# Molecular Evidence for Deep Precambrian Divergences Among Metazoan Phyla

Gregory A. Wray,* Jeffrey S. Levinton, Leo H. Shapiro†

A literal reading of the fossil record suggests that the animal phyla diverged in an "explosion" near the beginning of the Cambrian period. Calibrated rates of molecular sequence divergence were used to test this hypothesis. Seven independent data sets suggest that invertebrates diverged from chordates about a billion years ago, about twice as long ago as the Cambrian. Protostomes apparently diverged from chordates well before echinoderms, which suggests a prolonged radiation of animal phyla. These conclusions apply specifically to divergence times among phyla; the morphological features that characterize modern animal body plans, such as skeletons and coeloms, may have evolved later.

Darwin (*1*) recognized that the sudden appearance of animal fossils in the Cambrian posed a problem for his theory of natural selection. He suggested that fossils might eventually be found documenting a protracted unfolding of Precambrian metazoan evolution. Many paleontologists today interpret the absence of Precambrian animal fossils that can be assigned to extant clades not as a preservational artifact, but as evidence of a Cambrian or late Vendian origin and divergence of metazoan phyla (*2*–*6*). This would make the Cambrian the greatest evolutionary cornucopia in the history of the Earth. Definitive representatives of all readily fossilizable animal phyla (with the exception of bryozoans) have been found in Cambrian rocks, as have representatives of several soft-bodied phyla (*6*). Recent geochronological studies have reinforced the impression of a "big bang of animal evolution" by narrowing the temporal window of apparent divergences to just a few million years (*4*).

The evidence for a Cambrian explosion of animal phyla is based on the absence of fossils of triploblastic metazoans from rocks predating the Cambrian. This negative evidence is not entirely convincing. Tiny unskeletonized animals with no possibility of preservation in the fossil record may have existed before the Cambrian (*7*, *8*). Even if larger, soft-bodied animals were present, conditions appropriate for their preservation may not have existed for much of the half-billion years preceding the Cambrian (*6*, *8*, *9*). In particular, the famous Lagerstätten of the Cambrian (*8*, *10*) resulted from taphonomic conditions that are exceptionally rare at other times in the rock record (*9*). Nevertheless, some Vendian trace fossils and body fossils suggest that animals with coeloms existed before the Cambrian (*6*, *8*, *11*, *12*).

Calibrated rates of gene sequence divergence provide another avenue for dating divergence times between animal phyla (*13*). An early study by Runnegar, based on hemoglobin, suggested Precambrian divergences (*14*) but was criticized for not testing assumptions of rate constancy (*15*). A more recent study based on 18S ribosomal RNA (rRNA) sought evidence of rapid divergences in the inability of sequence data to resolve phylogenetic relationships (*16*). The burgeoning database of gene sequences provides an opportunity to examine the divergence times of metazoan phyla from large data sets based on several genes and many taxa. We present such an analysis here. Our results cast doubt on the prevailing notion that the animal phyla diverged explosively during the Cambrian or late Vendian, and instead suggest that there was an extended period of divergence during the mid-Proterozoic, commencing about a billion years ago.

**Calibrating sequence divergence rates.** Our approach to estimating divergence times between metazoan phyla is based on the tendency for nucleotide and amino acid sequences to diverge over time (*17*–*19*). Although rates of sequence divergence vary through time and among taxa (*19*, *20*), in long sequences derived from many phylogenetically dispersed taxa, these heterogeneities average into a mean rate of divergence (*14*, *19*, *21*). Mean rates of sequence divergence, calibrated with the use of taxa with well-established divergence times during the Phanerozoic, can then be used to estimate unknown divergence times (*13*, *21*).

We applied this approach to seven genes: those encoding adenosine triphosphatase (ATPase) subunit 6, cytochrome c, cytochrome c oxidase subunits I and II, hemoglobin, NADH dehydrogenase subunit 1, and 18S rRNA. These genes were chosen for analysis because full-length (or nearly full-length) sequences are currently available from numerous phylogenetically dispersed metazoans. In addition, they do not belong to related gene families, some are nuclear and others mitochondrial, their final products include both RNA and proteins, they encompass a diversity of biochemical functions, and they evolve at different rates. Their evolution should therefore not be correlated, and divergence time estimates based on each gene should be independent.

We calculated mean rates of sequence divergence for each gene from a large number of taxa (Table 1) (*22*). Divergence times were based on a gene's first appearance in the fossil record (*23*, *24*), and gnathostome vertebrates were used for all sequences, with the addition of mollusks and echinoderms for 18S rRNA. For the six protein-coding genes, Kimura distances (*18*) were not significantly different from those obtained with Dayhoff's PAM matrix (*25*); both measures are designed to correct for among-site variation in substitution rate and for multiple substitutions. The Kimura distance for nucleotide sequences, which we used for 18S rRNA, accounts for differential rates of transitions versus transversions. We used the relation between Kimura distance and time to estimate divergence times (*13*, *21*) between the calibrating phyla and representatives of various other metazoan phyla.

Plots of sequence divergence versus divergence time for the seven genes are shown in Fig. 1. Each calibration plot incorporates $10^2$ to $10^3$ comparisons among species pairs (Table 1). For each gene, mean rates of sequence divergence were estimated as slopes with model I regression (*22*, *26*). Explained variation ($r^2$) ranged from 0.60 to 0.75 (Table 1). Separate calibrations with α and β hemoglobins (Fig. 1) yielded very similar mean rates of divergence (Table 1), although many of the species used for comparison differed; this indicates a high degree of repeatability in the calibration.

Gauging the statistical significance of, and confidence limits on, rates of sequence divergence is problematic. The rate is estimated from pairwise comparisons of taxa that are related by descent, which overestimates the number of degrees of freedom. We therefore used three rather different

The authors are in the Department of Ecology and Evolution, State University of New York at Stony Brook, Stony Brook, NY 11794–5245, USA.

*To whom correspondence should be addressed. E-mail: for G. A. Wray, gwray@life.bio.sunysb.edu; for J. S. Levinton, levinton@life.bio.sunysb.edu.

†Present address: Molecular Genetics Program, Department of Zoological Research, National Zoological Park, Washington, DC 20008, USA.

approaches. (i) We determined the significance of the correlation between the divergence time and genetic distance matrices with a Mantel test, which uses sampled randomization to test the correlation between two similarity matrices (*26*). For the six protein-coding genes, this test indicated a correlation between genetic distance and divergence time that is significant at the $P < 0.0001$ level, and for 18S rRNA at the $P < 0.004$ level (Table 1). This test demonstrates rigorously what is clear from visual inspection, namely, that there is a strong positive correlation between sequence divergence and time. (ii) We calculated 95% confidence limits on the slopes of the genetic/time divergence plots using model I regression (*26*), with the degrees of freedom reduced to the number of taxa (approximately the number of nodes on a dichotomous tree). (iii) We bootstrapped amino acid sequences (200 replicates) with PHYLIP (*27*), computed a mean slope, and identified the range of 95% of the slopes. Both the mean bootstrap slope and the 95% range corresponded closely to those calculated from our regression approach (Table 1), which suggests that these are robust estimates of sequence divergence rates.

**Estimating divergence times**. For each gene, we estimated divergence times between phyla by averaging the Kimura distance between each invertebrate and all the vertebrates; we then calculated the implied divergence time from the mean sequence divergence (*13*, *21*) (Fig. 1). Whenever possible, we estimated divergences as the mean of several distantly related species per invertebrate phylum. The invertebrate phyla for which the most sequences are available for comparison with vertebrates are echinoderms, arthropods, annelids, and mollusks; consequently, these phyla are the focus of our analysis.

All mean divergence time estimates between these four phyla and chordates, based on all seven genes, substantially predate the beginning of the Cambrian period (Table

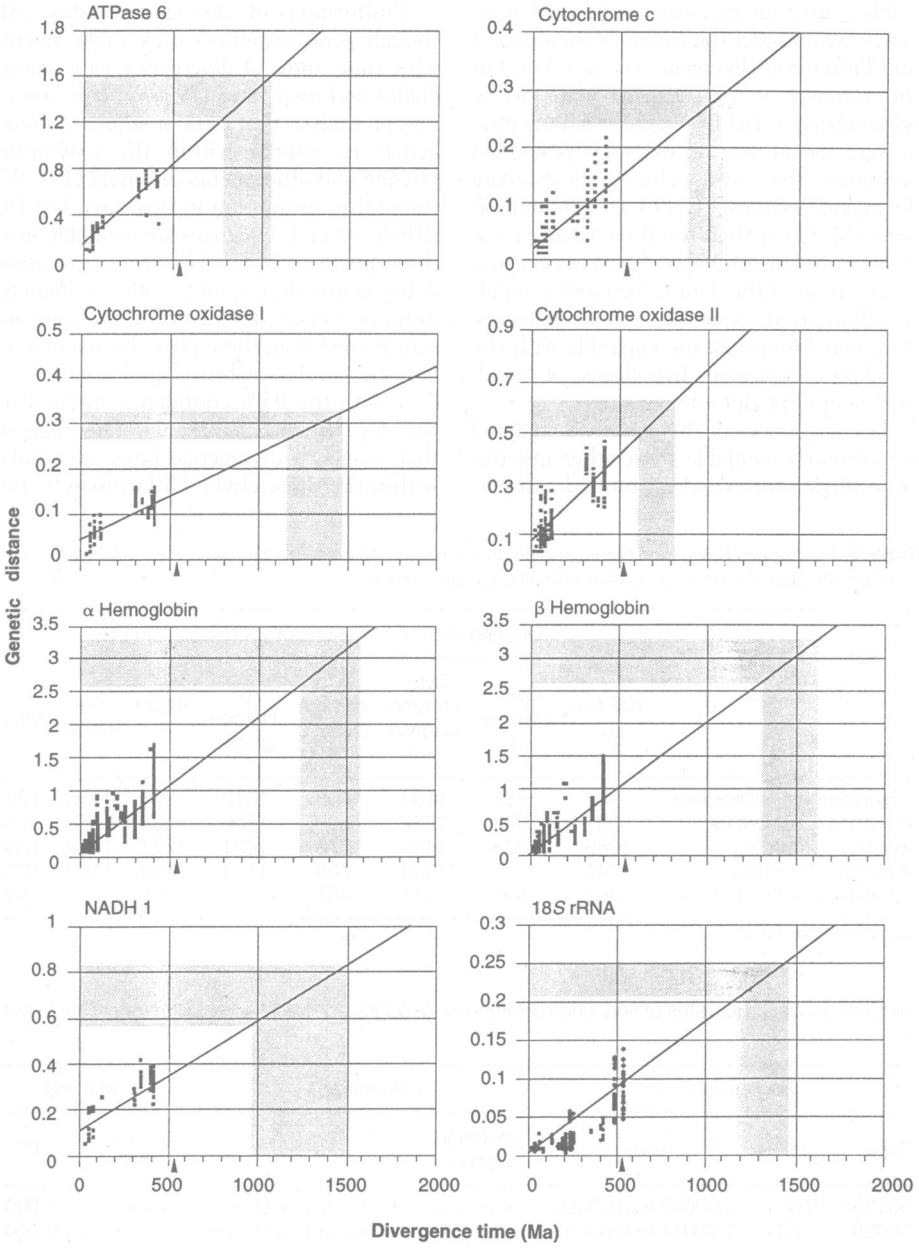


**Fig. 1.** Sequence divergence rates and estimated interphylum divergence times. We calibrated mean rates of sequence divergence (*22*) by plotting pairwise genetic distance versus divergence times from the fossil record (*23*, *24*). The vertebrate fossil record was used to calibrate the six protein-coding genes; vertebrates, echinoderms, and mollusks were used to calibrate 18S rRNA (represented on the plot as squares, circles, and diamonds, respectively). For all seven genes, sequence divergence is strongly correlated with time (Table 1, Mantel test). Hemoglobins α and β diverged just after the origin of chordates (*43*), providing a test of repeatability; regression slopes for these paralogous genes are very similar (see also Table 1). Shaded regions indicate the entire range of invertebrate-vertebrate genetic distances (*y* axis) and the entire implied range of invertebrate-vertebrate divergence times (*x* axis). All estimated invertebrate-vertebrate divergence times are in the Middle to Late Proterozoic, well before the Cambrian (see also Table 2). The base of the Cambrian period (*4*) is marked with an arrowhead.

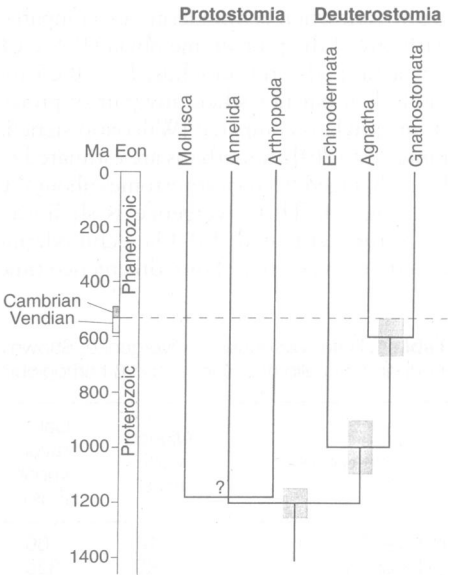


**Fig. 2.** Estimated divergence times for selected metazoan phyla. Mean divergence times based on the seven genes (Table 2) are shown, with standard errors indicated by shaded bars. The three estimated divergence times nest in agreement with well-corroborated phylogenetic relationships (*28*, *31*). The chordate-echinoderm and chordate-protostome divergence times are significantly different from each other (see text). Divergence times among the three protostome phyla were not estimated in our analysis.

2). No individual comparison between the vertebrates and any echinoderm, arthropod, annelid, or mollusk species for any of the seven genes implies a divergence during or after the Cambrian. This alone is significant, in that 89 such comparisons were made. The averaged data (means or medians) imply divergence times of ~1.0 to 1.2 billion years ago (Ga) between these four phyla and the Chordata; in contrast, the base of the Cambrian era lies about 0.54 Ga (*4*).

The mean divergence time estimates between chordates and the three protostome phyla (arthropods, annelids, and mollusks) were all about 1.2 Ga and differed by less than 5% among phyla (Table 2, last column). These three phyla are expected to yield similar estimates, given that they belong to a distinct clade, the Protostomia (*28*) and should therefore share a common divergence time from chordates (Fig. 2). The mean echinoderm-chordate divergence estimate was more recent, about 1.0 Ga. Echinoderm-chordate divergence estimates were shallower than the protostome-chordate divergence for six genes and were similar for 18*S* rRNA (Table 2). Because echinoderms and chordates belong to the same clade, the Deuterostomia (*28*), the shallower divergence estimate between these phyla is again consistent with phylogenetic relationships (Fig. 2). The contrast of divergence times from chordates is statistically significant when echinoderms are compared with any of the protostome phyla ($P < 0.01$ for arthropods and annelids, $P < 0.05$ for mollusks), but not when any pair of protostome phyla is compared [Wilcoxon signed-ranks test (*26*)]. Agnathans are estimated to have diverged from gnathostomes about 0.6 Ga (Fig. 2). This divergence is shallower than the estimated 1.0 Ga echinoderm-chordate divergence. Thus, divergence time estimates based on different genes and different taxa are consistent with each other and with well-corroborated phylogenetic relationships (*28*).

The divergence time estimates between chordates and the phyla just discussed point to two conclusions. (i) The triploblastic metazoan phyla had begun to diverge by the mid-Proterozoic, about twice as long ago as is commonly accepted. Because there is appreciable scatter among estimates from different genes, we consider the general conclusion of mid-Proterozoic divergences to be robust but the accuracy of the estimated dates to be relatively poor. (ii) Divergences among phyla were spread over an extended period. In particular, the mean echinoderm-chordate divergence estimate is 172 to 224 million years (My) later than the three mean protostome-chordate estimates. Again, we consider the trend of the data to be more compelling than are the exact numerical estimates. Both conclusions are incompatible with the Cambrian explosion hypothesis of rapid, shallow interphylum divergences.

Few sequences of the genes we analyzed are currently available from other invertebrate phyla. Nevertheless, limited comparisons were possible for 12 additional phyla, involving 25 additional species-gene combinations. In all cases, divergence time estimates between invertebrates and vertebrates imply mid-Proterozoic divergences. In total, 114 individual invertebrate-vertebrate divergence time estimates were made, spanning 16 invertebrate phyla, and each one indicates a deep Precambrian divergence time.

**Uniformity of divergence rates.** Although gene sequences inexorably diverge with time, rates of divergence vary among clades and over time (*19*, *20*). It is crucial for our analysis that rates of sequence divergence are similar within the calibrating phylum and other metazoan phyla (*15*). We tested this assumption in two ways. For 18*S* rRNA, several sequences are available from three phyla with good fossil records: chordates, echinoderms, and mollusks. Genetic distance versus divergence time comparisons pooled from these phyla form a reasonably tight relationship (Fig. 1) with $r^2 = 0.71$, and the 95% confidence interval on the slope is small (Table 1). This suggests that sequence divergence rates are similar within the three phyla. Unfortunately, this

**Table 2.** Estimated divergence times, shown as the mean of the mean divergence times between each invertebrate species within a phylum and all chordate species.

| Divergence | Divergence time in millions of years according to: | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | ATPase 6 | Cytochrome c | Cytochrome oxidase I | Cytochrome oxidase II | β Hemoglobin | NADH 1 | 18*S* rRNA | Mean |
| Echinodermata—Chordata | 786 | 883 | 1160 | 608 | 1312 | 971 | 1288 | 1001 |
| Arthropoda—Chordata | 887 | 953 | 1272 | 803 | 1506 | 1338 | 1453 | 1173 |
| Annelida—Chordata | 1059 | 1078 | 1465 | 773 | 1621 | 1221 | 1214 | 1204 |
| Mollusca—Chordata | 1045 | – | 1333 | 788 | 1511 | 1492 | 1183 | 1225 |
| Agnatha—Gnathostomata | 462 | 895 | 511 | 487 | –* | 638 | –* | 599 |

*Sequence unavailable.

**Table 1.** Rates of sequence divergence. Shown are calibration statistics for rates of sequence divergence for genes encoding seven different products (with independent calibrations for α and β hemoglobins).

| Gene product | Aligned positions* | Calibration comparisons† | Regression‡ | | | Bootstrap§ | | Mantel¶ | |
|---|---|---|---|---|---|---|---|---|---|
| | | | Slope | $r^2$ | 95% | Average slope | 95% | Correlation | *P* |
| ATPase 6 | 216 | 66 | 0.00139 | 0.66 | 0.00092 to 0.00189 | 0.00141 | 0.00107 to 0.00173 | 0.90 | <0.0001 |
| Cytochrome c | 85 | 325 | 0.00026 | 0.60 | 0.00017 to 0.00035 | 0.00025 | 0.00013 to 0.00039 | 0.78 | <0.0001 |
| Cytochrome oxidase I | 492 | 120 | 0.00019 | 0.69 | 0.00012 to 0.00026 | 0.00019 | 0.00015 to 0.00025 | 0.82 | <0.0001 |
| Cytochrome oxidase II | 206 | 325 | 0.00067 | 0.75 | 0.00051 to 0.00083 | 0.00066 | 0.00051 to 0.00087 | 0.87 | <0.0001 |
| α Hemoglobin | 101 | 1711 | 0.00207 | 0.69 | 0.00171 to 0.00243 | 0.00208 | 0.00174 to 0.00253 | 0.83 | <0.0001 |
| β Hemoglobin | 96 | 1176 | 0.00201 | 0.70 | 0.00163 to 0.00240 | 0.00205 | 0.00158 to 0.00278 | 0.84 | <0.0001 |
| NADH 1 | 273 | 91 | 0.00048 | 0.69 | 0.00028 to 0.00069 | 0.00049 | 0.00037 to 0.00062 | 0.83 | <0.0001 |
| 18*S* rRNA | 1181 | 274 | 0.00015 | 0.71 | 0.00014 to 0.00018 | ND‖ | ND‖ | 0.74 | <0.0039 |

*Number of unambiguously aligned positions (of nucleotides for 18*S* rRNA and of amino acids for all others), excluding all gaps resulting from alignments and missing data. †Number of sequence/divergence time points used for calibration and plotted in Fig. 1. For protein-coding genes, calibrations are based on chordates alone; for 18*S* rRNA, calibration is based on 8 chordates, 21 echinoderms, and 9 mollusk species (regression line fit through three independently derived sets of points). ‡Calculated with the use of model I regression (*26*), with 95% confidence limits on the slope, assuming degrees of freedom equal to the number of calibrating taxa. §Average slope from 200 bootstrap replicates with PHYLIP 3.5 (*27*). ¶Mantel test (*26*) for significance of a positive relation between genetic distance and divergence time, with matrix correlations and significance levels from 10,001 permutations. For 18*S* rRNA, values are for chordate species alone. ‖Not done because calibration involved comparisons within three phyla.

kind of comparison was not possible for the other genes because of the limited number of invertebrate sequences available.

As a more generally applicable approach, we gauged the uniformity of sequence divergence rates among phyla with a simple modification of the relative rate test (*29*), which compares genetic distances from an outgroup species to various ingroup species (Fig. 3A). We used three phylogenetically dispersed nonmetazoans as outgroups to minimize the possibility that an idiosyncratic sequence in the reference taxon might bias the test. Rate variation, as indicated by standard errors of means, was generally low, ranging from 0.5 to 2.0% of the mean, depending on the gene (Table 3, "all metazoans"). In general, calculating the mean interphylum sequence divergences between several species pairs, as we did wherever possible, tends to minimize the effects of rate variation.

Differences between invertebrates and vertebrates in mean rates of sequence divergence are a particular concern because they could introduce a systematic bias in divergence time estimates. Measured against nonmetazoan outgroups, mean invertebrate sequence divergences averaged 1.03 times vertebrate divergences, ranging from 0.88 for cytochrome c to 1.20 for 18S rRNA (Table 3). These gene-specific rate heterogeneities distort invertebrate-chordate divergence estimates based on the vertebrate calibration by −149 to +244 My, depending on the gene and phylum. Furthermore, if average interphylum divergences are estimated with an adjustment for gene-specific rate heterogeneities, they are only 26 to 63 My shallower (depending on the phylum) than those calculated with unadjusted slopes.

The relative rate test turned up significant rate heterogeneities in a few specific comparisons. We conservatively eliminated from consideration those invertebrates that showed consistently faster rates by the relative rate test. The most dramatic case concerned nematodes: Sequences for all seven genes are very divergent relative to all other metazoan phyla, as had been noted earlier for 18S rRNA (*16*). As a result, we did not estimate a nematode-chordate divergence time. Only a few scattered cases of large rate heterogeneities appeared elsewhere. In each case, other species from the phylum or other genes from the same species had more typical sequence divergences and could be used to estimate divergence times.

These results suggest that either the metazoan phyla began to diverge well before the Cambrian or exceptionally high rates of sequence divergence occurred throughout the metazoa (but not in other groups) for a very brief interval during the early Cambrian. Several functionally diverse genes distributed among the nuclear and mitochondrial genomes would have had to experience a very elevated but brief burst of sequence divergence simultaneously and proportionally scaled to the different rates of divergence for each gene. This would have to have occurred independently in several metazoan phyla. In addition, this putative burst of sequence divergence would have to be at least equivalent to all subsequent sequence changes during the rest of the Phanerozoic. Some proponents of a Cambrian explosion have suggested that the metazoan phyla diverged in as little as 8 My (*4*, *6*), which would require sequence divergence rates in Cambrian and post-Cambrian times to differ by at least 65-fold. This is an order of magnitude greater than those of the most rate-variable genes known (*19*).

Theoretical considerations aside, the relative rate test provides empirical evidence against sharply elevated rates of sequence divergence during the Cambrian. We computed sequence divergences from a prokaryote to various metazoan and nonmetazoan eukaryotes for five of the study genes (Fig. 3B and Table 4) (globin and 18S rRNA lack known prokaryotic orthologs). This test is easily sensitive enough to detect a putative Cambrian spike in sequence divergence rates, for the simple reason that metazoan branches would consistently have to be about twice as long as other eukaryote branches in order to explain away the mid-

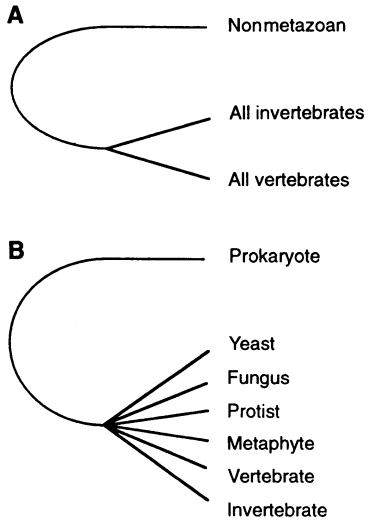


**Fig. 3.** Relative rate tests. Diagrams illustrate the scheme used to test for homogeneity of sequence divergence rates among metazoan phyla (**A**) and among eukaryotic kingdoms (**B**). The test results (Tables 3 and 4) indicate similar branch lengths and therefore similar rates of sequence divergence in both cases.

**Table 3.** Relative rate test among metazoan phyla. For test scheme, see Fig. 3A; numbers are Kimura distances (*18*) from reference taxon.

| Reference taxon | All metazoans | | | Vertebrates | | | Invertebrates | | |
|---|---|---|---|---|---|---|---|---|---|
| | *n* | Mean | SEM | *n* | Mean | SEM | *n* | Mean | SEM |
| | | | | *ATPase 6* | | | | | |
| Metaphyte | 22 | 1.50 | 0.032 | 12 | 1.54 | 0.034 | 10 | 1.60 | 0.058 |
| Yeast | | 1.75 | 0.025 | | 1.68 | 0.030 | | 1.83 | 0.087 |
| Eubacterium | | 1.73 | 0.030 | | 1.65 | 0.030 | | 1.82 | 0.039 |
| | | | | *Cytochrome c* | | | | | |
| Metaphyte | 41 | 0.48 | 0.006 | 27 | 0.50 | 0.006 | 14 | 0.44 | 0.008 |
| Fungus | | 0.49 | 0.011 | | 0.52 | 0.004 | | 0.42 | 0.003 |
| Eubacterium | | 0.77 | 0.006 | | 0.79 | 0.008 | | 0.75 | 0.011 |
| | | | | *Cytochrome oxidase I* | | | | | |
| Metaphyte | 26 | 0.40 | 0.003 | 16 | 0.43 | 0.003 | 10 | 0.40 | 0.007 |
| Fungus | | 0.44 | 0.004 | | 0.45 | 0.003 | | 0.42 | 0.006 |
| Eubacterium | | 0.48 | 0.003 | | 0.47 | 0.003 | | 0.49 | 0.008 |
| | | | | *Cytochrome oxidase II* | | | | | |
| Metaphyte | 48 | 0.75 | 0.009 | 26 | 0.74 | 0.007 | 22 | 0.76 | 0.016 |
| Yeast | | 0.91 | 0.013 | | 0.95 | 0.016 | | 0.86 | 0.017 |
| Eubacterium | | 1.20 | 0.014 | | 1.22 | 0.021 | | 1.17 | 0.017 |
| | | | | *Hemoglobin* | | | | | |
| Metaphyte | 65 | 4.41 | 0.062 | 49 | 4.66 | 0.084 | 16 | 3.21 | 0.083 |
| Protist | | 5.63 | 0.071 | | 5.34 | 0.084 | | 6.29 | 0.475 |
| Eubacterium | | 3.83 | 0.060 | | 3.83 | 0.063 | | 3.83 | 0.155 |
| | | | | *NADH 1* | | | | | |
| Metaphyte | 27 | 0.83 | 0.021 | 14 | 0.76 | 0.013 | 13 | 0.91 | 0.028 |
| Slime mold | | 0.87 | 0.018 | | 0.81 | 0.010 | | 0.94 | 0.024 |
| Eubacterium | | 1.04 | 0.021 | | 0.96 | 0.013 | | 1.13 | 0.022 |
| | | | | *18S rRNA* | | | | | |
| Metaphyte | 51 | 0.31 | 0.004 | 8 | 0.26 | 0.021 | 43 | 0.32 | 0.003 |
| Fungus | | 0.31 | 0.004 | | 0.26 | 0.018 | | 0.32 | 0.003 |
| Alga | | 0.33 | 0.007 | | 0.30 | 0.018 | | 0.34 | 0.008 |

Proterozoic divergence time estimates. This was not the case for any of the five genes tested (Table 4). Given the consistent interphylum divergence time estimates obtained from seven different genes that all seem to have relatively constant rates of sequence divergence, the only reasonable interpretation is that the metazoan phyla began to diverge long before the Cambrian.

A second line of empirical support comes from molecular phylogenies. Although there are substantial difficulties in resolving metazoan relationships with the use of molecular data (*16*), the fact that it is possible to recover even crude phylogenies from sequence data should raise suspicions about the possibility of divergences compressed into as little as 8 My. A Cambrian explosion would result in short internodes followed by long terminal nodes (branch length ratio ~1:65), well into the "Felsenstein zone" from which it is nearly impossible to reconstruct branch order (*30*). Yet phylogenetic analyses of molecular data sets consistently recover echinoderms and chordates as a clade, and typically unite the protostome phyla examined here (*31*). Variation in rates of sequence divergence, which are particularly evident over relatively short intervals of time, would compound the branch length ratio problem, making it even harder to recover topology.

The hypothesis of deep Precambrian divergences makes specific testable predictions. In particular, divergence time estimates based on other genes and taxa should be comparable to those presented here. In addition, new estimates of divergence times should not violate well-corroborated phylogenetic relationships. Few genes have been sequenced at this time in enough species or from a sufficiently broad phylogenetic range to allow estimation of interphylum divergence times. Several genes, however, would become useful with an additional 10 to 20 phylogenetically strategic sequences. This places tests of our hypothesis using other genes within the range of a modest project. The hypothesis of deep Precambrian divergences also makes predictions about the fossil record of metazoans. The stratigraphic ranges of chordates, echinoderms, arthropods, annelids, and mollusks (and, by implication, many other phyla) should be considerably expanded (Fig. 2).

**Implications**. Deep Precambrian divergence times make interpretations of some Neoproterozoic body and trace fossils less problematic. These include interpretations of *Dickinsonia* and *Spriggina* as coelomate triploblasts (*11*, *12*), *Arkarua* as an echinoderm (*32*), *Parvancorina* and *Diplichnites* as arthropods (*10*, *33*), various carbonaceous compressions as annelids and pogonophorans (*34*), and certain microfossils as metazoan fecal pellets (*35*). Our results are also compatible with several indirect lines of evidence that point to Precambrian divergences among metazoan phyla (*8*, *36*). For example, the earliest trilobite fossils fall into distinct biogeographic provinces (*37*) and have morphologies that on well-resolved cladograms place them as highly derived arthropods (*38*).

The existence of an extended but cryptic Precambrian history of metazoans also has some interesting implications for understanding the origin and diversification of animal body architecture. In particular, the rapid appearance of diverse skeletonized taxa in the fossil record during the middle Early Cambrian may reflect an exceptional period of simultaneous morphological innovation within distinct lineages rather than a rapid branching of phyla. It has long seemed likely, for example, that mineralized skeletons evolved independently in several phyla at this time (*39*). It is unlikely, however, that all "body plan" features evolved during the Cambrian. A cephalized, bilaterally symmetrical body composed of three germ layers predates the protostome-deuterostome split (*28*, *40*, *41*) and thus probably evolved much earlier than is generally recognized. Coeloms are shared by the two deuterostome phyla we examined and may predate the Cambrian by several hundred million years (the coeloms of protostomes may have an independent origin, and dating their appearance will require more information about the divergence times of the various protostome phyla).

The genetic regulatory apparatus that is so strikingly and extensively shared by protostomes and deuterostomes must also have evolved long before the Cambrian. This includes the cluster of *Hox* genes that are responsible for regional specification along the anteroposterior axis (*40*), as well as many other genes responsible for patterning and specifying cell fates in various organ systems (*41*). The antiquity of these genetic regulatory circuits suggests that their appearance was not sufficient to trigger the morphological diversification that occurred during the Cambrian, as recently suggested (*42*), although their presence may have been a necessary precondition.

**Table 4.** Relative rate test among eukaryotic kingdoms. For test scheme, see Fig. 3B; numbers are Kimura distances (*18*) from reference taxon.

| Gene product | Yeast | Fungus | Protist | Metaphyte | Chordate | Insect | Echinoderm |
|---|---|---|---|---|---|---|---|
| ATPase 6 | 1.38 | 1.24 | 4.05 | 0.82 | 1.50 | 1.79 | 1.84 |
| Cytochrome c | 0.90 | 0.93 | 0.89 | 0.90 | 0.84 | 0.90 | 0.81 |
| Cytochrome oxidase I | 0.60 | 0.45 | 1.05 | 0.39 | 0.48 | 0.48 | 0.46 |
| Cytochrome oxidase II | 0.99 | 1.18 | 1.97 | 1.01 | 1.16 | 1.05 | 1.13 |
| NADH 1 | 1.27 | 0.85 | 0.93 | 0.71 | 1.06 | 1.16 | 1.17 |

## REFERENCES AND NOTES


1. C. Darwin, *On The Origin of Species by Means of Natural Selection* (Murray, London, 1859).
2. S. J. Gould, *Wonderful Life* (Norton, New York, 1989).
3. J. H. Lipps and P. W. Signor, Eds., *Origin and Early Evolution of Metazoa* (Plenum, New York, 1992).
4. S. A. Bowring *et al.*, *Science* **261**, 1293 (1993).
5. J. W. Schopf and C. Klein, Eds., *The Proterozoic Biosphere. A Multidisciplinary Study* (Cambridge Univ. Press, Cambridge, 1992).
6. J. W. Valentine, S. M. Awramik, P. W. Signor, P. M. Sadler, *Evol. Biol.* **25**, 279 (1991).
7. P. J. S. Boaden, *Zool. J. Linn. Soc.* **96**, 217 (1989); E. H. Davidson, R. A. Cameron, K. J. Peterson, *Science* **270**, 1319 (1995).
8. S. Conway Morris, *Nature* **361**, 219 (1993).
9. N. J. Butterfield, *Lethaia* **28**, 1 (1995).
10. H. B. Whittington, *The Burgess Shale* (Yale Univ. Press, New Haven, CT, 1985).
11. B. Runnegar, *Alcheringa* **6**, 223 (1982).
12. M. F. Glaessner, *The Dawn of Animal Life* (Cambridge Univ. Press, Cambridge, 1984).
13. B. Runnegar, *Palaeontology* **29**, 1 (1986).
14. ———, *Lethaia* **15**, 199 (1982).
15. D. H. Erwin, *ibid.* **22**, 251 (1989).
16. H. Phillipe, A. Chenuil, A. Adoutte, *Dev. Suppl.* **1994**, 15 (1994).
17. E. Zuckerkandl and L. Pauling, in *Evolving Genes and Proteins*, V. Bryson and H. J. Vogel, Eds. (Academic Press, New York, 1965), pp. 97–166.
18. M. Kimura, *The Neutral Theory of Molecular Evolution* (Cambridge Univ. Press, Cambridge, 1983).
19. J. H. Gillespie, *The Causes of Molecular Evolution* (Oxford Univ. Press, Oxford, 1991).
20. R. J. Britten, *Science* **231**, 1393 (1986); W.-H. Li and C.-I. Wu, *Mol. Biol. Evol.* **4**, 74 (1987).
21. R. F. Doolittle, D.-F. Feng, T. Tsang, G. Cho, E. Little, *Science* **271**, 470 (1996); S. B. Hedges, P. H. Parker, C. G. Sibley, S. Kumar, *Nature* **381**, 226 (1996).
22. Sequences were obtained from the National Center for Biotechnology Information repository of the GenBank and SwissProt databases and aligned with ClustalW [J. D. Thompson, D. G. Higgins, T. J. Gibson, *Nucleic Acids Res.* **22**, 4673 (1994)]; alignments were verified by eye, and positions not represented by all species were eliminated from further analysis. Each calibration data set included sequences from a maximum of three taxonomic families per order and one species per family. Sequence divergence matrices were calculated according to Kimura's method (*18*) with PHYLIP 3.5 (*27*). Sequence divergence rates were calculated as the model I regression through all pairwise sequence divergences within the chordates (or chordates, echinoderms, and mollusks for 18*S* rRNA), plotted against divergence times based on first appearances in the fossil record. Regressions were not forced through the origin; the slightly positive intercepts are probably a simple consequence of having to rely primarily on first appearances in the fossil record for our rate calibrations, which tends to cause underestimation of divergence times. Forcing regressions through the origin results in average invertebrate-chordate divergence estimates approximately 100 to 150 My less than

those presented in Table 2, which would not alter our conclusions. Invertebrate-chordate sequence divergences were calculated by taking the mean of the sequence divergence between each invertebrate and all the chordates. Complete species lists, accession numbers, alignments, distance matrices, and divergence time matrices are available on request or from our World Wide Web page (http://life.bio.sunysb.edu/ee/precambrian).
23. A. B. Smith, *Systematics and the Fossil Record* (Blackwell, Oxford, 1994).
24. Stratigraphic ranges were obtained from M. J. Benton, Ed., *The Fossil Record 2* (Chapman and Hall, London, 1993). Space limitations preclude listing all pairwise divergence time estimates based on the fossil record, but some examples follow: Chondrichthyes versus Sarcopterygii or Actinopterygii = 415 million years ago (Ma); Dipnoi or *Latimeria* versus Tetrapoda = 408 Ma; Amphibia versus Archosauria, Lepidosauria, or Mammalia = 349 Ma; Crocodylia versus Aves = 235 Ma; Teleostei versus Chondrostei = 223 Ma; Urodela versus Anura = 208 Ma; Monotrema versus Marsupialia or Eutheria = 160 Ma; Serpentes versus Sauria = 157 Ma; Salmoniformes or Cypriniformes versus Anguilliformes = 152 Ma; Marsupialia versus Eutheria = 112 Ma; Varanidae versus Iguanidae = 83 Ma; Cetacea versus Artiodactyla = 56 Ma; Muridae versus Sciuridae or Caviidae = 42 Ma; Phocidae versus Felidae = 38 Ma; and Hominidae versus Tarsiidae = 38 Ma.
25. M. O. Dayhoff, *Atlas of Protein Sequence and Structure*, vol. 5, suppl. 3 (National Biomedical Research Foundation, Washington, DC, 1978).
26. R. R. Sokal and F. J. Rohlf, *Biometry* (Freeman, New York, ed. 3, 1995).
27. J. Felsenstein, *PHYLIP (Phylogeny Inference Package)*, version 3.5c (distributed by the author, Dept. of Genetics, University of Washington, Seattle, WA, 1993).
28. R. C. Brusca and G. J. Brusca, *Invertebrates* (Sinauer, Sunderland, MA, 1990); C. Nielsen, *Animal Evolution* (Oxford Univ. Press, Oxford, 1995).
29. V. M. Sarich and A. C. Wilson, *Science* **179**, 1144 (1973).
30. J. Felsenstein, *Syst. Zool.* **27**, 401 (1978); J. P. Huelsenbeck and D. M. Hillis, *Syst. Biol.* **42**, 247 (1993).
31. K. G. Field *et al.*, *Science* **239**, 748 (1988); R. Christen *et al.*, *EMBO J.* **10**, 499 (1991); J. M. Turbeville, K. G. Field, R. A. Raff, *Mol. Biol. Evol.* **9**, 235 (1992); H. Wada and N. Satoh, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 1801 (1994); K. M. Halanych *et al.*, *Science* **267**, 1641 (1995).
32. J. G. Gehling, *Alcheringa* **11**, 337 (1987).
33. _____, *Mem. Geol. Soc. India* **20**, 181 (1991); R. J. F. Jenkins, in *Origin and Early Evolution of the Metazoa*, J. H. Lipps and P. W. Signor, Eds. (Plenum, New York, 1992), pp. 181–223.
34. W. Sun, G. Wang, B. Zhou, *Precambrian Res.* **31**, 377 (1986); H. J. Hoffmann, in *Early Life on Earth*, S. Bengtson, Ed. (Columbia Univ. Press, New York, 1994), pp. 342–357.
35. E. I. Robbins, K. G. Porter, K. A. Haberyan, *Proc. Natl. Acad. Sci. U.S.A.* **82**, 5809 (1985).
36. R. A. Fortey, D. E. G. Briggs, M. A. Wills, *Biol. J. Linn. Soc.* **57**, 13 (1996).
37. R. A. Fortey and R. M. Owens, in *Major Evolutionary Radiations*, P. D. Taylor and G. Larwood, Eds. (Clarendon, Oxford, 1990), pp. 139–164.
38. D. E. G. Briggs and R. A. Fortey, *Science* **246**, 241 (1989).
39. B. Runnegar and S. Bengtson, in *Paleobiology: A Synthesis*, D. E. G. Briggs and P. R. Crowther, Eds. (Blackwell, Oxford, 1990), pp. 24–29.
40. D. Duboule and P. Dollé, *EMBO J.* **8**, 1497 (1989); W. McGinnis and R. Krumlauf, *Cell* **68**, 283 (1992).
41. E. M. DeRobertis, in *Guidebook to the Homeobox Genes*, D. Duboule, Ed. (IRL, Oxford, 1994), pp. 13–23; M. P. Scott, *Cell* **79**, 1121 (1994); N. Patel, *Science* **266**, 581 (1994).
42. J. W. Valentine, D. H. Erwin, D. Jablonski, *Dev. Biol.* **173**, 373 (1996).
43. M. Goodman, W. G. Moore, G. Matsuda, *Nature* **253**, 603 (1975).
44. We thank J. Felsenstein, J. Rohlf, R. Sokal, B. Runnegar, A. Bely, J. Boore, and R. Zardoya for invaluable help with these analyses, and A. Bely, D. Futuyma, D. Jablonski, and A. Knoll for helpful comments on the manuscript. Supported by grants from NSF to all three authors and from the A. P. Sloan Foundation to G.A.W.

28 May 1996; accepted 27 August 1996