Syst. Biol. 49(1):130–142, 2000

Early Evolution of the Bilateria

BERNHARD HAUSDORF

Zoologisches Institut und Zoologisches Museum der Universität Hamburg, Martin-Luther-King-Platz 3, D-20146 Hamburg, Germany; E-mail: hausdorf@zoologie.uni-hamburg.de

Abstract.—The phylogeny of the Bilateria and especially the early steps in the evolution of the bilaterian bauplan are still a controversial topic. In this context the relationships of the platyhelminths and the nematodes play a crucial role. Previous molecular studies of the relationships of these groups, which were based on 18S ribosomal DNA sequences, yielded conflicting results. In the present study a new framework is developed for the phylogenetic analysis of bilaterian relationships, using concatenated amino acid sequences of several nuclear genes. In this analysis, the rhabditophoran platyhelminths are probably the sister group of all other analyzed Bilateria, the Eubilateria, which are characterized by a one-way intestine with an anus. The Eubilateria are split into the nematode lineage and the coelomates. The phylogenetic results of the present study indicate that genetic features found in the model organisms *Caenorhabditis* and *Drosophila* might be found in all Eubilateria. Estimations of the divergence times show that the major bilaterian phyla did not originate in an explosive radiation during the Cambrian but rather that the Bilateria have a several hundred million years long Precambrian history. [Bilateria; Coelomata; molecular clock; molecular phylogeny; Nematoda; Platyhelminthes.]

The phylogeny of the Bilateria and especially the early steps in the evolution of the bilaterian bauplan are still a controversial topic. In this context the relationships of the platyhelminths and the nematodes play a crucial role because of their supposed basal position. But even if only the phylogenetic relationships of these groups and two other major bilaterian phyla, the arthropods and the chordates, are considered, the tree topologies of the various phylogenetic reconstructions do not correspond, neither between those reconstructions based on morphological characters (Figs. 1a–e) nor between those based on 18S ribosomal DNA (rDNA) (Figs. 1f–j).

Because of problems with the interpretation and establishment of homology for morphological characters across metazoan phyla, the phylogenetic analysis of conserved genes offers a promising approach. So far, only 18S rDNA has been used widely for phylogenetic analyses of the metazoan phyla (e.g., Riutort et al., 1993; Philippe et al., 1994; Raff et al., 1994; Winnepenninckx et al., 1995; Aguinaldo et al., 1997). However, the inconsistent results of analyses based on 18S rDNA (Figs. 1f-j) are disappointing. The reliability of 18S rDNA as a phylogenetic marker for deep divergences has repeatedly been questioned because of differences in the base composition among taxa (Hasegawa and Hashimoto, 1993; Abouheif et al., 1998), drastic differences in substitution rates among taxa that cause long branch attraction (Carmean and Crespi, 1995), and conflict between the trees based on 18S rDNA and trees estimated by using other data (Huelsenbeck and Bull, 1996). Moreover, 18S rDNA does not have enough informative positions for a robust reconstruction of the metazoan phylogeny (Philippe et al., 1994), the alignment of the rDNAs is often more ambiguous than that of protein-coding genes (Winnepenninckx and Backeljau, 1996), and the strong influence of the rRNA secondary structure on the evolution of its sequence is difficult to model (Dixon and Hillis, 1993).

Conservative protein-coding genes might be more suitable for the reconstruction of the metazoan phylogeny (Hasegawa and Hashimoto, 1993; Maley and Marshall, 1998), and a growing number of studies are based on individual protein-coding genes (e.g., Sidow and Thomas, 1994; McHugh, 1997; Nikoh et al., 1997; Borchiellini et al., 1998). However, these studies show that individual protein-coding genes do not include enough information for a robust reconstruction of metazoan phylogeny.

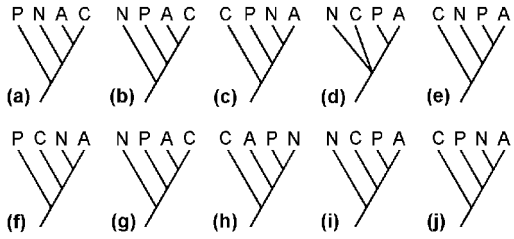

FIGURE 1. Comparison of the topologies of platyhelminths (P), nematodes (N), arthropods (A), and chordates (C) in published trees based on morphological characters (a–e) or on 18S rDNA sequences (f–j). (a) Hennig, 1979. (b) Schram, 1991. (c) Eernisse et al., 1992. (d) Backeljau et al., 1993. (e) Nielsen, 1995. (f) Riutort et al., 1993. (g) Philippe et al., 1994 (in this analysis the flies do not cluster with the other arthropods but with the nematodes). (h) Raff et al., 1994. (i) Winnepenninckx et al., 1995. (j) Aguinaldo et al., 1997.

In the present study a new framework for the phylogenetic analysis of bilaterian relationships is developed, using concatenated amino acid sequences of several nuclear genes. Furthermore, the divergence times of the major bilaterian groups are estimated by means of maximum likelihood trees constructed with the clocklike evolving sequences of the data set.

MATERIALS AND METHODS

Because there are fewer sequences of platyhelminths than of nematodes, arthropods, or vertebrates, the GenBank was screened for proteins from platyhelminths first. The fluke *Schistosoma* is the platyhelminth with most entries. The sequences available from *Schistosoma* were screened for genes that are also known in *Arabidopsis*, *Caenorhabditis*, *Drosophila*, *Mus*, and *Homo*. *Arabidopsis* is used as outgroup. *Caenorhabditis*, *Drosophila*, and *Mus* and *Homo* are the representatives of the nematodes, the arthropods, and the chordates, respectively, having the most entries in the data bases. About 30 genes were found to be known in all of the mentioned taxa and are so conserved that an unambiguous alignment is possible. Many of these genes have several paralogs, at least in some taxa. All genes with paralogs were analyzed phylogenetically to determine whether the paralogs from each phylum form monophyletic groups. When this is the case, the gene duplications do not affect the topology of the tree, and one paralog was chosen randomly for further analyses; otherwise, the genes were discarded. Further analyses with *Saccharomyces* as an additional outgroup were performed to test the robustness of the tree. The accession numbers of the analyzed sequences and the number of amino acids used for the phylogenetic analyses are listed in Table 1.

The protein sequences were aligned by using the divide-and-conquer multiple sequence alignment program version 1.0 (Tönges et al., 1996) with the BLOSUM 62 substitution matrix (Henikoff and Henikoff, 1992). All positions with gaps and all positions adjacent to gaps to the first site with a conserved amino acid in all analyzed taxa were excluded in the phylogenetic analyses. The sequence alignments may be obtained from the author, or from the *Systematic Biology* web site (www.utexas.edu/ftp/depts/systbiol/).

The program package PHYLIP version 3.57c (Felsenstein, 1995) was used for maximum parsimony (Fitch, 1971) and neighbor-joining (Saitou and Nei, 1987) analyses. The distances used for the neighbor-joining analyses were calculated with the PAM 250 matrix (Dayhoff et al., 1978). The program package PUZZLE version 4.0 (Strimmer and von Haeseler, 1996) was used for maximum likelihood (Felsenstein, 1981; Kishino et al., 1990) analyses, either with the PAM 250 matrix or with the BLOSUM 62 matrix. Unless otherwise stated, the calculations were performed by assuming uniform rates over all sites. If rate heterogeneity was allowed for, the rates were assumed to be gamma-distributed with eight rate categories; the gamma distribution parameter alpha was estimated from the data set. All other program options, unless noted, were default. Confidence values for internal branches were established by bootstrapping (Felsenstein, 1985; with 1,000 replications) for maximum parsimony and neighbor-joining analyses. Quartet puzzling provided the reliability values for maximum likelihood analyses (Strimmer and von Haeseler, 1996). Furthermore, alternative topologies were evaluated by the Kishino-Hasegawa test (Kishino and Hase-

TABLE 1. GenBank/EMBL accession numbers of the sequences and numbers of amino acids used for the phylogenetic analyses.

	Amino acids used in analyses								
	Without *Saccharomyces*	With *Saccharomyces*	*Arabidopsis*	*Saccharomyces*	*Schistosoma*	*Caenorhabditis*	*Drosophila*	*Mus*	*Homo*
Aldolase	297	297	X53058	—	L38658	D83738	M98351	Y00516	M11560
Calreticulin	253	253	U27698	—	L24159	X59589	X64461	X14926	M84739
Elongation factor 1α	437	428	X16432	U51033	Y08487	U40935	X06869	L26479	J04617
Enolase	384	361	X58107	J01322	U33177	Z68318	X17034	X52379	M14328
Glyceraldehyde-3-phosphate dehydrogenase	267	264	X98130	J01324	L09549	X52674	M11254	M32599	X01677
3-Hydroxy-3-methylglutaryl coenzyme A reductase	168	160	L19261	M22002	M27294	U28991	M21329	M62766	M11058
Phosphoglycerate kinase	342	332	U37701	J01342	L36833	U88169	Z14029	M15668	V00572
Ribosomal protein L13E	165	123	X75162	Z47071	U57003	U88308	X77926	U28917	X64707
t-Complex polypeptide-1α	511	433	D11351	X85021	U55769	U07941	M21159	D10606	X52882
Triose phosphate isomerase	223	201	U02949	J01366	M83294	U23081	X57576	X53333	X69723

gawa, 1989) with use of the program package PUZZLE (Strimmer and von Haeseler, 1996) and the PAM 250 matrix.

The hypothesis that the amino acid substitution rates are equal among lineages was tested with a likelihood ratio test (Felsenstein, 1981), again using the program PUZZLE for each gene and for different substitution and rate heterogeneity models. Only those genes for which the clock hypothesis cannot be rejected at a significance level of 5% were used for the calculations of divergence times. The calculations of divergence times were based on the maximum likelihood branch lengths of the trees estimated under the clock hypothesis. The divergence between rodents and primates was the only available reference for the datings. Several recent molecular clock estimations of the date of this divergence, based on different data and assumptions, vary from 100 million years before present (MYBP) (Li et al., 1990), 104 MYBP (Hedges et al., 1996), and 115 MYBP (Janke et al., 1997) to 125 MYBP (Janke et al., 1997). To get conservative estimates of divergence times, the rodent-primate divergence is set at 100 MYBP. The 85-million-year-old fossils of hoofed mammals (Archibald, 1996), which originated only after the divergence of rodents and primates, show that 100 MYBP is a reasonable estimate of the divergence time of rodents and primates.

Of course, the available data do not allow exact datings because of the small number of clocklike evolving genes known from all examined taxa and because only a single calibration point was available. Moreover, the true divergence times might be 25% older than those calculated from the conservative estimate of the rodent-primate divergence at 100 MYBP.

Relationships of Major Bilaterian Phyla

Ten nuclear protein-coding genes were found in the GenBank that fulfill the following conditions: (1) They must be known in at least *Arabidopsis*, *Schistosoma*, *Caenorhabditis*, *Drosophila*, *Mus*, and *Homo;* (2) they must be so conserved that an unambiguous alignment is possible; (3) no paralogs are allowed that do not form clades with other paralogs of the same phylum. The 10 genes are fructose-bisphosphate aldolase (class I), calreticulin, elongation factor 1α, enolase, glyceraldehyde-3-phosphate dehydrogenase, 3-hydroxy-3-methylglutaryl coenzyme A reductase, phosphoglycerate kinase, ribosomal protein L13E, *t*-complex polypeptide-1α, and triose phosphate isomerase. The aldolase and the enolase genes have been duplicated in the evolution of the chordates. Mice as well as humans exhibit several clearly identifiable paralogs of these genes. Therefore, the same ortholog from mice and humans has been included in the analysis. No class I fructose-bisphosphate aldolase and no calreticulin orthologs are known from *Saccharomyces.* Therefore, these two genes were coded as unknown for yeast in the analyses with *Saccharomyces* as an additional outgroup.

The maximum likelihood and neighbor-joining analyses of the concatenated amino acid sequences of the 10 proteins resulted in the same tree topology (Figs. 2, 3). This tree topology is also found in the maximum parsimony analysis with *Arabidopsis* as the only outgroup. If *Saccharomyces* is included as an additional outgroup, *Schistosoma* and *Caenorhabditis* are joined in the most-parsimonious tree. However, the majority-rule consensus tree of 1,000 bootstrap replicates with *Saccharomyces* as an additional outgroup analyzed with the maximum parsimony method corresponds to the tree found by other methods and not to the most-parsimonious tree.

The quartet puzzling reliability values obtained in the maximum likelihood analyses imply a strong support for all internal branches of this tree. In contrast, the bootstrap values for the neighbor-joining and parsimony analyses indicate that especially the monophyly of the group that includes the nematodes, arthropods, and chordates needs further corroboration. According to the Kishino-Hasegawa test (Table 2), many of the alternative topologies cannot be refuted at the 5% significance level. However, the test shows that the log-likelihood of the tree proposed by Aguinaldo et al. (1997) (Fig. 1j) is significantly ($p < 0.05$) smaller than that of the maximum likelihood tree, if *Saccharomyces* is included as an additional outgroup.

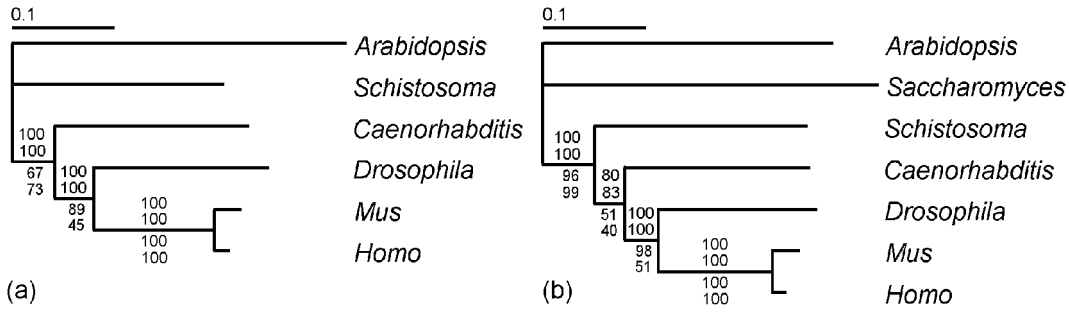

FIGURE 2. Bilaterian phylogeny based on amino acid sequences of 10 nuclear genes. The branch lengths shown are maximum likelihood estimations calculated with the PAM 250 matrix and assuming uniform rates over all sites. The numbers on the internal branches represent (from top to bottom) the quartet puzzling reliability values for the maximum likelihood analyses with the PAM 250 matrix and the BLOSUM 62 matrix and the bootstrap values for neighbor-joining and maximum parsimony analyses. (a) With *Arabidopsis* as only outgroup. (b) With both *Arabidopsis* and *Saccharomyces* as outgroups.

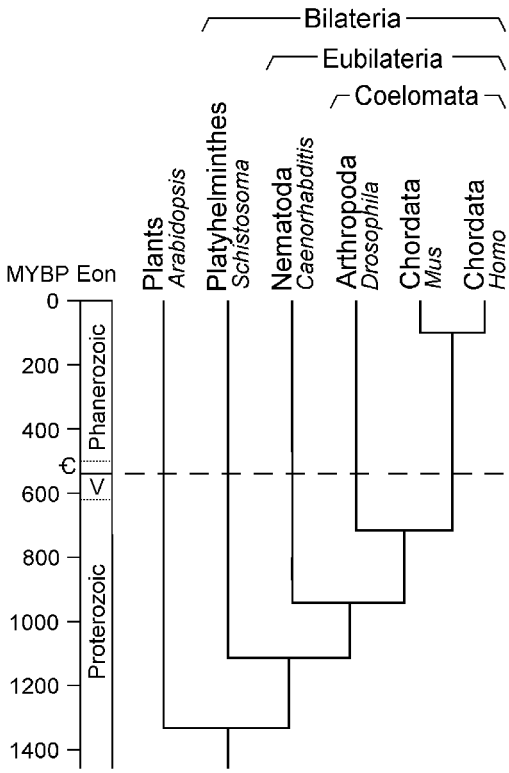

FIGURE 3. Temporal frame of the bilaterian phylogeny. The estimations of the divergence times are based on the maximum likelihood branch lengths calculated with the PAM 250 matrix and assuming a uniform rate over all sites (see Table 4). The name Coelomata is used only as a taxon name and should not necessarily imply homology of the coelom cavities. = Cambrian; V = Vendian; MYBP = million years before present.

Considering the sequence length, the low bootstrap values may appear disillusioning. However, one must consider that the divergences between the few taxa for which the analyzed sequences are available are very deep. Therefore, the terminal branches are very long and, thus, many multiple and parallel substitutions may have occurred along these branches. The statistical support for the internal branches may increase if additional taxa will be sampled that break the long terminal branches (Graybeal, 1998). The next steps in a suitable sampling strategy might be to include a rhabditophoran "turbellarian", a nematomorph, a mollusc or an annelid, and an echinoderm in the analyses.

The poor taxon sampling is the major disadvantage of the amino acid sequence data set presented in comparison with the available 18S rDNA data set. The major advantage of the concatenated amino acid sequences is that they might include enough informative positions for a robust reconstruction of the metazoan phylogeny, if sequences from additional taxa become available. If the current data set does not include enough informative positions, it can easily be expanded by inclusion of the amino acid sequences of hundreds of other nuclear protein-coding genes. In contrast, the 18S rDNA does not provide enough information for a robust reconstruction of the metazoan phylogeny despite a dense taxon sampling (Philippe et al., 1994; Abouheif et al., 1998), and that data set can be expanded without problems by only the few genes of

TABLE 2. Maximum likelihood analyses (with the PAM 250 matrix) of the 15 possible topologies of the platyhelminths (P), nematodes (N), arthropods (A), and chordates (C). Δlog L, log-likelihood differences between the maximum likelihood tree (for which the log-likelihood is given in parentheses) and the 14 other topologies as well as the standard deviation (s.d.) of these differences. Differences that are statistically significant at the 5% level according to the Kishino–Hasegawa test are marked with an asterisk.

	Analyses without *Saccharomyces*		Analyses with *Saccharomyces*	
	Δlog L	s.d.	Δlog L	s.d.
(P,(N,(A,C)))	(−23428.95)		(−24538.82)	
(P,(A,(N,C)))	−13.58	21.89	−23.10	21.14
(P,(C,(A,N)))	−19.44	21.25	−22.01	21.27
(A,(N,(P,C)))	−72.71*	28.96	−67.79*	28.25
(A,(P,(N,C)))	−45.12	30.12	−56.00*	28.09
(A,(C,(P,N)))	−38.35	27.39	−44.64	27.12
(N,(A,(P,C)))	−61.71*	26.72	−49.41	26.91
(N,(P,(A,C)))	−12.57	16.96	−4.15	15.88
(N,(C,(P,A)))	−14.12	30.72	−12.03	30.12
((P,N),(A,C)))	−19.63	16.36	−0.58	16.44
((P,A),(N,C))	−30.76	31.66	−27.28	30.75
((N,A),(P,C))	−72.40*	28.54	−54.63	28.85
(((P,N),A),C)	−43.56	27.42	−52.95*	26.60
(((P,A),N),C)	−36.78	32.40	−44.82	30.60
(((N,A),P),C)	−56.01	30.03	−61.94*	28.29

the rDNA cluster, because only these genes evolve according to the same model as the 18S rDNA.

If the individual protein-coding genes of the present data set are analyzed separately, the phylogenetic results differ from gene to gene, similar to the analyses of Wang et al. (1999). This confirms that the information conserved in single protein-coding genes is insufficient for the phylogenetic analysis of metazoan relationships and that the only promising approach to a robust resolution of metazoan relationships is to analyze a set of several sequences from many taxa.

According to the maximum likelihood tree based on the present set of amino acid sequences, the rhabditophoran platyhelminth lineage is the sister group of all other analyzed Bilateria, the Eubilateria (Figs. 2, 3). Recently, several analyses based on morphological characters and 18S rDNA sequences have challenged the monophyly of the platyhelminths (Haszprunar, 1996; Carranza et al., 1997; Zrzavý et al., 1998; Littlewood et al., 1999; Ruiz-Trillo et al., 1999). According to these analyses, the Acoela and perhaps also the Catenulida and the Nemertodermatida as well have to be separated from the main group of the platyhelminths, the Rhabditophora. However, the inconclusiveness of the 18S rDNA data is shown by the analysis of Campos et al. (1998), who favor the monophyly of the Platyhelminthes and include the Acoela and the Catenulida on the basis of almost the same data. In the present analysis, only a representative of the Rhabditophora, *Schistosoma,* is included. Therefore, no conclusions about the monophyly of the platyhelminths can be drawn, and the conclusions presented about the phylogenetic relationships of the "plathyhelminths" apply, strictly speaking, only to the Rhabditophora.

As are the Cnidaria and the Ctenophora, which belong to the stem group of the Metazoa, the platyhelminths are characterized by a compact (acoelomate) organization (Ruppert, 1991), a blind-ending intestine, and protostomy. Thus, the basal position of the platyhelminths within the Bilateria is consistent with the hypothesis that the mentioned character states are ancestral within the Bilateria and that the one-

way intestine is a derived character state of the Eubilateria (Hennig, 1979; Ax, 1985). It is not necessary to assume that platyhelminths are derived from ancestors with spacious coelom cavities by a secondary reduction of these cavities in the adult (Siewing, 1980) or by progenesis from larval or juvenile stages (Rieger, 1985; Balavoine, 1998).

Often the platyhelminths have been classified with several other phyla (e.g., molluscs, annelids, arthropods) as Spiralia or Eutrochozoa because of the spiral quartet cleavage and the mesoderm formation by the 4d mesentoblast or one of its daughter cells (Siewing, 1980; Nielsen, 1995; Valentine, 1997). However, the tree favored in the present analyses is not consistent with a clade Spiralia. If this tree is correct, the spiral quartet cleavage is either a synapomorphy of the Bilateria (perhaps except the Acoela) or has been achieved convergently in the platyhelminths and the Eutrochozoa. The convergent origin of the spiral cleavage is less likely, even if one considers that this cleavage pattern results in the thermodynamically most stable arrangement of the blastomeres, because the formation of the mesoderm by the same blastomere can hardly be explained. Therefore, a determinate spiral quartet cleavage was probably ancestral within the Bilateria and has secondarily been modified in several lineages of the Eubilateria (Ax, 1985).

In the maximum likelihood tree, the Eubilateria split into the nematode lineage with a compact organization on the one hand and the lineage of arthropods and chordates on the other. This result is also supported by a four-cluster analysis of concatenated amino acid sequences of 18 nuclear genes by Wang et al. (1999).

The arthropods, the chordates, and several other bilaterian phyla (e.g., molluscs, annelids) often have been classified as Coelomata (Hennig, 1979). However, it is still questionable whether the various coelom cavities (or their rudiments) are homologous (Ruppert, 1991; Nielsen, 1995).

Ecdysis, which has been taken for a synapomorphy of a supposed major clade of the Metazoa, the Ecdysozoa (Eernisse et al., 1992; Aguinaldo et al., 1997), has probably evolved convergently in the nematode lineage (probably comprising some of the smaller groups formerly included in the aschelminths) and in the arthropods.

The phylogenetic relationships of the platyhelminths and of the nematodes could not be resolved on the basis of the 18S rDNA sequences (Figs. 1f-j), partly because of the unusually high substitution rates of the 18S rDNA in platyhelminths and nematodes (Aguinaldo et al., 1997; Balavoine, 1997). Aguinaldo et al. (1997) tried to solve this problem by looking for platyhelminths and nematodes with apparently low substitution rates. They calculated the number of substitutions per position from the last common ancestor of protostomes, using a cnidarian and an echinoderm as outgroups. That means they assumed the protostomes are monophyletic and the echinoderms are an outgroup to the protostomes. However, several previous phylogenetic analyses—morphological (Hennig, 1979; Ax, 1985; Schram, 1991) as well as molecular (Riutort et al., 1993; Philippe et al., 1994; Sidow and Thomas, 1994; Winnepenninckx et al., 1995)—have shown that the protostomes are not monophyletic and that the echinoderms are not an outgroup of the protostomes. This conclusion is also supported by the present data. Therefore, the distances calculated by Aguinaldo et al. (1997) are not distances between the last common ancestor of the protostomes (which is the last common ancestor of all Bilateria) and the respective taxa but instead reflect to a certain degree the convergent similarities of the analyzed platyhelminths and nematodes with the annelids and priapulids, which Aguinaldo et al. used for the calculations. Their results, that the Protostomia are monophyletic, is a consequence of their biased taxon selection and reflects their basic assumption. The exclusion of the apparently more rapidly evolving taxa may have resulted in systematic errors in tree reconstruction. In accordance with other recent analyses (Abouheif et al., 1998), the result of the present study—of several nuclear protein-coding genes that are characterized by more homogeneous substitution rates across the analyzed taxa—suggests that the 18S rDNA sequences alone are insufficient for a stable phylogenetic reconstruction of the deep branches within the Metazoa.

The phylogenetic results of the present analysis imply that genetic features found in the model organisms *Caenorhabditis* and *Drosophila* might be found in all Eubilateria, if the features have not been modified secondarily. In contrast, common features of *Drosophila* and mouse or human are not necessarily present in other metazoans. This is true, even if the phylogenetic position of the platyhelminths in the favored tree is incorrect. These important predictions of the generality of genetic features found in model organisms are contrary to the predictions of Aguinaldo et al. (1997) and highlight how important the knowledge of the correct topology of the tree of life is.

The phylogenetic results also allow some conclusions on the evolution of the *Hox* complex. At least seven genes in the *Hox* cluster of the rhabditophoran platyhelminths are orthologous to cognate groups of the coelomates (Bartels et al., 1993; Balavoine, 1997, 1998). Thus, there were at least seven genes in the *Hox* cluster of the common ancestor of the rhabditophoran platyhelminths and the coelomates, because the rhabditophoran platyhelminths are inferred to be the sister group of the other analyzed Bilateria (Fig. 2). Consequently, at least three *Hox* genes must have been lost in the lineage of the nematode *Caenorhabditis,* of which the *Hox* cluster includes only four genes (Bürglin et al., 1991; Kenyon and Wang, 1991). These conclusions are in agreement with the results of a phylogenetic analysis of the homeodomain sequences of the *Hox* genes of nematodes and coelomates (Zhang and Nei, 1996). Although the nematodes are more closely related to the coelomates than are the platyhelminths (Fig. 2), the distances between the homeodomain sequences of the *Caenorhabditis Hox* genes and the *Hox* genes of either the platyhelminths or the coelomates are greater than those between the respective sequences of the platyhelminths and the coelomates (Bartels et al., 1993). The loss of *Hox* cluster genes and the divergent evolution of the remaining *Hox* genes in the nematode lineage might be the consequence of a different selection acting on the *Hox* cluster as a result of the acquisition of a different, strongly determinative development mode in the nematodes.

Time Frame of Bilaterian Evolution

Although some divergence times of metazoan groups have been estimated by means of the molecular clock (Wray et al., 1996; Feng et al., 1997; Nikoh et al., 1997; Ayala et al., 1998; Gu, 1998; Wang et al., 1999), the divergence time between platyhelminths and eubilaterians has not been investigated so far. However, this divergence is of extraordinary importance, because it marks the beginning of the bilaterian radiation. Therefore, datings of this and other evolutionary divergences were estimated on the basis of the clocklike evolving genes of the present data set.

The results of the likelihood ratio tests for equal amino acid substitution rates among lineages are listed in Table 3. In the analyses with *Arabidopsis* as the only outgroup, the clock hypothesis cannot be rejected for five or seven genes, the specific genes depending on the chosen substitution and rate heterogeneity models. In the analyses with *Saccharomyces* as an additional outgroup, the aldolase and the calreticulin genes have to be omitted, because they are not known from *Saccharomyces.* Considering the remaining genes, the clock hypothesis cannot be rejected for only two, three, or five genes. This is partly because of an accelerated substitution rate in *Saccharomyces,* which becomes apparent also in the phylogram (Fig. 2b). Because of the few suitable genes available, the analyses that included *Saccharomyces* are less reliable than those without *Saccharomyces* and are not considered further.

The estimations of the divergence times based on the maximum likelihood branch lengths of the tree estimated under the clock hypothesis with the amino acid sequences of the clocklike evolving genes (Table 4) vary widely, depending on the chosen substitution and rate heterogeneity models. Nevertheless, all estimates agree in two important points: The major divergences occurred in the Proterozoic before the Vendian, and these divergences were spread over several hundred million years. The estimations based on uniform rates among sites and those based on the BLOSUM 62 substitution matrix and gamma-distributed rates suggest that the diver-

TABLE 3. Results of the likelihood ratio tests for equal amino acid substitution rates among lineages. The first symbol refers to the analyses without *Saccharomyces,* the second to those with *Saccharomyces.* +, the clock hypothesis has not been rejected on a significance level of 5%; −, the clock hypothesis has been rejected on a significance level of 5%; *, data not available.

	Uniform rate, all sites		Gamma-distributed rates (8 categories)	
	PAM 250	BLOSUM 62	PAM 250	BLOSUM 62
Aldolase	+/*	+/*	+/*	+/*
Calreticulin	+/*	+/*	+/*	+/*
Elongation factor 1α	+/−	+/−	−/−	+/−
Enolase	−/+	−/−	−/−	−/+
Glyceraldehyde-3-phosphate dehydrogenase	+/+	+/+	+/+	+/+
3-Hydroxy-3-methylgutaryl coenzyme A reductase	+/−	+/−	−/−	−/−
Phosphoglycerate kinase	−/−	−/−	−/−	−/−
Ribosomal protein L13E	+/+	+/+	+/+	+/+
t-complex polypeptide-1α	+/−	+/+	−/−	+/+
Triose phosphate isomerase	−/−	−/−	+/−	+/+
Clocklike evolving genes	7/3	7/3	5/2	7/5

gence between the rhabditophoran platyhelminths and the Eubilateria is more than a billion years old, that the divergences between the nematode lineage and the coelomates occurred at the beginning of the Neoproterozoic, and that the divergences between the arthropod and the chordate lineages occurred in the middle Neoproterozoic. The estimations based on the PAM 250 substitution matrix and gamma-distributed rates with only five clocklike genes suggest even older divergence times. The estimations are underestimates rather than overestimates because of the conservative fixing of the rodent-primate divergence at 100 MYBP. Thus, the general conclusion of a Proterozoic origin of the Bilateria long before the Cambrian and a prolonged radiation of bilaterian phyla is reliable, even if the confidence intervals might be large because of uncertainties in the calibration, which could not be calculated because only a single calibration point was available.

The estimation of a middle Neoproterozoic divergence between the arthropod and the chordate lineages is in accord with some of the previous datings based on

TABLE 4. Estimation of divergence times (±s.d.), MYBP, based on the maximum likelihood branch lengths of the trees estimated under the clock hypothesis with the amino acid sequences of the clocklike evolving genes. The given standard deviations are only the errors in the estimation of the branch length and do not take into account uncertainties in the dating of the divergence between rodents and primates, which was used as reference.

	Uniform rate, all sites		Gamma-distributed rates (8 categories)	
	PAM 250	BLOSUM 62	PAM 250	BLOSUM 62
Rodents–primates	100	100	100	100
Arthropods–chordates	715 ± 24	700 ± 23	1,171 ± 53	810 ± 28
Nematodes–coelomates	941 ± 23	914 ± 22	1,445 ± 48	1,002 ± 26
Platyhelminths–eubilaterians	1,114 ± 27	1,061 ± 25	1,678 ± 60	1,156 ± 31
Plants–animals	1,334 ± 37	1,263 ± 35	2,575 ± 123	1,544 ± 53

other data, namely, 730 MYBP (Feng et al., 1997), 610 MYBP (Ayala et al., 1998), 736 MYBP (Ayala et al., 1998), and 830 MYBP (Gu, 1998). However, still other studies resulted in older divergence times. The divergence of arthropods and chordates has been estimated at 993 MYBP by Wang et al. (1999) and ~1,200 MYBP by Wray et al. (1996). The divergence times of the protostomes and deuterostomes estimated by Bromham et al. (1998), using 18S rDNA sequences, are even distinctly higher.

The estimation of an early Neoproterozoic divergence between the nematode lineage and the coelomates is intermediary between the estimate by Feng et al. (1997), 815 MYBP, and that by Wang et al. (1999), 1,177 MYBP. The same is true for the estimation of a middle or early Mesoproterozoic animal-plant divergence, which was estimated at 1,200 MYBP by Feng et al. (1997) and at 1,576 MYBP by Wang et al. (1999). However, the reliability of the present estimate of the divergence time of animals and plants is questionable, because of the inability to test whether the outgroup evolved at the same rate as the ingroups.

The divergence time estimates calculated by Wang et al. (1999) that are based on those genes for which their relative rate tests could not show rate heterogeneity among lineages are always smaller than their estimates based on all genes, i.e., including genes shown to evolve with different rates in different lineages. This indicates a bias in the evolution of the substitution rates. The substitution rates within the vertebrates, which were used for the calibrations, are generally smaller than those in the other groups examined (Ayala et al., 1998). This can also be seen in the phylograms based on all 10 amino acid sequences used in this study (Fig. 2) and in comparison of the branch length between the trees calculated without clock assumption and with clock assumption, if only the clocklike evolving genes are considered. If the substitution rates within the vertebrates are generally smaller than those in the other examined groups, the divergence times based on calibrations with vertebrate divergences are overestimations. The overestimation will be the greater, the greater the difference between vertebrate and invertebrate rates for a given gene. The more stringent the criteria for the exclusion of genes that are not clocklike evolving, the less the described bias will influence the estimation of divergence times.

The global tree-based likelihood ratio test used in this study to examine the clocklike evolution of a gene is a stringent test (Sanderson, 1998), because it examines the substitution rates in all lineages at the same time, whereas the relative rate tests used in many other studies (Wray et al., 1996; Gu, 1998; Wang et al., 1999) compare only the rates in two lineages at one time and have low statistical power (Ayala et al., 1998).

The differences between the results of the various studies of metazoan divergence times can partly be explained by the different exclusion criteria for the genes used in these studies. The studies using DNA sequences of the 18S rDNA and the mitochondrial genes, which are known not to evolve in a clocklike way (Ayala et al., 1998), resulted in the oldest divergence dates (Wray et al., 1996; Bromham et al., 1998), whereas the estimates based on amino acid sequences of nuclear genes (Feng et al., 1997; Nikoh et al., 1997; Ayala et al., 1998; Gu, 1998; Wang et al., 1999), which have been examined for clocklike substitution rates, are in better agreement with the estimates of the present study. The estimates based on rDNA are additionally biased by the correlated evolution of paired sites, which probably results in an increase in the count of independent changes between two sequences and thus causes an overestimation of divergence times (Bromham et al., 1998).

The present divergence time estimations differ from previous estimations in that they are based on branch length. The use of branch length has the advantage, that deviations from the assumption of rate homogeneity between lineages are corrected for by the maximum likelihood method. If divergence time estimations are based on pairwise distances, as in most previous studies, differences between the pairwise distances relevant for one divergence time (e.g., *Mus-Drosophila, Homo-Drosophila*) can be corrected for only by simple averaging. Calculations based on branch length result in coherent estimates of divergence times, whereas other methods may result in para-

doxical estimates. For example, Gu (1998) found that if the hydroxy-3-methylglutaryl coenzyme A reductase is calibrated with an animal-fungus clock, the calculated divergence of arthropods and vertebrates is older than the animal-fungus divergence. Similarly, the average of the estimated divergence times of echinoderms and vertebrates based on mitochondrial DNA in the study of Bromham et al. (1998) is greater than the average of their estimations based on mitochondrial DNA for the split of protostomes and deuterostomes.

Despite the differences in the applied methods and the differences in the absolute time estimates, the previous molecular clock studies are compatible with the two major conclusions of the present study, namely, the occurrence of the divergences between the major lineages of the Bilateria in the Proterozoic before the Vendian, and the extension of these divergences over several hundred million years.

The divergence time of the rhabditophoran platyhelminths and the Eubilateria more than a billion years ago represents the minimum age of the start of the bilaterian radiation, because the evolution of the Bilateria probably began with a radiation of compact (acoelomate) phyla, of which only a few (e.g., rhabditophoran platyhelminths) are still extant.

The early emergence of bilaterians is consistent with the interpretation of trace fossils more than one billion year old as evidence for the presence of bilaterians (Breyer et al., 1995; Seilacher et al., 1998) and with the hypothesis that the decline in the diversity of stromatolites starting about one billion years ago was due to grazing bilaterians (Walter and Heys, 1985). Furthermore, the estimated divergence time of the arthropods and chordates more than 700 MYBP is consistent with the interpretation of some Ediacaran fossils as arthropods and echinoderms (Conway Morris, 1993).

The realization of a more than 500-million-year-long Precambrian history of the Bilateria exposes the Cambrian explosion as the near coincidence of the appearance of large soft-bodied animals, the mineralization of preexisting organic substrates in already complex soft-bodied animals of diverse bilaterian lineages, and a secondary radiation of the arthropods, some eutrochozoans, and the deuterostomes (Runnegar, 1982).

Acknowledgments

I am grateful to the referees and the editors for helping me to develop a more critical outlook to the results and conclusions of this study. Furthermore, I thank Martin Hingston (Hamburg) for correcting the English text.

References

Abouheif, E., R. Zardoya, and A. Meyer. 1998. Limitations of metazoan 18S rRNA sequence data: Implications for reconstructing a phylogeny of the animal kingdom and inferring the reality of the Cambrian explosion. J. Mol. Evol. 47:394–405.

Aguinaldo, A. M. A., J. M. Turbeville, L. S. Linford, M. C. Rivera, J. R. Garey, R. A. Raff, and J. A. Lake. 1997. Evidence for a clade of nematodes, arthropods and other moulting animals. Nature 387:489–493.

Archibald, J. D. 1996. Fossil evidence for a late Cretaceous origin of "hoofed" mammals. Science 272: 1150–1153.

Ax, P. 1985. The position of the Gnathostomulida and Platyhelminthes in the phylogenetic system of the Bilateria. Pages 168–180 *in* The origins and relationships of lower invertebrates (S. Conway Morris, J. D. George, R. Gibson, and H. M. Platt, eds.). Clarendon Press, Oxford, UK.

Ayala, F. J., A. Rzhetsky, and F. J. Ayala. 1998. Origin of the metazoan phyla: Molecular clocks confirm paleontological estimates. Proc. Natl. Acad. Sci. USA 95:606–611.

Backeljau, T., B. Winnepenninckx , and L. De Bruyn. 1993. Cladistic analysis of metazoan relationships: A reappraisal. Cladistics 9:167–181.

Balavoine, G. 1997. The early emergence of platyhelminths is contradicted by the agreement between 18S rRNA and *Hox* genes data. C.R. Acad. Sci. Paris Life Sci. 320:83–94.

Balavoine, G. 1998. Are Platyhelminthes coelomates without a coelom? An argument based on the evolution of *Hox* genes. Am. Zool. 38:843–858.

Bartels, J. L., M. T. Murtha, and F. H. Ruddle. 1993. Multiple *Hox/HOM*-class homeoboxes in Platyhelminthes. Mol. Phylogenet. Evol. 2:143–151.

Borchiellini , C., N. Boury-Esnault, J. Vacelet, and Y. Le Parco. 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.

Breyer, J. A., A. B. Busbey, R. E. Hanson, and E. C. Roy III. 1995. Possible new evidence for the origin of metazoans prior to 1 Ga: Sediment-filled tubes from the Mesoproterozoic Allamoore Formation, Trans-Pecos Texas. Geology 23:269–272.

Bromham, L., A. Rambaut, R. Fortey, A. Cooper, and D. Penny. 1998. Testing the Cambrian explosion hypothesis by using a molecular dating technique. Proc. Natl. Acad. Sci. USA 95:12386–12389.

BÜRGLIN, T. R., G. RUVKUN, A. COULSON, N. C. HAWKINS, J. D. MCGHEE, D. SCHALLER, C. WITTMANN, F. MÜLLER, AND R. H. WATERSTON. 1991. Nematode homeobox cluster. Nature 351:703.

CAMPOS, A., M. P. CUMMINGS, J. L. REYES, AND J. P. LACLETTE. 1998. Phylogenetic relationships of Platyhelminthes based on 18S ribosomal gene sequences. Mol. Phylogenet. Evol. 10:1–10.

CARMEAN, D., AND B. J. CRESPI. 1995. Do long branches attract flies? Nature 373:666.

CARRANZA, S., J. BAGU À, AND M. RIUTORT. 1997. Are the Platyhelminthes a monophyletic primitive group? An assessment using 18S rDNA sequences. Mol. Biol. Evol. 14:485–497.

CONWAY MORRIS, S. 1993. The fossil record and the early evolution of the Metazoa. Nature 361:219–225.

DAYHOFF, M. O., R. M. SCHWARTZ, AND B. C. ORCUTT. 1978. A model of evolutionary change in proteins. Pages 345–352 *in* Atlas of protein sequence and structure, Volume 5, Suppl. 3 (M. O. Dayhoff, ed.). National Biomedical Research Foundation, Washington, D.C.

DIXON, M. T., AND D. M. HILLIS. 1993. Ribosomal RNA secondary structure: compensatory mutations and implications for phylogenetic analysis. Mol. Biol. Evol. 10:256–267.

EERNISSE, D. J., J. S. ALBERT, AND F. E. ANDERSON. 1992. Annelida and Arthropoda are not sister taxa: A phylogenetic analysis of spiralian metazoan morphology. Syst. Biol. 41:305–330.

FELSENSTEIN, J. 1981. Evolutionary trees from DNA sequences: A maximum likelihood approach. J. Mol. Evol. 17:368–376.

FELSENSTEIN, J. 1985. Confidence limits on phylogenies: An approach using the bootstrap. Evolution 39:783–791.

FELSENSTEIN, J. 1995. PHYLIP, version 3.57c. Department of Genetics, Univ. Washington, Seattle.

FENG, D.-F., G. CHO, AND R. F. DOOLITTLE. 1997. Determining divergence times with a protein clock: Update and reevaluation. Proc. Natl. Acad. Sci. USA 94:13028–13033.

FITCH, W. M. 1971. Toward defining the course of evolution: Minimum change for a specific tree topology. Syst. Zool. 20:406–416.

GRAYBEAL, A. 1998. Is it better to add taxa or characters to a difficult phylogenetic problem? Syst. Biol. 47:9–17.

GU, X. 1998. Early metazoan divergence was about 830 million years ago. J. Mol. Evol. 47:369–371.

HASEGAWA, M., AND T. HASHIMOTO. 1993. Ribosomal RNA trees misleading? Nature 361:23.

HASZPRUNAR, G. 1996. Plathelminthes and Plathelminthomorpha-Paraphyletic taxa. J. Zool. Syst. Evol. Res. 34:41–48.

HEDGES, S. B., P. H. PARKER, C. G. SIBLEY, AND S. KUMAR. 1996. Continental breakup and the ordinal diversification of birds and mammals. Nature 381:226–229.

HENIKOFF, S., AND J. G. HENIKOFF. 1992. Amino acid substitution matrices from protein blocks. Proc. Natl. Acad. Sci. USA 89:10915–10919.

HENNIG, W. 1979. Wirbellose I (ausgenommen Gliedertiere). Taschenbuch der Speziellen Zoologie, 4th edition, Volume 2. Fischer, Jena, Germany.

HUELSENBECK, J. P., AND J. J. BULL. 1996. A likelihood ratio test to detect conflicting phylogenetic signal. Syst. Biol. 45:92–98.

JANKE, A., X. XU, AND U. ARNASON. 1997. The complete mitochondrial genome of the wallaroo (*Macropus robustus*) and the phylogenetic relationship among Monotremata, Marsupialia, and Eutheria. Proc. Natl. Acad. Sci. USA 94:1276–1281.

KENYON, C., AND B. WANG. 1991. A cluster of *Antennapedia*-class homeobox genes in a nonsegmented animal. Science 253:516–517.

KISHINO, H., AND M. HASEGAWA. 1989. Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in Hominoidea. J. Mol. Evol. 29:170–179.

KISHINO, H., T. MIYATA, AND M. HASEGAWA. 1990. Maximum likelihood inference of protein phylogeny and the origin of chloroplasts. J. Mol. Evol. 31:151–160.

LI, W.-H., M. GOUY, P. M. SHARP, C. O'HUIGIN, AND Y.-W. YANG. 1990. Molecular phylogeny of Rodentia, Lagomorpha, Primates, Artiodactyla, and Carnivora and molecular clocks. Proc. Natl. Acad. Sci. USA 87:6703–6707.

LITTLEWOOD, D. T. J., K. ROHDE, AND K. A. CLOUGH. 1999. The interrelationships of all major groups of Platyhelminthes: Phylogenetic evidence from morphology and molecules. Biol. J. Linnean Soc. 66: 75–114.

MALEY, L. E., AND C. R. MARSHALL. 1998. The coming of age of molecular systematics. Science 279:505–506.

MCHUGH, D. 1997. Molecular evidence that echiurans and pogonophorans are derived annelids. Proc. Natl. Acad. Sci. USA 94:8006–8009.

NIELSEN, C. 1995. Animal evolution. Interrelationships of the living phyla. Oxford Univ. Press, Oxford, UK.

NIKOH, N., N. IWABE, K. KUMA, M. OHNO, T. SUGIYAMA, Y. WATANABE, K. YASUI, Z. SHI-CUI, K. HORI, Y. SHIMURA, AND T. MIYATA. 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.

PHILIPPE, H., A. CHENUIL, AND A. ADOUTTE. 1994. Can the Cambrian explosion be inferred through molecular phylogeny? Development 1994 (suppl.): 15–25.

RAFF, R. A., C. R. MARSHALL, AND J. M. TURBEVILLE. 1994. Using DNA sequences to unravel the Cambrian radiation of the animal phyla. Annu. Rev. Ecol. Syst. 25:351–375.

RIEGER, R. M. 1985. The phylogenetic status of the acoelomate organization within the Bilateria: A histological perspective. Pages 101–122 *in* The origins and relationships of lower invertebrates (S. Conway Morris, J. D. George, R. Gibson, and H. M. Platt, eds.). Clarendon Press, Oxford, UK.

RIUTORT, M., K. G. FIELD, R. A. RAFF, AND J. BAGU A. 1993. 18S rRNA sequences and phylogeny of Platyhelminthes. Biochem. Syst. Ecol. 21:71–77.

RUIZ-TRILLO, I., M. RIUTORT, D. T. J. LITTLEWOOD, E. A. HERNIOU, AND J. BAGU À. 1999. Acoel flatworms: Earliest extant bilaterian metazoans, not members of Platyhelminthes. Science 283:1919–1923.

RUNNEGAR, B. 1982. The Cambrian explosion: Animals or fossils? J. Geol. Soc. Aust. 29:395–411.

RUPPERT, E. E. 1991. Introduction to the aschelminth phyla: A consideration of mesoderm, body cavities, and cuticle. Pages 1–17 *in* Microscopic anatomy of invertebrates, Volume 4 (F. W. Harrison and E. E. Ruppert, eds.). Wiley-Liss, New York.

SAITOU, N., AND M. NEI. 1987. The neighbor-joining method: A new method for reconstructing phylogenetic trees. Mol. Biol. Evol. 4:406–425.

SANDERSON, M. J. 1998. Estimating rate and time in molecular phylogenies: beyond the molecular clock? Pages 242–264 *in* Molecular systematics of plants II. DNA sequencing (D. E. Soltis, P. S. Soltis, and J. J. Doyle, eds.). Kluwer, Boston.

SCHRAM, F. R. 1991. Cladistic analysis of metazoan phyla and the placement of fossil problematica. Pages 35–46 *in* The early evolution of Metazoa and the significance of problematic taxa (A. M. Simonetta and S. Conway-Morris, eds.). Cambridge Univ. Press, Cambridge, UK.

SEILACHER, A., P. K. BOSE, AND F. PFLÜGER. 1998. Triploblastic animals more than 1 billion years ago: Trace fossil evidence from India. Science 282:80–83.

SIDOW, A., AND W. K. THOMAS. 1994. A molecular evolutionary framework for eukaryotic model organisms. Curr. Biol. 4:596–603.

SIEWING, R. 1980. Das Archicoelomatenkonzept. Zool. Jahrb. Anat. 103:439–482.

STRIMMER, K., AND A. VON HAESELER. 1996. Quartet puzzling: A quartet maximum likelihood method for reconstructing tree topologies. Mol. Biol. Evol. 13:964–969.

TÖNGES, U., S. W. PERREY, J. STOYE, AND A. W. M. DRESS. 1996. A general method for fast multiple sequence alignment. Gene 172:GC33–GC41.

VALENTINE, J. W. 1997. Cleavage patterns and the topology of the metazoan tree of life. Proc. Natl. Acad. Sci. USA 94:8001–8005.

WALTER, M. R., AND G. R. HEYS. 1985. Links between the rise of the Metazoa and the decline of stromatolites. Precambrian Res. 29:149–174.

WANG, D. Y.-C., S. KUMAR, AND S. B. HEDGES. 1999. Divergence time estimates for the early history of animal phyla and the origin of plants, animals and fungi. Proc. R. Soc. London B 266:163–171.

WINNEPENNINCKX , B., AND T. BACKELJAU. 1996. 18S rRNA alignments derived from different secondary structure models can produce alternative phylogenies. J. Zool. Syst. Evol. Res. 34:135–143.

WINNEPENNINCKX , B., T. BACKELJAU, L. Y. MACKEY, J. M. BROOKS, R. DE WACHTER, S. KUMAR, AND J. R. GAREY. 1995. 18S rRNA data indicate that Aschelminthes are polyphyletic in origin and consist of at least three distinct clades. Mol. Biol. Evol. 12:1132–1137.

WRAY, G. A., J. S. LEVINTON, AND L. H. SHAPIRO. 1996. Molecular evidence for deep precambrian divergences among metazoan phyla. Science 274:568–573.

ZHANG, J., AND M. NEI. 1996. Evolution of Antennapedia-class homeobox genes. Genetics 142:295–303.

ZRZAVÝ, J., S. MIHULKA, P. KEPKA, A. BEZDĚK, AND D. TIETZ. 1998. Phylogeny of the Metazoa based on morphological and 18S ribosomal DNA evidence. Cladistics 14:249–285.

Received 8 September 1998; accepted 23 February 1999
Associate Editor: S. Edwards