupled from confidence in phylogenetic trees. Thus, it is imperative that future studies account for among-site rate variation when computing an ''overall'' Cl and applying it as a measure of confidence or phylogenetic information contained within a cladogram.

A significantly negative, but nonlinear, relationship exists between the transition/transversion ratio and the inferred number of nucleotide substitutions over the entire alignment (again in 25-bp windows) (Fig. 5). Since these are counts of transitions and tranversions, the expected transition/tranversion ratio for a site that is at or near saturation is 0.5 given that there are twice as many possible transversions as there are transitions. Figure 5 shows that, at this phylogenetic level, those regions which are most conserved (<100 nucleotide changes in a 25-bp window) contain the lowest degree of substitutional saturation, whereas those regions which evolve rapidly only contain sites which are or are nearly saturated (Desalle et al. 1987).

Although substitutional saturation of sites is present throughout the data set, and may potentially hinder phylogenetic analyses, if the more rapidly evolving sites are removed, then substitutional saturation of sites should not be considered a major source of bias. This result is consistent with that of Philippe et al. (1994) in their finding that there are only weak saturational effects due to mutation in the 18S rRNA molecule across their limited data set of metazoan phyla.

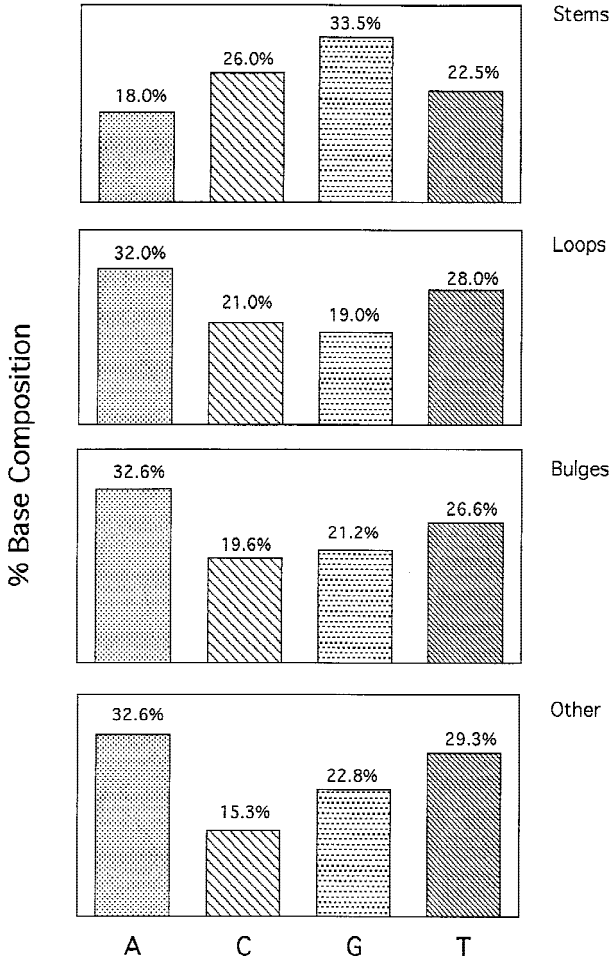

Fig. 6. Base composition within each of the secondary structural classes, i.e., stems, loops, bulges, and other.

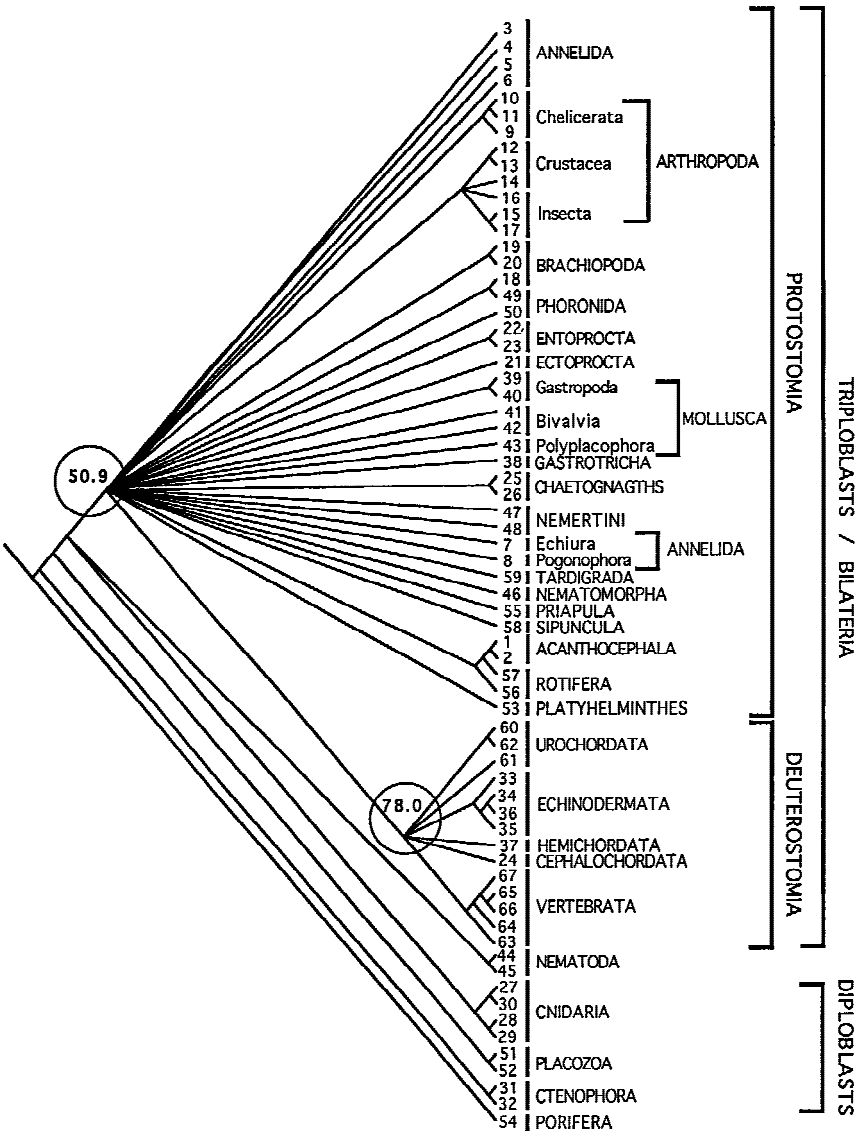

Fig. 7. Bootstrap MP tree after the removal of exclusion set A. The tree is based on 100 bootstrap replicates, and all nodes with bootstrap values less than 50% were collapsed into polytomies. The bootstrap support for nodes at the base of the triploblastic grade animals and the deuterostomes is indicated directly on the tree. The numbers at the tips of the branches correspond to those in Table 1 and serve as labels for species names. The phylum Porifera was used as the outgroup taxon. The same procedures apply to Figs. 8 and 9.

There are base compositional differences within and between taxa (Table 2). The test of homogeneity of base frequencies presented in Table 2 is only a heuristic given the low statistical power of the test within any given taxon and the confounding problem of historical nonindependence of data points between taxa. It is interesting that not all taxa show the same patterns of base compositional differences; those taxa above the horizontal line in Table 2 have a similar pattern, which differs from that seen in taxa below the horizontal line. These differences in base compositional patterns may be correlated in part with the possession of long branches by particular lineages, such as the Chaetognatha, Acanthocephala, Rotifera, and Nematoda (Table 2). This pattern also seems to be associated with the phylogenetic position of particular lineages, such as the Placozoa, Cnidaria, and Ctenophora. These results reveal the biases which can potentially be introduced by base compositional differences within and between taxa and may be controlled for by the omission of particular lineages (or species taken to represent those lineages) prior to any given phylogenetic analysis.

There are also base compositional differences within each of the structural classes (Fig. 6), which are a reflection of the structural folding constraints put upon the 18S rRNA molecule. The paired stem regions are GC-rich, reflecting the enhanced thermodynamic stability of GC pairs (Frier et al. 1986). The unpaired regions (loops, bulges, and other) are A-rich, reflecting the possible interactions of these regions with rRNA subunits and ribosomal proteins (Gutell et al. 1985). These results are consistent with previous rDNA analyses performed at different phylogenetic scales (Vawter and Brown 1993; Orti et al. 1996; Zardoya and Meyer 1996; Friedrich and Tautz 1997) and reveal that the functional constraints on various regions within the 18S rRNA molecule may render these regions nonindependent from one another (Wheeler and Honeycutt 1988; Orti et al. 1996). This

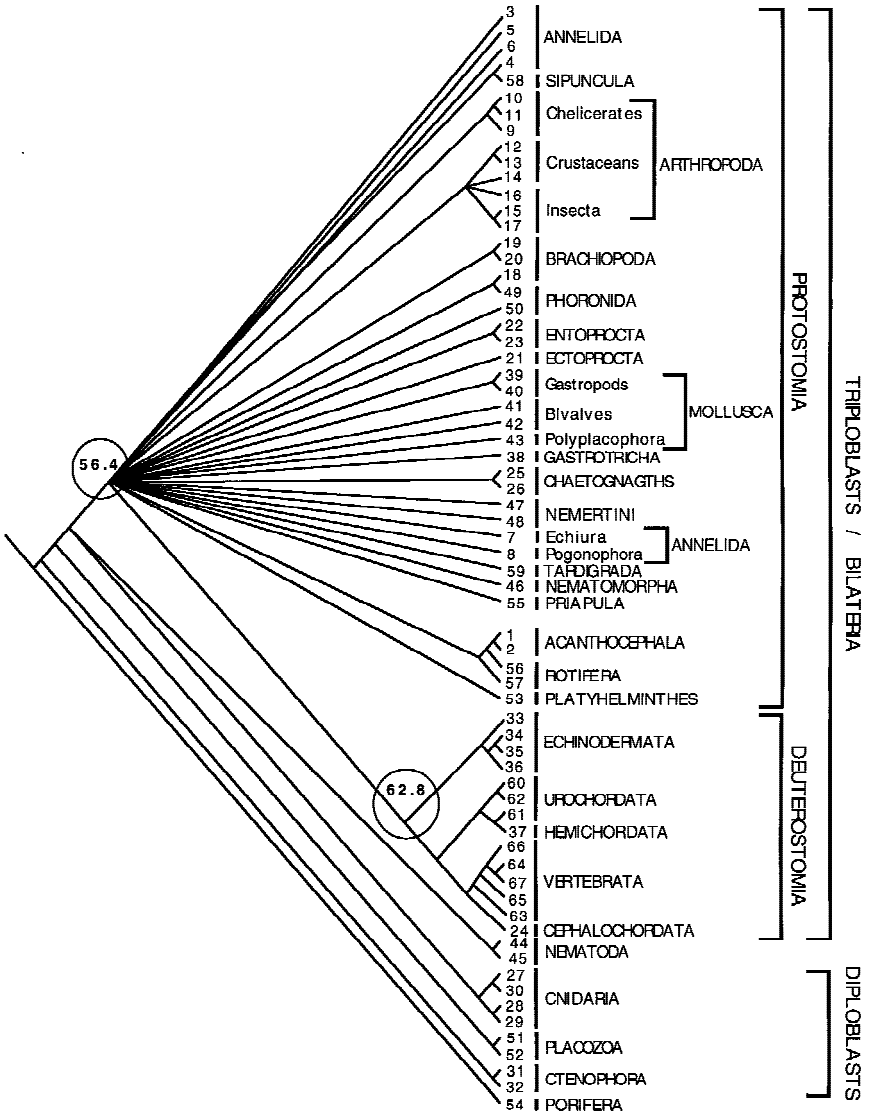

Fig. 8. Bootstrap MP tree after the removal of exclusion sets A and B.

introduces yet another difficulty that may, in some cases, require these regions to be analyzed separately from one another in any given phylogenetic analysis with the possibility of combining results afterward. For example, Smith (1989) investigated the phylogenetic relationships among echinoderms using 18S rRNA sequences and found that paired regions produce more reliable results than unpaired regions, while combining the data from paired and unpaired regions produces intermediate results. Partitioning phylogenies by secondary structure, however, are inappropriate in the context of reconstructing phylogenies at the metazoan scale, because the sequences become too short to be informative.

Molecular Phylogenetic Analysis

The bootstrap (50% majority rule consensus) MP tree in Fig. 7 is based on the multiple sequence alignment of 67 taxa from 28 phyla after the removal of exclusion set A (Fig. 1) and contains 714 informative sites. The major nodes within the tree are largely consistent with previous 18S rRNA analyses of metazoan relationships (Raff et al. 1994). The diploblastic animals branch off prior to triploblastic animals, and the triploblasts subsequently split into a deuterostome and a protostome clade. However, protostome monophyly is unsupported, and the monophyly of the deuterostome clade is supported but its topology is largely unresolved.

Although these polytomies and weakly supported nodes may be the result of a rapid radiation of animal phyla, the phylogenetic reconstruction based on the removal of exclusion sets A and B, containing 615 informative sites (Fig. 8), seems to present a qualitatively and quantitatively different picture. The relationships among phyla within the deuterostome clade have now become fully resolved. This is an important result, since a number of Cambrian fossils have been interpreted as representatives of several deuterostome phyla (i.e., echinoderms, cephalochordates, and hemichordates) (Fortey et al. 1997). The fact that these deuterostomes are present

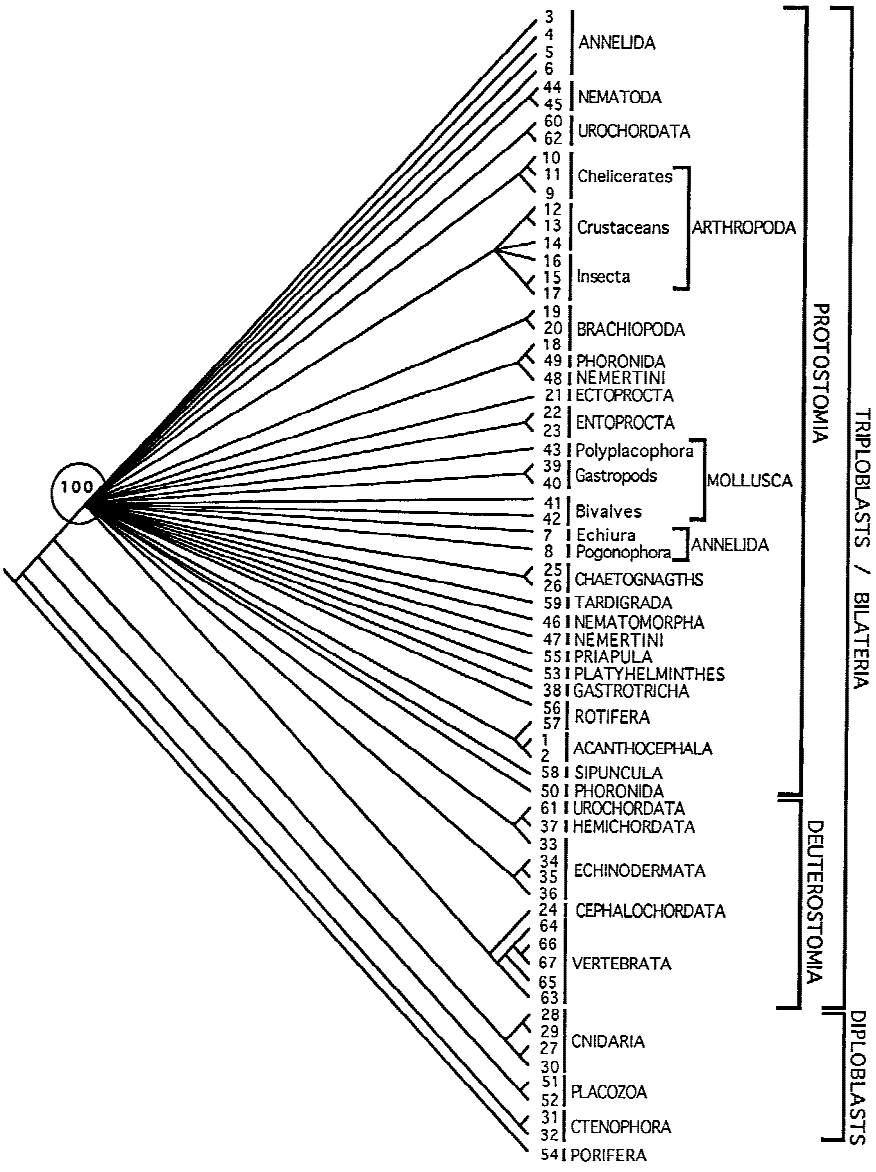

Fig. 9. Bootstrap MP tree after the removal of exclusion sets A, B, and C.

in the Cambrian and that a deuterostome clade can be fully recovered at this phylogenetic scale is consistent with a Precambrian divergence of these phyla. With respect to the protostomes, Wray and co-workers' (1996) scenario of Precambrian divergences suggests that the protostome node may be difficult to recover due to the hypothesized proximity of divergence times between the triploblast and protostome nodes. Conversely, the node at the base of the triploblasts, the deuterostomes, and the chordates may well be within the range of 18S rRNA's resolving power since their hypothesized divergence dates were found to be significantly different from one another. Furthermore, the full resolution of phylum-level relationships within the deuterostome clade reflects the artifactual nature of some of the polytomies contained within the tree (Figs. 7 and 8).

The resulting bootstrap MP tree based on the alignment after the removal of exclusion sets A, B, and C (350 phylogenetically informative sites) (Fig. 9) effectively demonstrates the existence of conflicting phylogenetic signals contained within this data set. Since many of the most variable sites have been removed from the alignment, many higher nodes become unstable, whereas the bootstrap support for some of the deeper nodes in the tree increases dramatically (Fig. 10). For example, bootstrap support for the triploblast node increases to 100%, while the echinoderms become paraphyletic, and urochordates fall among the protostomes. Although the latter two results are almost surely incorrect, the novel placement of the nematodes among the protostomes is consistent with the recent and independent findings of Aguinaldo et al. (1997). This novel placement of the nematode lineage may be a result of correcting for the two- to threefold faster sequence evolution within particular nematode lineages by the removal of the quickly evolving sites within exclusion sets A, B, and C.

Comparing the bootstrap support at particularly important nodes across the three phylogenetic trees in Figs.

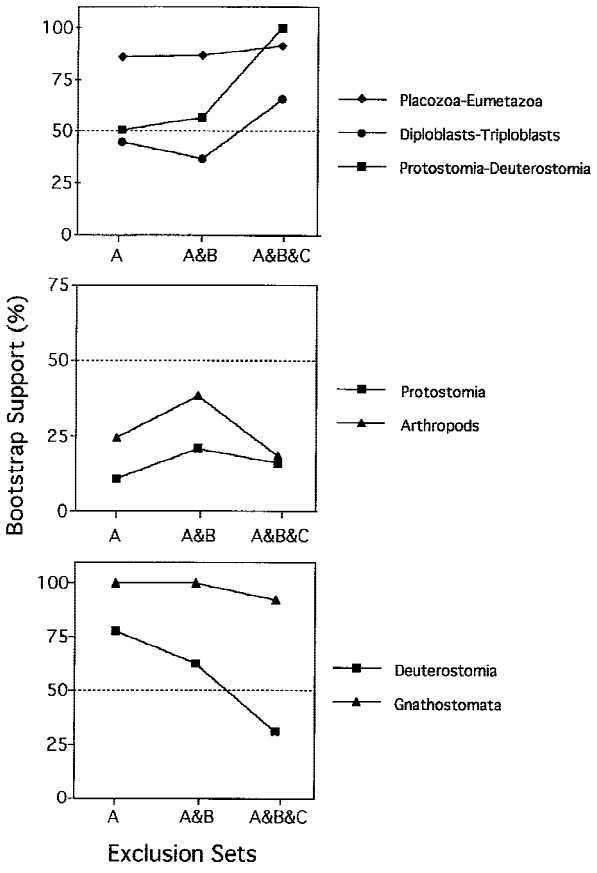

Fig. 10. Comparison of bootstrap values for particular nodes across the MP trees in Figs. 7 (based on the removal of exclusion set A), 8 (based on the removal of exclusion sets A and B), and 9 (based on the removal of exclusion sets A, B, and C).

7, 8, and 9 further reveals conflicting phylogenetic signals in the data set, which may be caused by the large effects of among-site rate variation (Fig. 10). Fig. 10A shows that, at the deeper nodes in the tree, the phylogenetic signal is concentrated in those regions in the alignment that remain after the removal of exclusion sets A, B, and C, i.e., those with low levels of variation. Conversely, the phylogenetic signal in the protostome node and a subtaxon, the Arthropoda (Fig. 10B), is concentrated in those regions in the alignment that remain after the removal of exclusion sets A and B. Finally, at the deuterostome node and a subtaxon, the Gnathostomata, (Fig. 10C) the phylogenetic signal is concentrated in those regions of the alignment that remain after removing exclusion set A.

The phylogenetic trees based on the NJ method uncorrected for among-site rate variation are largely consistent with those obtained with MP (data not shown). The taxonomic placement among the diploblastic grade animals are unchanged, and a triploblast clade is supported regardless of which of the exclusion sets are removed. A deuterostome clade is supported in exclusion sets A and B, with the degree of resolution and bootstrap support following approximately the same pattern as the MP trees; 8 of 14 nodes within the deuterostomes are resolved at the 50% bootstrap level after the removal of exclusion set A, but 10 of 14 are resolved after the removal of exclusion sets A and B. After the removal of exclusion set C, the deuterostome clade becomes unsupported. The only notable discrepancy in the uncorrected NJ trees compared to MP trees is the placement of the Platyhelminthes, Gastrotricha, Ancanthocephala, and Rotifera, all of which fall to a basal position relative to the rest of the triploblastic grade animals. However, the basal phylogenetic placement of these phyla may be due to long branch attraction (Table 2). Furthermore, the trees based on the NJ method with corrections for among-site rate variation place the Platyhelminthes, Gastrotricha, Ancanthocephala, and Rotifera among the protostomes. The remaining taxonomic placements within the rate-corrected NJ trees are mostly consistent with the MP trees but are poorly resolved. For example, regardless of which of the exclusion sets is removed, a deuterostome clade is never supported. It is also important to note that in both MP and NJ analyses, the phylogenetic placements of species which have diverged relatively recently are not well resolved by 18S rRNA at this phylogenetic scale. For example, the relationships among the Vertebrata in all the phylogenies presented here fail to recover the expected topology derived both by molecules and morphology (Wada and Satoh 1994; McClintock Turbeville et al. 1994). This variability in degree of resolution and taxonomic placements using the MP and NJ (with and without corrections for among-site rate variation) methods further reveals the pronounced effects of rate heterogeneity among sites and the remaining level of uncertainty of metazoan relationships.

Methodological Issues and Conclusions

Any attempts to reconstruct the phylogenetic history of the metazoa from a single molecular data set must take into account the problem of among-site rate variation. The data presented here clearly demonstrate that rate heterogeneity among sites has significant impact on reconstructing metazoan phylogeny, and creates conflicting phylogenetic signals within the molecule. This reinforces the need to apply phylogenetic methods which can account for this problem (Tateno et al. 1994; Kuhner and Felsenstein 1994; Yang 1995). Here we used a parsimony-based method to document and assess the effects of rate heterogeneity among sites. This approach is related to threshold parsimony (Felsenstein 1981) in which sites evolving extremely rapidly are given less, but not zero weight. These parsimony methods can potentially be used to account for the effects of among-site rate variation but will have to be evaluated both empirically and by computer simulation before they can be applied with confidence. This approach may also be used in combination with the NJ method when correcting for rate

heterogeneity via the γ distribution, since the γ correction may not adequately alleviate severe cases of among-site rate variation (Yang 1996), as is the case for 18S rRNA.

Future attempts to account for among-site rate variation at the metazoan scale will be informative as a means of both empirically evaluating phylogenetic methods and improving the accuracy of reconstructing metazoan relationships. These analyses make it clear that the 18S rRNA molecule alone is an unsuitable candidate for reconstructing a phylogeny of the metazoa and that the polytomies commonly observed within 18S rRNA phylogenies are not reliable evidence for inferring the existence of the Cambrian explosion. Independent data from several molecules are required to ascertain the biological reality of polytomies in phylogenies.

Acknowledgments. We thank Ernst Mayr for providing the impetus for this study, and A. Bely, M. Friedrich, J. Felsenstein, G. Bharathan, E. Knott, and D. McHugh for insightful discussions or comments on early drafts of the manuscript. David Swofford kindly granted permission to publish results based on the test version d55 of his PAUP* program. This work was supported by a graduate fellowship from the Fonds pour la Formation de Chercheurs et l'Aide à la Recherche (FCAR) to E.A. and a postdoctoral fellowship to R.Z. from the Ministerio de Education y Ciencia of Spain. This work received partial financial support from grants to A.M. from the National Science Foundation (DEB-9615178) and from grant from the Max-Planck Society, Germany.

References

Aguinaldo AMA, Turbeville JM, Linford LS, Rivera MC, Garey JR, Raff RA, Lake JA (1997) Evidence for a clade of nematodes, arthropods and other moulting animals. Nature 387:489–493

Bowring SA, Grotzinger JP, Isachsen CE, Knoll AH, Pelechaty SM, Kolosov P (1993) Calibrating rates of early Cambrian evolution. Science 261:1293–1298

Brusca RC, Brusca GJ (1990) Invertebrates. Sinauer, Sunderland, MA

Conway Morris S (1993) The fossil record and the early evolution of the Metazoa. Nature 361:219–225

Conway Morris S (1994) Why molecular biology needs paleontology. Development 120 (Suppl):1–13

Desalle R, Freedman T, Prager EM, Wilson AC (1987) Tempo and mode of sequence evolution in mitochondrial DNA of Hawaiian *Drosophila.* J Mol Evol 26:157–164

Doolittle RF, Feng DF, Tsang S, Cho G, Little E (1996) Determining divergence times of the major kingdoms of living organisms with a protein clock. Science 271:470–477

Erwin DH (1991) Metazoan phylogeny and the Cambrian radiation. Trends Ecol Evol 6:131–134

Felsenstein J (1981) A likelihood approach to character weighting and what it tells us about parsimony and compatibility. Biol J Linn Soc 16:183–196

Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39:783–791

Felsenstein J (1989) Phylogenetic inference programs (PHYLIP). University of Washington, Seattle, and University Herbarium, University of California, Berkeley

Field KG, Olsen GJ, Lane DJ, Giovannoni SJ, Ghiselin MT, Raff EC, Pace NR, Raff RA (1988) Molecular phylogeny of the animal kingdom. Science 239:748–753

Fortey RA, Briggs DEG, Wills MA (1997) The Cambrian evolutionary ''explosion'' recalibrated. BioEssays 19:429–434

Freier SM, Kierzek R, Jaeger JA, Sugimoto N, Caruthers MH, Neilson T, Turner DH (1986) Improved free-energy parameters for predictions of RNA duplex stability. Proc Natl Acad Sci USA 83:9373–9377

Friedrich M, Tautz D (1997) An episodic change of rDNA nucleotide substitution rate has occurred during the emergence of the insect order Diptera. Mol Biol Evol 14:644–653

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

Gutell RR, Weiser B, Woese CR, Noller HF (1985) Comparative anatomy of 16S-like ribosomal RNA. Prog. Nucleic Acid Res Mol Biol 32:155–216

Hillis DM, Dixon MT (1991) Ribosomal DNA: molecular evolution and phylogenetic inference. Q Rev Biol 66:411–453

Kuhner MK, Felsenstein J (1994) A simulation comparison of phylogeny algorithms under equal and unequal evolutionary rates. Mol Biol Evol 11:459–468

Li CW, Chen J-Y, Hua T-E (1998) Precambrian sponges with cellular structures. Science 279:879–882

Maddison WP, Maddison DR (1992) MacClade. Sinauer, Sunderland, MA

Maley LE, Marshall CR (1998) The coming age of molecular systematics. Science 279:505–506

Marshall CR (1992) Substitution bias, weighted parsimony, and amniote phylogeny as inferred from 18S rRNA sequences. Mol Biol Evol 9:370–373

McClintock Turbeville J, Schulz JR, Raff RA (1994) Deuterostome phylogeny and the sister group of the chordates: evidence from molecules and morphology. Mol Biol Evol 11:648–655

Nielsen C (1995) Animal evolution. Oxford University Press, Oxford, UK

Nielsen R (1997) Site-by-site estimation of the rate of substitution and the correlation of rates in mitochondrial DNA. Syst Biol 46:346–353

Nikoh N, Iwabe N, Kuma K, Ohno M, Sugiyama T, Watanabe Y, Yasui K, Shi-cui Z, Hori K, Shimura Y, Miyata T (1997) An estimate of divergence time of Parazoa and Eumetazoa and that of Cephalochordata and Vertebrata by Aldolase and triose phosphate isomerase clocks. J Mol Evol 45:97–106

Olsen GJ, Woese CR (1993) Ribosomal RNA: a key to phylogeny. FASEB J 7:113–123

Ortí G, Petry P, Porto JIR, Jégu M, Meyer A (1996) Patterns of nucleotide change in mitochondrial ribosomal RNA genes and phylogeny of piranhas. J Mol Evol 42:169–182

Philippe H, Chenuil A, Adoutte A (1994) Can the Cambrian explosion be inferred through molecular phylogeny? Development 120 (Suppl):15–25

Raff RA, Marshall CR, Turbeville JM (1994) Using DNA sequences to unravel the Cambrian radiation of the animal phyla. Annu Rev Ecol Syst 25:351–375

Runegar B (1982) A molecular clock date for the origin of the animal phyla. Lethaia 15:199–205

Sanderson MH, Donoghue M (1996) The relationship between homoplasy and confidence in a phylogenetic tree. In: Sanderson MJ, Hufford L (eds) Homoplasy: the recurrence of similarity in evolution. Academic Press, San Diego, pp 67–89

Signor PW (1990) The geologic history of diversity. Annu Rev Ecol Syst 21:509–539

Smith AB (1989) RNA sequence data in phylogenetic reconstruction: testing the limits of its resolution. Cladistics 5:321–344

Sokal RR, Rohlf FJ (1995) Biometry, 3rd ed. Freedman, New York

Swofford DL (1993) PAUP: phylogenetic analysis using parsimony. Illinois Natural History Survey, Champaign

Swofford DL (1997) PAUP*: phylogenetic analysis using parsimony (* and other methods), Version d55. Sinauer Associates, Sunderland, MA

Tateno Y, Takezaki N, Nei M (1994) Relative efficiencies of the maximum-likelihood, neighbor-joining, and maximum parsimony methods when substitution rate varies with site. Mol Biol Evol 11:261–277

Valentine JW, Erwin DH, Jablonski D (1996) Developmental evolution of metazoan body plans: the fossil evidence. Dev Biol 173:373–381

Van de Peer Y, Neefs JM, De Rijk P, De Wachter R (1993) Reconstructing evolution from eukaryotic small-ribosomal-subunit RNA sequences: calibration of the molecular clock. J Mol Evol 37:221–232

Van de Peer Y, 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

Vawter L, Brown WM (1993) Rates and patterns of base change in the small subunit ribosomal RNA gene. Genetics 134:597–608

Wada H, Satoh N (1994) Details of the evolutionary history from invertebrates to vertebrates, as deduced from sequences of 18S rDNA. Proc Natl Acad Sci USA 91:1801–1804

Wheeler WC, Honeycutt RL (1988) Paired sequence difference in ribosomal RNAs: evolutionary and phylogenetic implications. Mol Biol Evol 5:90–96

Wray GA, Levinton JS, Shapiro LH (1996) Molecular evidence for deep Precambrian divergences among metazoan phyla. Science 274:568–573

Xiao S, Zhang Y, Knoll AH (1998) Three dimensional preservation of algae and animal embryos in a Neoproterozoic phosphorite. Nature 391:553–558

Yang Z (1995) Evaluation of several methods for estimating phylogenetic trees when substitution rates differ over nucleotide sites. J Mol Evol 40:689–697

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

Yang Z, Kumar S (1996) Approximate methods for estimating the pattern of nucleotide substitution and the variation of substitution rates among sites. Mol Biol Evol 13:650–659

Zardoya R, Meyer A (1996) Evolutionary relationships of the coelacanth, lungfishes, and tetrapods based on the 28S ribosomal RNA gene. Proc Natl Acad Sci USA 93:5449–5454