The Age and Relationships of the Major Animal Phyla

Michael Lynch

Evolution, Vol. 53, No. 2. (Apr., 1999), pp. 319-325.

Stable URL:
http://links.jstor.org/sici?sici=0014-3820%28199904%2953%3A2%3C319%3ATAAROT%3E2.0.CO%3B2-S

Evolution is currently published by Society for the Study of Evolution.

Evolution, 53(2), 1999, pp. 319–325

THE AGE AND RELATIONSHIPS OF THE MAJOR ANIMAL PHYLA

MICHAEL LYNCH
Department of Biology, University of Oregon, Eugene, Oregon 97403
E-mail: mlynch@oregon.uoregon.edu

Abstract.—Given the uncertainties in the fossil record and the paucity of informative morphological characters, there is still considerable uncertainty as to the phylogenetic affinities and times of origins of essentially all of the phyla of animals. A multilocus analysis of amino-acid sequence data for mitochondrial genes suggests that the major triploblast phyla began diverging approximately 630 million years ago. These results support the hypothesis that the so-called Cambrian radiation of animals actually initiated about 100 million years prior to the Cambrian, as the fossil evidence suggests. In addition, phylogenetic analysis supports the monophyly of animals, an early (~900 million years ago) branching off of the cnidarian lineage, the monophyly of deuterostomes and protostomes, and the inclusion of nematodes in the protostome lineage. The results of this study suggest that, with appropriate levels of taxon sampling and a focus on conserved regions of protein-coding sequence, complete mitochondrial genome analysis may be sufficiently powerful to elucidate the genealogical relationships of many of the animal phyla.

Key words.—Animal phyla, Annelida, Arthropoda, Cambrian radiation, Cnidaria, invertebrate phylogeny, Mollusca.

Received November 12, 1997. Accepted December 7, 1998.

A common interpretation of the sudden appearance of most animal phyla in the Cambrian fossil record is the occurrence of an explosive phase of phylogenetic diversification during this period (Gould 1989; Lipps and Signor 1992), initiating about 550 million years ago and coming to completion in as few as 10 million years ago (Bowring et al. 1993). However, an absence of fossils prior to the Cambrian does not preclude the possibility that most phyla were represented by soft-bodied or microscopic forms in their early stages of evolution with a very low likelihood of preservation. The existence of fossils of soft-bodied forms from Vendian deposits of about 560 million years ago as well as of a few forms approaching 900 million years in age is consistent with a Precambrian presence of animals (McMenamin and McMenamin 1989; Weiguo 1994). Nevertheless, there is considerable uncertainty as to whether these ancient fossils of multicellular organisms even represent animals, and assuming they do, it is debatable as to whether they are the progenitors of our current-day phyla or simply remnants of extinct lineages (Bergstrom 1994; Seilacher 1994).

Quantitative resolution of the timing and patterning of the divergence of the animal phyla is highly desirable because a correct phylogenetic framework is essential for the development of hypotheses for the evolution of body plans of these diverse lineages. Given that some uncertainties will always exist as to the affinities of fossil remains and current-day taxa, that a detailed fossil record may simply not exist for many phyla, and that the various phyla are so distinct that few anatomical or embryological features can be used to confidently infer their relationships, the use of historical information inherent in gene sequences provides a useful alternative approach to the problem.

Drawing from comparative data on six mitochondrial and two nuclear genes, Wray et al. (1996) recently challenged the idea of a Cambrian radiation, arguing that the diversification of the animal phyla may have begun 1.0 to 1.2 billion years ago. This conclusion is somewhat surprising in that fossil evidence suggests that the divergence of the animal, plant, and fungal clades occurred approximately 1.1 billion years ago (Knoll 1992; Shixing and Huineng 1995), observations that are at least qualitatively consistent with molecular analyses (Wainwright et al. 1993; Nikoh et al. 1994, 1997; Philippe et al. 1994; Gupta 1995; Saccone et al. 1995; Doolittle et al. 1996; Feng et al. 1997). However, Wray et al. (1996) are not alone in suggesting an ancient origin of animals. Observations of complex fossilized burrows in ancient sediments from India have led Seilacher et al. (1998) to argue that triploblast animals existed at least a billion years ago. The latter study raises obvious questions as to why no other such trace fossils are apparent for the nearly 500 subsequent million years, and one possibility is that the fossils recorded by Seilacher et al. (1998) are actually contained within Cambrian deposits (Brasier 1998). Nevertheless, other indirect arguments support a Precambrian origin of metazoans (Conway Morris 1997, 1998; Fortey et al. 1997), and the central question now appears to be just how far back the diversification of the major animal lineages extends. If the extreme conclusions of Wray et al. (1996) and Seilacher et al. (1998) are correct, then either the radiation of the animal phyla occurred shortly after the divergence of the eukaryotic kingdoms or the age of the major eukaryotic kingdoms has also been greatly underestimated.

The study of Wray et al. (1996) is much more ambitious in scope than previous molecular investigations of the age of the animal phyla. Nevertheless, several aspects of their analyses may have led to substantial errors in the estimated times of divergence. For example, to calibrate molecular clocks for their eight genes, sequence comparisons were made between various taxa with a fairly well-understood paleontological record. The authors acknowledge that rates of sequence divergence may vary over time and among lineages, but they argue that the noise from such heterogeneity will average out over long periods of divergence. However, this is only true if the deviations from the average pattern are randomly distributed over the time scale of clock calibration, which is unlikely to be the case. Of particular concern is the authors' use of mammalian sequences, which are known to evolve at substantially elevated rates in the mitochondrial genome (see review in Lynch and Blanchard 1998). Because the times separating the mammalian taxa were relatively short

compared to those separating other pairs of taxa, the slope of the regression of sequence divergence on time is expected to be reduced by the presence of such sequences, and this would tend to lead to inflated estimates of divergence times for more distantly related taxa. In addition, the regressions employed by Wray et al. (1996) did not account for the phylogenetic nonindependence of comparative data. Both types of problems may be responsible for the substantial positive intercepts (up to 15% protein sequence divergence at time zero) that are present in most of the molecular clock calibrations of Wray et al. (1996). Of equal concern is the fact that the authors obtained their divergence time estimates by linearly extending their regressions to ages up to three times greater than the range of existing data.

The analyses of Wray et al. (1996) were taxonomically limited to vertebrates, echinoderms, arthropods, molluscs, and annelids, presumably because of the relative abundance of sequence data for species from these phyla. Because the genome of *Caenorhabditis elegans* is nearly completely sequenced, nematodes could have been included, but the authors eliminated them because of concerns about their elevated rate of evolution. Moreover, because no nonanimal sequences were included in the Wray et al. (1996) study, one can only wonder what the predicted date of divergence of the major eukaryotic domains would have been if plants and fungi had been included. Inclusion of sequences from these other eukaryotic lineages would have also reduced the problem of temporal extrapolation noted above.

Finally, there is some uncertainty as to the quality of the comparative sequence data in the Wray et al. (1996) study. For taxonomic groups as distant as animal phyla, there are typically many insertions and deletions in protein-coding genes, and this is definitely the case for the loci employed by Wray et al. (1996). Although the authors aligned their sequences with a reasonably powerful algorithm, ClustalW (Thompson et al. 1994), they eliminated from their final analyses only "those positions not represented by all species." Such treatment almost certainly leaves some ambiguously aligned sites in the final analysis because distant sequences with multiple insertion/deletions generally have extended stretches of amino acids (not just isolated sites) for which alignment is very difficult. In a recent reanalysis of the Wray et al. (1996) dataset, Ayala et al. (1998) had to completely exclude one locus, for which they were unable to obtain reliable alignments.

With all these potential uncertainties, it is very difficult to know whether the results of Wray et al. (1996) are biased in an upward or downward fashion or whether the various sources of error essentially cancel, leaving the projected times as reasonable estimates. However, in their reanalysis of the same data, Ayala et al. (1998) obtained substantially different results, dating the divergence of the major phyla to between 600 and 670 million years ago. Still other molecular analyses have suggested times of origins for the animal phyla that are intermediate to these two studies. Using hemoglobin as a molecular clock, Runnegar (1982) suggested that the phyla arose at least 900 to 1000 million years ago. Drawing from a study of 64 nuclear-encoded protein-coding genes, Feng et al. (1997) suggested that the deuterostome-protostome divergence occurred 730 to 850 million years ago, and in a study involving 22 such genes, Gu (1998) estimated the same event to have occurred 830 million years ago. Animal nuclear genomes contain many duplicate genes, and one potential limitation of these latter two studies is uncertainty that orthologous comparisons were made across taxa.

In this paper, I attempt to clarify the approximate times of origin and relationships of several of the major animal phyla by: (1) restricting the analysis to mitochondrial protein-coding genes, for which there is no uncertainty about orthology; (2) restricting the analysis to sites that are unambiguously aligned; (3) weighting the results from different genes by their information content; (4) expanding the analysis to include sequences from cephalochordates, hemichordates, nematodes, cnidarians, platyhelminthes, fungi, and plants; and (5) employing methods of phylogenetic analysis that are insensitive to unequal rates of molecular evolution in different terminal lineages.

Methods

Source of Data

This study is based on the amino-acid sequences for 10 mitochondrially encoded protein genes, all of which were retrieved from the Genbank data repository. Performing an analysis exclusively on mitochondrial genes eliminates all uncertainty about orthology because, in virtually all animals, all protein-coding genes in the mitochondrial genome are present in single copy and presumably always have been. Of the 13 genes encoded in the mitochondrial genome, all but those for ATP8, ND4L, and ND6 were used, the last being too divergent to align confidently. The genes involved were sorted into the phylogenetic groups that are central to this paper: amphibia, ray-finned fishes, cephalochordates, hemichordates, echinoderms, molluscs, annelids, arthropods, nematodes, and cnidarians. In addition, for purposes of rooting the animal phylogeny, all available sequences for fungi and chlorophytes were utilized. Except in the case of cephalochordates, hemichordates, and annelids, there were typically multiple sequences per phylogenetic group.

The use of sequence data in this study was highly conservative. All amino-acid sequences were aligned with the assistance of ClustalW (Thompson et al. 1994), and only portions of the molecules that could be unambiguously aligned in long stretches were retained for analysis. Such stretches were identified as continuous sequence (uninterrupted by insertions or deletions) sandwiched between sites that were invariant across all animals, fungi, and plants. The invariant sites were not included in the final analysis. The complete dataset is available from the author on request.

Estimation of the Number of Substitutions per Site

As an estimator for the number of substitutions per amino acid site separating two sequences, I used the formula

$$\hat{d} = \sum_{i=1}^{k} \frac{\left(\frac{1}{\alpha}+1\right)\cdots\left(\frac{1}{\alpha}+i-1\right)}{(1-I_\infty)^{i-1}i!}\cdot\frac{k^{(i)}}{n^{(i)}}, \qquad (1)$$

where n is the number of amino-acid sites compared, k is the number of differences between two sequences, I_∞ is the as-

ymptotic divergence, and $1/\sqrt{\alpha}$ is the among-site coefficient of variation of the substitution rate. This formula, which I derived as a generalization of that given by Rzhetsky and Nei (1994), eliminates the bias in the estimation of d due to finite numbers of sites as well as that due to variation in the substitution rate among sites. I set $I_\infty = 0.08$ based on previous observations (Lynch and Jarrol 1993), and α was taken to be approximated by 0.5 based on the average of observations in the literature (Kumar 1996; Yang 1996). When the observed fraction of different sites (k/n) is on the order of 0.15 or less, $\hat{d} \simeq k/n$ regardless of the values of I_∞ and α. As k/n increases beyond this point, $\hat{d}$ becomes increasingly greater than the observed divergence, as multiple substitutions per site are accounted for.

When multiple sequences existed for different species within a phylum, as was often the case, the estimate of d differentiating two phyla for that locus was obtained by averaging over all possible pairwise estimates across phyla. Such treatment assumes that the different groups employed in the analyses are truly monophyletic, a condition that was verified throughout by phylogenetic analysis of the data. The sampling variance for each individual estimate of $\hat{d}$ was obtained with

$$\mathrm{Var}(\hat{d}) = \left(1 + \frac{\hat{d}}{\alpha(1 - I_\infty)}\right)^{2(\alpha+1)} \left(\frac{k(n-k)}{n^2(n-1)}\right), \qquad (2)$$

again a generalization that I derived of an expression in Rzhetsky and Nei (1994). For each pair of species, the sampling variance was obtained for each locus using equation (2), and the actual sampling variance for the locus-specific $\hat{d}(x,y)$ for a pair of phyla was then taken to be the average of the species-specific estimates. This is an upwardly biased estimate of the sampling variance because it does not account for the fact that estimates of $\hat{d}$ obtained with different taxa are semi-independent. However, because the phylogenies of species within the phyla examined in this paper are not confidently known, I made no attempt to reduce the pooled sampling variances.

For pooling the data from all loci into a single estimate of divergence, each locus was weighted by the average reciprocal of the sampling variances for the estimates of d for that locus. Letting the weight for locus i be w_i and the estimate of the number of substitutions per site for the locus be $\hat{d}_i(x,y)$, the pooled estimate of divergence for two taxonomic groups (x and y) is then

$$\hat{d}(x, y) = \frac{\sum w_i \hat{d}_i(x, y)}{\sum w_i}. \qquad (3)$$

Weighting with the inverse of the sampling variances for the individual loci yields pooled estimates that should have close to minimal variance.

Phylogenetic Analysis

All phylogenetic analyses were performed with the neighbor-joining algorithm (Saitou and Nei 1987), as implemented by MEGA (Kumar et al. 1993). This method allows for heterogeneity in rates of molecular evolution on different branches of a phylogeny. Because of the composite nature of the analyses, involving multiple loci and with different numbers of taxa within the various phyla for each locus, a conventional bootstrap analysis (over individual amino-acid sites) was not performed. Rather, the data were bootstrapped over loci.

Estimation of Divergence Times

Two temporal benchmarks were used to estimate the divergence times of the animal phyla. First, the divergence of tetrapods and ray-finned fishes is assumed to have begun 430 million years ago (Ahlberg and Milner 1994). Second, fungi and animals are assumed to have initiated divergence 1100 million years ago (Knoll 1992; Shixing and Huineng 1995). The times associated with nodes of the phylogenetic tree intermediate to these two points were then estimated by linear interpolation. This approach requires no quantitative assumptions about the rate of molecular evolution, nor is it influenced by heterogeneity of evolutionary rates in terminal branches of the phylogeny. It does, however, assume an essentially constant rate of evolution on the internal branches separating the animal-fungal and tetrapod-fish nodes. Even with this approach, it is still necessary to place a root at the base of the animal-fungal tree. This was accomplished by including plants in the analysis as an outgroup.

RESULTS

Phylogenetic analysis of the mitochondrial encoded proteins reveals several patterns of interest (Fig. 1). First, the positions of the branch tips for the various lineages indicate a huge degree of heterogeneity in the rate of molecular evolution (Fig. 1, top). Relative to members of the triploblast lineage, cnidarians have evolved at a very reduced rate, the total amount of protein sequence evolution in this lineage being substantially less than that in the triploblast lineage prior to the origin of its component phyla. The rate of evolution in the protostome lineage (containing molluscs, annelids, and arthropods) substantially exceeds that in the deuterostome lineage (containing echinoderms, hemichordates, cephalochordates, and vertebrates). The total amount of sequence evolution that occurred in the protostome lineage prior to the divergence of its component phyla exceeds the total amount of evolution in the deuterostome lineages. The amount of evolution in the nematode lineage is particularly high, exceeding the total amount of evolution remaining in the eukaryotic phylogeny (i.e., since the origin of plants, fungi, and animals).

Second, although several of the internal branch lengths separating the major phyla are very small, bootstrap analysis indicates that the cumulative data from the mitochondrial protein-coding genes are capable of elucidating some genealogical relationships (Fig. 1, bottom). Not surprisingly, the tetrapod-fish clade is supported unequivocally and a terminal echinoderm-hemichordate clade is strongly supported as well. The monophyly of the deuterostome lineages is marginally supported by the analysis, whereas the monophyly of the protostome lineages is very strongly supported. Contained within the latter clade of coelomate phyla is the pseudocoelomate nematode lineage. Finally, a basal position for the diploblast cnidarians is fully supported.

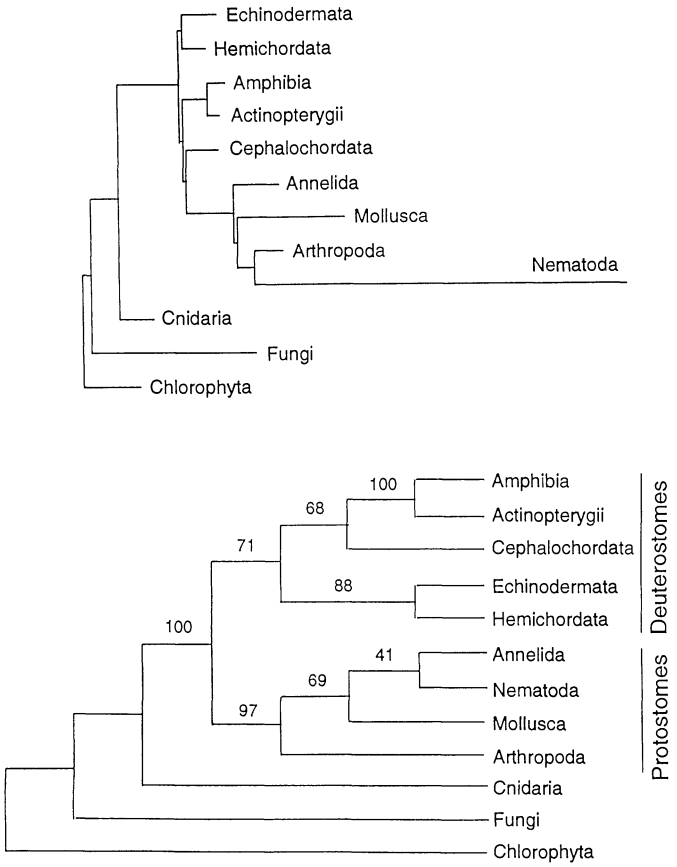

FIG. 1. Top: Phylogenetic tree based on the neighbor-joining methods applied to the pooled distance statistic derived from mitochondrial protein-coding loci. The branch lengths are proportional to the weighted estimates of numbers of amino-acid substitutions per site. Bottom: The most common topology obtained by bootstrapping over loci, with the percent bootstrap support denoted on the internal branches. Here, only the form of the phylogenetic tree is given; the branch lengths are not proportional to evolutionary distance.

Third, a qualitative examination of the phylogenetic tree indicates that a very substantial fraction of the sequence divergence in animals occurred prior to the diversification of this lineage into its major triploblast phyla (Fig. 1, top). Given the huge variation among lineages in the rate of evolution, it is difficult to extrapolate the dates separating the various phyla. For example, the amount of evolution in the amphibian lineage is twice that in the fish lineage since their divergence from a common ancestor, and as noted above, the total amount of evolution in the vertebrate lineage is quite low relative to that in the protostome lineage. Thus, it is difficult to justify the use of evolutionary distances between any terminal groups of species in the animal phylogeny in predicting the dates of the earlier nodes in the phylogenetic tree. For example, if one were willing to assume that the average rate of evolution along the internal branches from the base of the triploblast lineage to the basal vertebrate is equivalent to the average rate of divergence between amphibians and fishes, then the predicted times for the beginning of the triploblast radiation, the diploblast-triploblast divergence, and the animal-fungal divergence are, respectively, 1270, 3310, and 4220 million years, all of which seem wildly unrealistic.

Alternatively, if one is willing to assume that the rate of molecular evolution was roughly constant between the fungal-animal node and the tetrapod-fish node, then estimates of the dates of the intervening nodes can be obtained by interpolation, without making any quantitative assumptions about the rate of molecular evolution. Applying such an analysis to bootstrapped phylogenies, the node ($\pm$ SE) at the base of the triploblast phyla is then estimated to be 627 $\pm$ 51 million years ago, and the node separating diploblasts (cnidarians) and triploblasts is estimated to be 902 $\pm$ 29 million years ago.

DISCUSSION

The phylogenetic tree obtained in this study suggests that a relatively small amount of molecular divergence separates the base of the triploblast animal phylogeny from the split between tetrapods and ray-finned fishes. The latter time point seems to be fairly well established at approximately 430 million years ago. Thus, assuming the animal-fungi divergence date to be 1100 million years ago, then the estimated time of initiation of the divergence of the triploblast animal phyla is approximately 630 million years ago, near the beginning of the Vendian period. This estimate was obtained without making any quantitative assumption about the rate of molecular evolution, and thus it is not dependent on the specific identities of taxa used to calibrate molecular clocks. This estimate does rely on the assumed animal-fungal divergence date being approximately correct, but this latter time would have to be greatly underestimated to place the triploblast divergence at the deep times suggested by Wray et al. (1996). For example, if we assume the animal-fungal divergence to have occurred two billion years ago, which is the current estimated time of the origin of eukaryotes (Han and Runnegar 1992; Knoll 1992; Schopf 1994), then the estimate of 630 million years expands to 900 million years. To achieve an estimate of 1.2 billion years for the initiation of the triploblast radiation, the divergence between animals and fungi would have to date to three billion years, which is not far from the estimated time of origin of life on earth (3.4 billion years in the past).

The validity of the results in this study also depends on the assumption that the rate of molecular evolution remained roughly constant over the time period separating the points of animal-fungal and tetrapod-fish divergence. If, for example, the rate of evolution on the internal branches separating the base of triploblasts from the tetrapod-fish split were unusually low compared to that on the branches leading back from the most recent triploblast to the animal-fungal split, then the time of the triploblast radiation will have been underestimated by linear interpolation. Again, the violation of this assumption would have to be quite extreme to place the divergence of triploblasts at about one billion years. Focusing on the results in Figure 1, a twofold reduction in the average rate of evolution would alter the estimated date from 630 million years to 740 million years. A 10-fold reduction in the rate of evolution is required to extend the date to 970 million years. However, if there was a progressive increase in the rate of evolution over time, then the time to the basal triploblast would be overestimated. Thus, unless the diver-

gence of animals and fungi is substantially more remote in the past than 1.1 billion years or unless the ancestral deuterostome experienced a very pronounced reduction in the rate of molecular evolution, the results of this study imply that the radiation of the triploblast phyla did not initiate much more than 100 million years prior to the Cambrian, whereas the ancestral triploblast dates to about 900 million years ago.

The estimated date for the triploblast radiation suggested here is lower than estimates obtained from several other analyses of molecular data, which range from 700 million years to 1200 million years (Runnegar 1982; Wray et al. 1996; Feng et al. 1997; Gu 1998). However, the 630-million-year date is remarkably compatible with results reported for the mitochondrial analysis in Ayala et al. (1998), which were obtained by a rather different approach. Their analyses considered only five mitochondrial genes, the same ones evaluated by Wray et al. (1996), but the average estimated time of the deuterostome-protostome split using gamma-corrected measure of amino-acid divergence (which is similar to the measure employed herein) is 600 million years (their average estimate based on a Poisson model of divergence is 580 million years). Contrary to the estimated time of divergence obtained in the present study, which ignores the rate of evolution in terminal branches and is not influenced by rate heterogeneity in those regions of the tree, the estimate of Ayala et al. (1998) was obtained by use of molecular clocks calibrated from the divergence of various vertebrate lineages. They made an attempt to eliminate sequences that have evolved at an aberrant rate, but they still assume that the evolutionary rate is constant over the entire phylogeny of triploblasts. That such similar estimates were obtained by such different procedures speaks to the robustness of this result, at least insofar as mitochondrial gene analysis is concerned.

In contrast to these similar results obtained with mitochondrial gene sequences, the average estimated time of the deuterostome-protostome split derived from the analysis of 14 nuclear genes in Ayala et al. (1998) is 780 million years, which is roughly comparable to, although somewhat lower than, estimates derived by Feng et al. (1997) and Gu (1998) using larger sets of such genes. Why do nuclear-gene analyses tend to give greater estimates of divergence times than mitochondrial-gene analyses? One possibility relates to the fact that animal nuclear genomes typically contain large numbers of gene duplicates. This is particularly true in vertebrates, because two complete genome duplications occurred prior to the divergence of tetrapods and ray-finned fishes and many of the duplicates remain functional to the present (Lundin 1993; Nadeau and Sankoff 1997). However, duplicate genes are common in invertebrates as well.

The presence of multiple copies of genes within organisms complicates phylogenetic analyses considerably, because it raises the possibility that nonorthologous genes will be inadvertently included. If, for example, true orthologues are used in the calibration of a molecular clock when the sequences of two species with an unknown divergence time are actually paralogues, then the estimated divergence time will be inflated. The opposite will occur if a molecular clock calibration based on paralogous sequences is applied to true orthologues in a pair of taxa of unknown divergence time. Guigo et al. (1996) provide numerous examples of inconsistent molecular trees for various sets of invertebrate phyla, suggesting that all such inconsistencies are a consequence of gene duplication and the inclusion of nonorthologues in analyses. Although this is an extreme view, it seems likely that analyses based on large numbers of nuclear genes, such as those of Feng et al. (1997), Ayala et al. (1998), and Gu (1998), are somewhat (and perhaps substantially) biased by the inclusion of nonorthologous genes, particularly because the molecular clocks in these studies were entirely or largely based on vertebrate sequences. None of these studies mention any formal attempt to verify orthology.

Although nuclear genomes, which harbor thousands of genes, contain much more information than 10 mitochondrial genes, the problem of orthology will not be easily solved. Correct assignment of orthology requires that sequences of all members of a gene family be available for all species of interest, which is seldom the case, and that the phylogenetic relationships of the species are known in advance, which is certainly not the case in studies of animal phyla with poorly understood affinities. Even then, the stochastic nature of the evolutionary process can result in misassignment of orthology based on sequence identity.

Clarification of the phylogenetic relationships of the major animal phyla has been an elusive problem, with analyses based on different genes and even different analyses based on the same genes yielding a diversity of phylogenetic trees (Table 1). For example, considerable effort has gone into resolving the relationships of the three major protostome lineages (annelids, arthropods, and molluscs). Analyses with 18S rRNA consistently identify annelids and molluscs as sister taxa, as does a morphology-based analysis and the analysis of Guigo et al. (1996). However, results based on partial 28S rRNA sequences are not even consistent with the monophyly of protostomes, suggesting instead a mollusc-deuterostome affiliation. Two studies employing elongation factor 1-α have also yielded different topological relationships involving annelids, arthropods, and molluscs. Similarly, the phylogenetic position of nematodes has been controversial. Of the two studies that are based on 18S rDNA sequences, one positions nematodes basally with respect to the coelomate lineages, while the other joins them with arthropods (Table 1). Aguinaldo et al. (1997) have argued that this inconsistency is an artifact of using nematode sequences that have evolved at unusually high rates. In still another example of phylogenetic inconsistency, in a study involving two protein-coding genes thought to be ideally suited to phylogenetic analysis, Nikoh et al. (1997) found amphioxus to cluster in its expected position with chordates in one case, but to branch off prior to the deuterostome-protostome divergence in the other case.

Given the substantial evolutionary time separating the animal phyla, it is not surprising that single-gene analyses yield such discordant results. Under such circumstances, the statistical noise associated with the substitution process leads to a high probability that phylogenetic analyses based on different molecules will yield different topologies (Philippe et al. 1994; Ruvolo 1997), so that inferences based on single genes can potentially be very misleading (leaving aside for now the additional problem of orthology). Although it is often

TABLE 1. Phylogenetic trees for various combinations of the major animal phyla obtained in previous studies. An, Annelida; Ar, Arthropoda; C, Cnidaria; E, Echinodermata; M, Mollusca; P, Platyhelminthes; and V, Vertebrata. The three results for the study of Carranza et al. (1997) were obtained with different methods of tree reconstruction. The root of the tree is denoted by a dot, whereas parentheses denote nestings of phylogenetic lineages.

Reference	Source of data	Phylogenetic tree
Aguinaldo et al. (1997)	18S rDNA	C.(((ArN)(AnM))E)
Ballard et al. (1992)	12S rDNA	((AnM)Ar).(EV)
Carranza et al. (1997)	18S rDNA	C.(P(Ar((AnM)(EV))))
		C.((Ar(AnM))(P(EV)))
		C.(((AnM)(ArP))(EV))
Halanych et al. (1995)	18S rDNA	C.((Ar(AnM))(EV))
Kim et al. (1996)	18S rDNA	P.(Ar(AnM))
Philippe et al. (1994)	18S rDNA	C.(P((ArM)(EV)))
Winnepenninckx et al. (1995)	18S rDNA	C.(N((Ar(AnM))(EV)))
Christen et al. (1991)	28S rDNA (partial)	C.((ArN)(V(EM))
McHugh (1997)	elongation factor 1-α	((ArM)An).V
Regier and Schultz (1997)	elongation factor 1-α	(AnM).Ar
Guigo et al. (1996)	multiple protein-coding loci	(Ar(AnM)).(EV)
Eernisse et al. (1992)	morphology	C.(((ArN)(AnM))(EV))

argued that the mitochondrial genome evolves too rapidly to be of much use in the analysis of distant phylogenetic relationships, the analyses presented above suggest that a multilocus analysis based on mitochondrial genes may have adequate power to elucidate the phylogenetic affinities of many of the animal phyla. The unambiguous orthology of mitochondrial genes is certainly a favorable property for phylogenetic analysis, although the number of utilizable genes (10, according to the experience of this study) is a limiting factor.

This study took a conservative approach with respect to use of mitochondrial sequence data, employing only those variable sites that could be aligned unambiguously across all animals, fungi, and plants. In addition, the results from each locus were weighted by their information content. The sampling variance of estimates of sequence divergence can vary dramatically from locus to locus depending on the length of the sequence and on the amount of divergence. Weighting by the inverse of the sampling variance helps ensure that the phylogenetic signal provided by the most informative loci is not overwhelmed by the noise associated with the less informative loci and should result in a composite measure with close to minimal sampling variance.

With this approach, and only one to a few taxa sampled per phylum, it was possible to verify the monophyly of animals, the triploblast phyla, and the protostomes. In addition, strong support was obtained for an affiliation between hemichordates and echinoderms and moderate support for the monophyly of the deuterostomes. Finally, the hypothesis of Aguinaldo et al. (1997) that nematodes are a member of the protostome clade was corroborated. The monophyly of molting animals suggested by these authors was not upheld, and additional data will be required to clarify this issue. In summary, the results of this study suggest that a broader taxonomic sampling of the conserved regions of mitochondrial protein-coding genes, perhaps combined with analyses of ribosomal and transfer RNA genes and with additional information on gene order and the genetic code, may be a useful approach to refining our understanding of the genealogical relationships of the animal phyla.

ACKNOWLEDGMENTS

Financial assistance was provided by National Institutes of Health grant RO1-GM36827. I thank A. Boucout, R. Fortey, J. Gray, and G. Wray for numerous helpful comments.

LITERATURE CITED

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 mammals. Nature 387:489–492.

AHLBERG, P. E., AND A. R. MILNER. 1994. The origin and early diversification of tetrapods. Nature 368:507–514.

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.

BALLARD, J. W. O., G. J. OLSEN, D. P. FAITH, W. A. ODGERS, D. M. ROWELL, AND P. W. ATKINSON. 1992. Evidence from 12S ribosomal RNA sequences that onychophorans are modified arthropods. Science 258:1345–1348.

BERGSTROM, J. 1994. Ideas on early animal evolution. Pp. 460–466 *in* S. Bengtson, ed. Early life on earth. Nobel Symp. No. 84. Columbia Univ. Press, New York.

BOWRING, S. A., J. P. GROTZINGER, C. E. ISACHSEN, A. H. KNOLL, S. M. PELECHATY, AND P. KOLOSOV. 1993. Calibrating rates of Early Cambrian evolution. Science 261:1293–1298.

BRASIER, M. 1998. From deep time to late arrivals. Nature 395: 547–548.

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

CHRISTEN, R., A. RATTO, A. BAROIN, R. PERASSO, K. G. GRELL, AND A. ADOUTTE. 1991. An analysis of the origin of metazoans, using comparisons of partial sequences of the 28S RNA, reveals an early emergence of triploblasts. EMBO J. 10:499–503.

CONWAY MORRIS, S. 1997. Molecular clocks: defusing the Cambrian "explosion." Curr. Biol. 7:R71-R74.

———. 1998. Eggs and embryos from the Cambrian. BioEssays 20:676–682.

DOOLITTLE, R. F., D.-F. FENG, S. TSANG, G. CHO, AND E. LITTLE. 1996. Determining divergence times of the major kingdoms of living organisms with a protein clock. Science 271:470–476.

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.

FENG, D.-F., G. CHO, AND R. F. DOOLITTLE. 1997. Determining di-

vergence times with a protein clock: update and reevaluation. Proc. Natl. Acad. Sci. USA 94:13028–13033.
FORTEY, R. A., D. E. G. BRIGGS, AND M. A. WILLS. 1997. The Cambrian evolutionary "explosion" recalibrated. BioEssays 19: 429–434.
GOULD, S. J. 1989. Wonderful life. Norton, New York.
GU, X. 1998. Early metazoan divergence was about 830 million years ago. J. Mol. Evol. 47:369–371.
GUIGO, R., I. MUCHNIK, AND T. F. SMITH. 1996. Reconstruction of ancient molecular phylogeny. Mol. Phylogen. Evol. 6:189–213.
GUPTA, R. S. 1995. Phylogenetic analysis of the 90 kD heat shock family of protein sequences and an examination of the relationship among animals, plants, and fungi species. Mol. Biol. Evol. 12:1063–1073.
HALANYCH, K. M., J. D. BACHELLER, A. M. A. AGUINALDO, S. M. LIVA, D. M. HILLIS, AND J. A. LAKE. 1995. Evidence from 18S ribosomal DNA that the lophophorates are protostome animals. Science 267:1641–1643.
HAN, T. H., AND B. RUNNEGAR. 1992. Megascopic eukaryotic algae from the 2.1-billion-year-old Negaunee iron-formation. Science 257:232–235.
KIM, C. B., S. Y. MOON, S. R. GELDER, AND W. KIM. 1996. Phylogenetic relationships of annelids, molluscs, and arthropods evidenced from molecules and morphology. J. Mol. Evol. 43:207–216.
KNOLL, A. H. 1992. The early evolution of eukaryotes: a geological perspective. Science 256:622–628.
KUMAR, S. 1996. Patterns of nucleotide substitution in mitochondrial protein coding genes of vertebrates. Genetics 143:537–548.
KUMAR, S., K. TAMURA, AND M. NEI. 1993. MEGA: Molecular evolutionary genetics analysis. Vers. 1.01. Pennsylvania State University, University Park, PA.
LIPPS, J. H., AND P. W. SIGNOR, EDS. 1992. Origin and early evolution of the metazoa. Plenum Press, New York.
LUNDIN, L. 1993. Evolution of the vertebrate genome as reflected in paralogous chromosomal regions in man and the house mouse. Genomics 16:1–9.
LYNCH, M., AND J. BLANCHARD. 1998. Deleterious mutation accumulation in organelle genomes. Genetica 102/103:29–39.
LYNCH, M., AND P. JARROL. 1993. A method for calibrating molecular clocks and its application to animal mitochondrial DNA. Genetics 135:1197–1208.
MCHUGH, D. 1997. Molecular evidence that echiurans and pogonophorans are derived annelids. Proc. Natl. Acad. Sci. USA 94: 8006–8009.
MCMENAMIN, M. A. S., AND D. L. S. MCMENAMIN. 1989. The emergence of animals. Columbia Univ. Press, New York.
NADEAU, J. H., AND D. SANKOFF. 1997. Comparable rates of gene loss and functional divergence after genome duplications early in vertebrate evolution. Genetics 147:1259–1266.
NIKOH, N., N. HAYASE, N. IWABE, K.-I. KUMA, AND T. MIYATA. 1994. Phylogenetic relationship of the kingdoms Animalia, Plantae, and Fungi, inferred from 23 different protein sequences. Mol. Biol. Evol. 11:762–768.
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 (Suppl.):15–25.
REGIER, J. C., AND J. W. SCHULTZ. 1997. Molecular phylogeny of the major arthropod groups indicates polyphyly of crustaceans and a new hypothesis for the origin of hexapods. Mol. Biol. Evol. 14:902–913.
RUNNEGAR, B. 1982. A molecular-clock date for the origin of the animal phyla. Lethaia 15:199–205.
RUVOLO, M. E. 1997. Molecular phylogeny of the hominids: inferences from multiple independent DNA sequence data sets. Mol. Biol. Evol. 14:248–265.
RZHETSKY, A., AND M. NEI. 1994. Unbiased estimates of nucleotide substitutions when substitution rate varies among different sites. J. Mol. Evol. 38:295–299.
SACCONE, C., C. GISSI, C. LANAVE, AND G. PESOLE. 1995. Molecular classification of living organisms. J. Mol. Evol. 40:273–279.
SAITOU, N., AND M. NEI. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. 4:406–425.
SCHOPF, J. W. 1994. The early evolution of life: solution to Darwin's dilemma. Trends Ecol. Evol. 9:375–377.
SEILACHER, A. 1994. Early multicellular life: late Proterozoic fossils and the Cambrian explosion. Pp. 389–400 *in* S. Bengtson, ed. Early life on earth. Nobel Symp. No. 84. Columbia Univ. Press, New York.
SEILACHER, A., P. K. BOSE, AND F. PFLUGER. 1998. Triploblastic animals more than 1 billion years ago—trace fossil evidence from India. Science 282:80–83.
SHIXING, Z., AND C. HUINENG. 1995. Megascopic multicellular organisms from the 1700-million-year-old Tuanshanzi formation in the Jixian area, North China. Science 270: 620–622.
THOMPSON, J. D., D. G. HIGGINS, AND T. J. GIBSON. 1994. ClustalW: Improving the sensitivity of practical multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucl. Acids Res. 22:4673–4680.
WAINWRIGHT, P. O., G. HINKLE, M. L. SOGIN, AND S. K. STICKEL. 1993. Monophyletic origins of the Metazoa: an evolutionary link with fungi. Science 260:340–342.
WEIGUO, S. 1994. Early multicellular fossils, Pp. 370–388 *in* S. Bengtson, ed. Early life on earth. Nobel Symp. No. 84. Columbia Univ. Press, New York.
WINNEPENNINCKX, 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. 12:967–979.
WRAY, G. A., J. S. LEVINTON, AND L. H. SHAPIRO. 1996. Molecular evidence for deep preCambrian divergences among metazoan phyla. Science 274:568–573.
YANG, Z. 1996. Among-site rate variation and its impact on phylogenetic analysis. Trends Ecol. Evol. 11:367–372.

Corresponding Editor: R. DeSalle