THE ROYAL SOCIETY

Divergence time estimates for the early history of animal phyla and the origin of plants, animals and fungi

Daniel Y.-C. Wang[1], **Sudhir Kumar**[1,2]† **and S. Blair Hedges**[1,2]*

[1]*Department of Biology and Institute of Molecular Evolutionary Genetics, 208 Mueller Laboratory, Pennsylvania State University, University Park, PA 16802, USA*
[2]*Astrobiology Research Center, Pennsylvania State University, University Park, PA 16802, USA*

In the past, molecular clocks have been used to estimate divergence times among animal phyla, but those time estimates have varied widely (1200–670 million years ago, Ma). In order to obtain time estimates that are more robust, we have analysed a larger number of genes for divergences among three well-represented animal phyla, and among plants, animals and fungi. The time estimate for the chordate–arthropod divergence, using 50 genes, is 993 ± 46 Ma. Nematodes were found to have diverged from the lineage leading to arthropods and chordates at 1177 ± 79 Ma. Phylogenetic analyses also show that a basal position of nematodes has strong support ($p > 99\%$) and is not the result of rate biases. The three-way split (relationships unresolved) of plants, animals and fungi was estimated at 1576 ± 88 Ma. By inference, the basal animal phyla (Porifera, Cnidaria, Ctenophora) diverged between about 1200–1500 Ma. This suggests that at least six animal phyla originated deep in the Precambrian, more than 400 million years earlier than their first appearance in the fossil record.

Keywords: metazoa; molecular clock; Cambrian explosion; plants; fungi

1. INTRODUCTION

Most animal phyla appear in the fossil record during a relatively short interval, the 'Cambrian explosion', about 530 Ma (figure 1). If phylogenetic divergences among the phyla also occurred at that time, it provides an unparalleled example of rapid organismal change. Alternatively, if animal phyla diverged over a long prior time interval then rapid rates of morphological change are not required. Molecular clocks can test these hypotheses, but previous attempts have yielded widely varying divergence times between protostomes and deuterostomes (1200–670 Ma) (Ayala *et al*. 1998; Brown *et al*. 1972; Feng *et al*. 1997; Gu 1998; Runnegar 1982, 1986; Wray *et al*. 1996). Although all time estimates have preceded the Cambrian, the divergence time (670 Ma) obtained in one recent study (Ayala *et al*. 1998) was considered to be consistent with palaeontological evidence and the Cambrian explosion theory. Because of this, there has been considerable uncertainty as to the actual time of divergence and its bearing on the 'Cambrian explosion' model of animal evolution.

Some of this variation among molecular time estimates is due to different methodologies used. For example, some studies (Ayala *et al*. 1998; Feng *et al*. 1997; Wray *et al*. 1996) have used an average of many calibration points, whereas others (Gu 1998) have used a small number of well-supported calibration points. Some studies (Wray *et al*. 1996) have included DNA sequences (18S ribosomal RNA) in their analyses whereas others have used only protein sequences. In some cases (Ayala *et al*. 1998; Wray *et al*. 1996), mitochondrial genes have been examined, whereas other researchers (Feng *et al*. 1997; Gu 1998) have focused on nuclear genes only. Although the methodological differences and their effects on time estimation have been discussed (Ayala *et al*. 1998), another important factor is the number of genes that has been used to estimate divergence time. Because the coefficient of variation of divergence time estimates among genes is large (25–35%), many independent genes should be used to estimate time (Kumar & Hedges 1998). For this reason, we have focused here on obtaining a larger sample of genes to address this evolutionary question.

The phylogenetic relationships of animal phyla are not well established, and most previous molecular studies have focused on sequences from a single gene, usually the 18S ribosomal RNA gene (Aguinaldo *et al*. 1997). Here, we use a larger sampling of genes to explore the phylogenetic relationships of three animal phyla (arthropods, nematodes and chordates) and three kingdoms (plants, animals and fungi).

2. MATERIALS AND METHODS

The protein sequence databases (Genbank) were searched for genes in which divergence times could be estimated between two or more animal phyla, or among plants, animals and fungi. In addition, short (<100 amino acids) sequences were omitted, and sequences of at least one mammal and one bird, for

*Author for correspondence (sbhl@psu.edu).
†Present address: Department of Biology, Arizona State University, Tempe, AZ 85287-1501, USA.

Received 21 August 1998 *Accepted* 14 October 1998

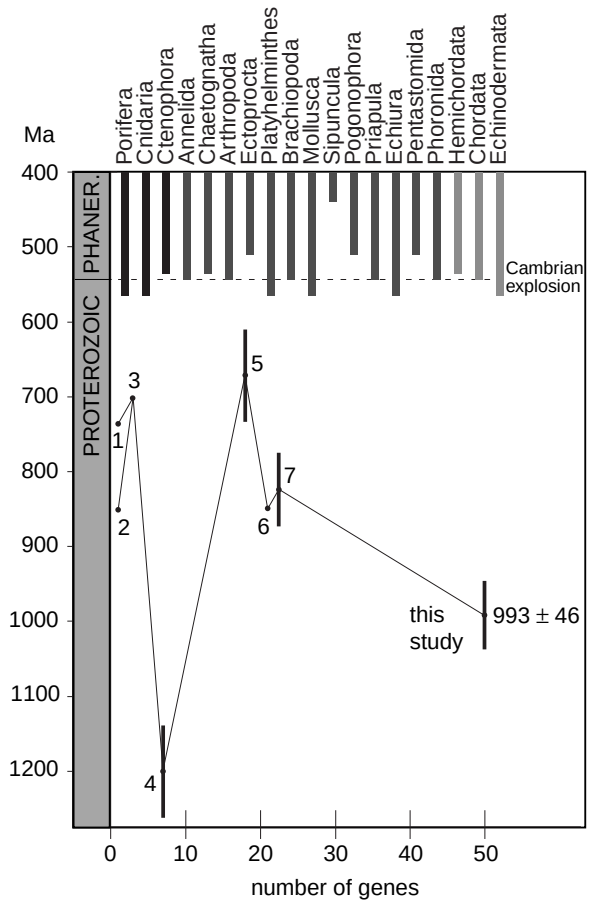

Figure 1. Time of divergence between animal phyla (protostome–deuterostome) as measured by different studies and here (numbers correspond to the following studies: 1, Brown *et al*. 1972; 2, Runnegar 1982; 3, Runnegar 1986; 4, Wray *et al*. 1996; 5, Ayala *et al*. 1998; 6, Feng *et al*. 1997; and 7, Gu 1998. Vertical bars at top indicate fossil record of living phyla (black, basal animal phyla; medium grey, protostomes; light grey, deuterostomes). Only those phyla known from the early Phanerozoic are shown. Standard errors are shown only for those studies where that statistic was reported.

primary calibration, or primate and rodent, for secondary calibration, were required for time calibration (see below). Obvious cases of paralogy were avoided by visually inspecting phylogenetic trees of each gene and omitting genes in which established monophyletic groups (e.g. plants, animals) were paraphyletic or polyphyletic, or where the rate of sequence evolution varied widely among taxa. The removal of outliers (see below) also provided a means of reducing the effect (if present) of gene paralogy.

In all, 75 nuclear genes met these criteria and could be used to estimate divergence times (table 1). Sequences were aligned using Clustal V (Higgins *et al*. 1992) and each alignment was visually inspected and adjusted (if necessary). Rate variation among sites was modelled using a gamma distance, and the gene-specific shape parameters (α) were estimated (Gu & Zhang 1997) for evolutionary distance calculations. In estimating these distances, the complete-deletion method (Kumar *et al*. 1993) was used in the primary analyses, whereby all insertion–deletion sites were excluded from the calculation of distance. For comparison only, the pairwise-deletion method (Kumar *et al*. 1993) also was used. For each taxonomic comparison, representative sequences were placed into clusters and average distances between clusters were calculated (Kumar 1996; Rzhetsky *et al*. 1995).

To estimate divergence time, the fossil-based bird–mammal divergence (310 Ma) was used as a primary calibration. It represents the split between synapsids (ancestors of mammals) and diapsids (ancestors of birds) as evidenced by the earliest representatives of those groups found in Carboniferous sediments at Joggins, Nova Scotia (Benton 1997). It is one of the best calibration points within vertebrates for several reasons: (i) the unique conditions of preservation at this site, in fossilized tree stump holes (natural pitfall traps), has yielded an unusually good fossil record of early amniotes; (ii) the earliest representatives of the two lineages are very similar in morphology, resembling small lizards, suggesting that the actual divergence was not significantly earlier; (iii) no earlier amniote fossils have been found since the discovery of this site in the mid-1800s; and (iv) the transition from fishes to stem amphibians (360–380 Ma) and the morphological gap between the first tetrapods and the first amniotes constrains this date from being significantly earlier (Benton 1997; Kumar & Hedges 1998).

The earliest synapsids and diapsids are well recognized as such and are connected to their descendants (e.g. mammals and birds, respectively) through a series of fossils. On the mammalian side, there is a well-documented transition of synapsids from pelycosaurs to therapsids, cynodonts and finally to mammals (Benton 1993, 1997). On the avian side, there is some disagreement as to which specific lineage of diapsid led to the origin of birds, but there is no dispute that birds descended from diapsid reptiles (Benton 1997). Thus there is strong palaeontological evidence that the divergence between birds and mammals was not more recent than 310 Ma.

For genes where a bird sequence was unavailable, the primate–rodent divergence (110 Ma) was used as a secondary calibration, as described elsewhere (Kumar & Hedges 1998). Although this time estimate comes from a molecular clock study, fossil evidence (Archibald 1996) also supports an early divergence time ($>$90 Ma) for the primate–rodent split. The use of multiple calibration points from the fossil record would be desirable if they were all close to the actual time of divergence. However, most fossil-based divergence times are underestimates, and thus a calibration based on an average of such times would result in a consistent underestimate of divergence time (Kumar & Hedges 1998). For practical reasons, multiple calibration points usually are not possible because the only sequences available for most genes are those of human, mouse, chicken or a few other commonly sequenced organisms.

Rate differences among lineages were examined for each gene (Kumar 1996; Takezaki *et al*. 1995) to identify pairwise comparisons showing significant rate variation (5% level). Single-gene divergence times were estimated using the average-distance method, where the rate was determined using the calibration point (and average distance) and applied to the distance between the two taxa in question (Kumar & Hedges 1998). Divergence times were averaged across all genes and across constant-rate genes (table 2). The upper and lower 5% (at least the highest and lowest) of single-gene time estimates in each comparison were excluded to minimize the influence of outliers (Kumar & Hedges 1998).

3. RESULTS

There were 571 sequences analysed across all 75 genes, and the average number of aligned residues per gene was 658.

Table 1. *Divergence time estimates for individual nuclear genes using a gamma distance with gene-specific shape parameter*

				divergence time (Ma)					
locus name	number of sequences	number of residues	gamma parameter	chordate–arthropod	chordate–nematode	arthropod–nematode	metazoa–fungi	metazoa–plant	fungi–plant
acetyl CoA acyltransferase	7	482	2.03	—	—	—	680[a]	590[b,c]	752[a]
aldehyde dehydrogenase	10	578	2.06	—	—	—	1293	1084	1313
aldolase	9	387	1.52	755	860	658	—	1209	—
alkaline phosphatase	7	622	4.75	1019	—	—	1744[a]	2651[a]	3100[a,b]
alpha actinin	5	937	1.75	1289	1720	1436	—	—	—
amidophosphoribosyltransferase	8	589	2.27	839[a]	—	—	2519	2041	2509[a]
aminolevulinate synthetase	8	755	1.48	1481	—	—	2442	—	—
argininosuccinate lyase	8	524	1.53	—	—	—	692	887	936
argininosuccinate synthetase	5	461	3.80	—	—	—	2146	—	—
asparagine synthase	8	775	2.58	—	—	—	1540	1398	985[a]
aspartate aminotransferase	10	462	2.29	—	—	—	1175	1120[a]	1423
aspartyl tRNA synthetase	5	679	4.06	—	1013	—	1444	—	—
beta adaptin	4	1032	∞[d]	1001	—	—	7126[a,b]	—	—
calcineurin A	5	647	1.09	1079[a]	—	—	2789[a]	—	—
calreticulin	13	434	1.52	940	1273	1309	—	2033	—
catalase	7	709	1.83	430[b]	491[b]	448[b,c]	862[a]	1103	1380
dihydrofolate reductase	8	241	2.72	1044[a]	—	—	1486[a]	—	—
disulfide isomerase	11	563	2.83	1257	1638	1847[a]	2417[a]	2160[a]	2490[a]
DNA polymerase alpha	5	1633	5.53	978[a]	—	—	1248	—	—
DNA polymerase delta	5	1152	1.98	621[a]	—	—	897	—	—
DNA polymerase gamma	5	1546	1.60	1301[a]	—	—	1816[a]	—	—
DNA topoisomerase 1	9	1317	2.43	1224	1565[a]	1683[c]	2921[b,c]	3080[c]	3044[c]
DNA topoisomerase 2	8	1661	1.88	1173	2412[a,b]	2567[a,b]	1999	1858[a]	1770
dopa decarboxylase	6	567	1.51	528	—	—	—	1028[a]	—
enolase	15	456	1.05	1385	—	—	2260	1627	2394
epoxide hydrolase	5	492	7.43	607	495	662	—	1199[a]	—
er60	10	566	3.76	—	—	—	1287[a]	1281	1359[a]
ferritin heavy chain	12	275	2.84	5220[a,b]	—	—	—	4712[b,c]	—
fructose-2,6-bisphosphatase	6	497	2.19	—	—	—	1254[a]	1426[a]	1538[a]
fumarase	7	516	0.71	—	—	—	1847	1830	1882
furin	5	1073	1.15	3119[a,b]	—	—	—	—	—
glucose 6-phosphate dehydrogenase	10	752	1.20	745	—	—	1405	1367	1367
glucose 6-phosphate isomerase	8	598	1.52	364[b,c]	—	—	557[a,b]	893[a]	918
glutamate dehydrogenase	5	561	1.65	1144	—	—	—	3722[a,b]	—
glutamic acid decarboxylase	6	641	60.16	1983[b,c]	—	—	—	6308[a,b]	—
glutamine synthetase	12	508	1.42	1070	—	—	1404	1627	1632
glyceraldehyde 3-phosphate dehydrogenase	18	353	0.61	1217	1159	1399	1827	1874	1856
glycine dehydrogenase	10	525	1.41	—	—	—	1472	1262	1428

(*Cont.*)

Table 1. (*Cont.*)

locus name	number of sequences	number of residues	gamma parameter	divergence time (Ma)					
				chordate–arthropod	chordate–nematode	arthropod–nematode	metazoa–fungi	metazoa–plant	fungi–plant
glycine hydroxymethyl transferase	6	1109	0.87	—	826	—	1073[a]	904[a]	854[a]
hexokinase	8	961	2.61	—	—	—	1571	1517	1534
histidine decarboxylase	4	855	∞[d]	587	—	—	—	1090[a]	—
HMG-CoA reductase	12	1189	0.82	822	—	—	1051	1129	1010
heat shock protein (Hsp) 70	7	659	0.54	148[b,c]	—	—	201[b,c]	193[a,b]	199[b,c]
inosine monophosphate dehydrogenase	6	557	1.88	1480	—	—	2014	3479[a]	3791[a,b]
lactate dehydrogenase	7	357	2.05	1078	1110	1047	—	1691	—
methionine adenosyl transferase	8	443	0.92	1146[a]	1146[a]	1137	1506	2049	2022
monoamine oxidase	5	772	∞[d]	—	—	—	1175	—	—
N-acetylglucosaminyltransferase I	4	479	∞[d]	—	1108	—	—	—	—
Na-K ATPase alpha chain	5	1046	0.61	1605[c]	—	—	—	—	—
Na-K ATPase beta chain	5	467	536.17	497	—	—	—	—	—
NADP-aldehyde reductase	4	379	∞[d]	—	—	—	1800	—	—
nucleoside diphosphate kinase	10	154	1.00	893	1897[b,c]	1683	2066	1983	2237
ornithine aminotransferase	7	486	1.85	444	—	—	976	701[a]	1328[a]
ornithine transcarbamylase	7	428	3.27	—	—	—	1922[a]	2060	2402
P53	6	532	4.19	—	—	—	—	1332[a]	—
P65	6	593	∞[d]	978[a]	—	—	—	1024	—
proliferating cell nuclear antigen	7	263	7.79	1373	—	—	3290[b,c]	1841	3049[c]
phenylalanine hydroxlyase	4	461	12.91	673	—	—	—	—	—
phosphoenolpyruvate carboxykinase	8	676	2.09	831	1041	1157	1501	—	—
phosphofructokinase	6	997	2.12	1514[a]	1393[a]	1443	2337[a]	—	—
phosphoglycerate kinase	10	426	0.86	997	—	—	1333	—	—
propionyl-CoA carboxylase alpha	4	739	2.12	—	647	—		—	—
pyruvate dehydrogenase alpha	10	457	1.73	—	249[b,c]	—	280[b,c]	296[b,c]	260[b,c]
pyruvate dehydrogenase beta	7	376	1.75	—	1157	—	1215	1390	1178
pyruvate kinase	10	584	1.76	—	—	—	1441	1788	1965
ribonucleotide reductase large subunit	6	925	1.65	—	1383	—	1802	—	—
ribonucleotide reductase small subunit	10	446	1.33	734	1174	1121	1369	1473	1685
S-adenosyl-L-homocysteine hydrolase	8	498	1.03	751	970	916	1139	1804	1745[a]
thymidylate synthase	9	570	0.98	—	—	—	839[a]	785[a]	895[a]
transcription factor Eryf1	8	698	4.12	514	785[a]	812[a]	1240[a]	—	—
transglutinamase	7	800	3.02	712	—	—	—	—	—
triosephosphate isomerase	16	268	1.15	1160	1347	1205	1882	1478	1817
tryptophan hydroxylase	5	464	2.37	1395	—	—	—	—	—
urate oxidase	9	387	3.19	823	—	—	1078	968	1131
valine-tRNA ligase	5	1298	2.10	—	—	—	1077	—	—

[a] Rate-constancy was rejected for this comparison.
[b] In the analysis of all genes, this comparison was an outlier (upper or lower 5%) and therefore was excluded (see methods).
[c] In the analysis of constant-rate genes, this comparison was an outlier and therefore was excluded.
[d] A gamma parameter of ∞ indicates no rate variation among sites. For these genes, a Poisson correction was used.

Table 2. *Multigene divergence times (mean ± 1 s.e.m.) between animal phyla and among plants, animals and fungi*

	divergence time (Ma)[a]					
	gamma correction ($\alpha = X$)		gamma correction ($\alpha = 2$)		number of genes	
comparison	all genes	constant rate genes	all genes	constant rate genes	total	constant rate
chordate–arthropod	993 ± 46	944 ± 52	997 ± 45	940 ± 50	50	39
chordate–nematode	1134 ± 68	1078 ± 80	1122 ± 69	1062 ± 81	25	20
arthropod–nematode	1220 ± 89	1167 ± 83	1189 ± 87	1111 ± 71	18	15
animal–fungi	1538 ± 72	1532 ± 75	1537 ± 72	1513 ± 75	55	39
animal–plant	1547 ± 89	1529 ± 69	1567 ± 97	1492 ± 66	49	32
fungi–plant	1642 ± 104	1604 ± 94	1608 ± 114	1520 ± 91	38	26

[a] Divergence times using gene-specific shape parameters ($\alpha = X$) and a fixed shape parameter ($\alpha = 2$) are shown.

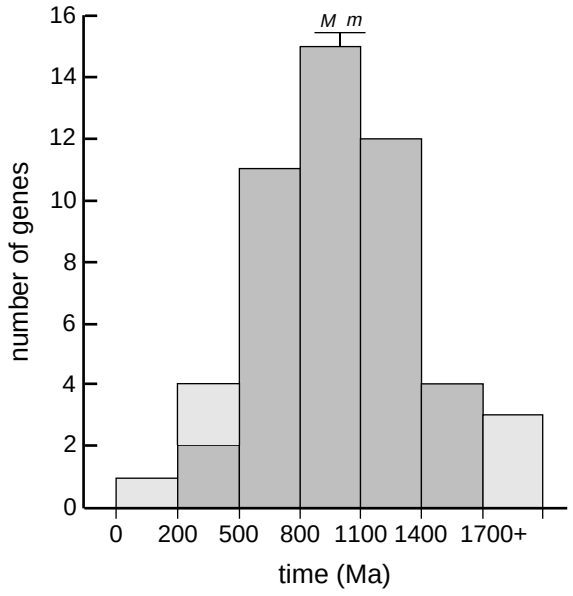

Figure 2. Distribution of divergence times between chordates and arthropods. Mean, *M*; median, *m*. Dark shading denotes time estimates used in analysis; light shading denotes outliers.

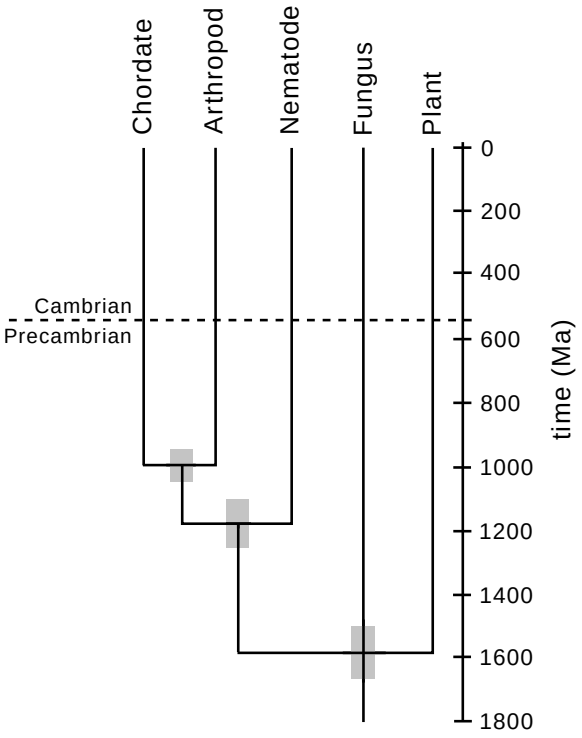

Figure 3. Divergence times and relationships of metazoan phyla and of plants, animals and fungi. Shaded bars denote 1 s.e.m.

The gene-specific gamma shape parameters ranged from 0.54 to infinity, but only nine (12%) were below 1.0 and the average value was 2.0 (excluding values > 5.0). A similar average value was obtained previously (Ota & Nei 1994) by analysis of an empirical matrix (Dayhoff *et al*. 1978), and in a study (Gu 1997) involving highly divergent protein sequence comparisons (prokaryotes and eukaryotes, metazoan phyla), suggesting some generality to this gamma shape parameter. For this reason, we also estimated evolutionary distances for each gene using a fixed shape parameter of 2.0. Average times estimates using the fixed shape parameter are similar (±5%) to estimates using gene-specific shape parameters (table 2).

A total of 50 genes could be used to estimate the chordate–arthropod divergence, and time estimates ranged from 148–5220 Ma (table 1). However, after the outliers were excluded, the time estimates ranged from 444–1605 Ma with a mean of 993 ± 46 Ma. After 11 genes failing the relative rate tests were excluded, the divergence time (944 ± 52 Ma) was within 5% of the time estimate from all genes. The same pattern was observed for the other comparisons here (table 2) and in a study of vertebrate sequences (Kumar & Hedges 1998), suggesting that rate variation present in some lineages and genes is not directionally biased. A similar pattern was found previously with fewer genes and more taxa (Wray *et al*. 1996). The single-gene time estimates for the chordate–arthropod comparison are normally distributed (figure 2) and the median (999 Ma) is close to the mean. With outliers excluded, the distribution is slightly (0.07), but not significantly, right-skewed.

The time estimates (all genes) using the pairwise-deletion method of estimating distances were about

Table 3. *Evolutionary relationships of three animal phyla (chordates, arthropods and nematodes) tested by four-cluster analysis*

(A gamma distribution with shape parameter=2 was used in all computations. Confidence probability (CP, in per cent) is the probability that the best tree is better than the given tree (Rzhetsky *et al*. 1995).)

	phylogenetic hypotheses				
	chordate +arthropod	nematode +arthropod		chordate +nematode	
gene name	CP	CP	$L_{nem.}/L_{art.}$ [a]	CP	$L_{nem.}/L_{chor.}$ [a]
DNA topoisomerase 2	best	>99%	1.60	>99%	1.68
phosphoenolpyruvate carboxykinase	best	92%	0.77	97%	1.12
nucleoside diphosphate kinase	best	87%	1.52	99%	1.42
calreticulin	best	89%	1.20	81%	1.26
S-adenosyl-L-homocysteine hydrolase	best	66%	1.02	93%	0.95
disulfide isomerase	best	>99%	1.32	53%	1.21
alpha actinin	best	31%	1.02	97%	1.08
ribonucleotide reductase small subunit	best	32%	1.07	97%	1.52
catalase	best	5%	0.95	88%	1.02
triosephosphate isomerase	best	5%	1.13	56%	1.02
phosphofructokinase	99%	best	0.87	97%	1.45
methionine adenosyl transferase	95%	best	0.75	99%	1.54
aldolase	57%	best	0.81	97%	0.84
lactate dehydrogenase	42%	best	0.71	93%	1.02
epoxide hydrolase	96%	88%	1.45	best	1.64
glyceraldehyde 3-phosphate dehydrogenase	69%	99%	1.09	best	1.02
transcription factor Eryf1	79%	18%	4.72	best	4.79
DNA topoisomerase 1	13%	77%	1.57	best	1.41
all genes combined	best	>99%		>99%	
rate-constant genes combined	best	>99%		>99%	

[a] Ratio of branch lengths (*L*). nem., nematode; art., arthropod; chor., chordate.

5–10% younger: arthropod–chordate, 942±49 Ma; chordate–nematode, 1036±65 Ma; arthropod–nematode, 1098±88 Ma; animal–fungi, 1408±71 Ma; animal–plant, 1407±85 Ma; fungi–plant, 1484±98 Ma. However, because that method may be biased by variation within large insertions, we have used time estimates from the complete-deletion method. The time estimates (all genes; $\alpha=X$, where X is the gene-specific shape parameter) for the six comparisons can be represented easily in a phylogenetic tree (figure 3). The chordate–arthropod divergence (993±46 Ma) is the most recent, and the average divergence time and average standard error between those two phyla and nematode is 1177±79 Ma.

We further examined the evolutionary relationships of arthropods, chordates and nematodes by using genes from table 1 for which the data contained at least one representative sequence from each of these groups and an outgroup. To explicitly test all three alternative hypotheses for the four groups of sequences, we employed four-cluster analysis, which is based on the minimum evolution principle (Rzhetsky *et al*. 1995). In this case, the best tree is the minimum evolution (ME) tree (tree requiring the least amount of evolutionary change). This ME tree is compared to the alternative trees and the confidence probability that the ME tree is better than the alternative tree is computed for each tree.

The results from four-cluster analysis for 18 genes are shown in table 3. In these results, the chordate +arthropod cluster appears as the best tree in ten genes compared with only four genes supporting nematode +arthropod and four genes supporting chordate +nematode. However, most genes do not show significant (confidence probability, CP>95%) support for any one particular phylogeny (exclusive of the other two phylogenies) for these three animal phyla. Such a mixed result was found with previous studies of single genes, where (for example) cytochrome *c* was found to support protostome polyphyly (Vanfleteren *et al*. 1994), and 18S rRNA was found to support protostome monophyly (Aguinaldo *et al*. 1997). When sequences from all 18 genes are combined (concatenated) into a single alignment of 8081 amino acids and analysed, the chordate+arthropod tree is significantly better (CP>99%) than either alternative hypothesis (table 3). The bootstrap confidence value for this relationship was >99% in a neighbour-joining analysis (gamma distance; $\alpha=2$; 2000 replications). The possibility that the basal position of the nematode is the result of an unusually long branch length, as suggested for 18S rRNA (Aguinaldo *et al*. 1997), was examined by calculating relative branch lengths (table 3). Although the nematode lineage is longer than the other two lineages in most gene comparisons, the ratios of branch lengths show that this difference is not great (*ca*. 10%). Also, combined analysis of only the 12 genes showing rate constancy in all three comparisons (4291 aligned amino acids) still results in significant support (CP>99%; bootstrap confidence value > 99% in neighbour-joining analysis) for the chordate–arthropod clade (table 3). This suggests that the

Table 4. *Evolutionary relationships of animals, plants, and fungi tested by four-cluster analysis*

	phylogenetic hypotheses		
gene name	animal + fungus	animal + plant	fungus + plant
argininosuccinate lyase	best	>99%	>99%
catalase	best	>99%	>99%
pyruvate kinase	best	94%	>99%
ornithine transcarbamylase	best	95%	87%
glucose 6-phosphate isomerase	best	99%	81%
heat shock protein (Hsp)70	best	81%	91%
methionine adenosyl transferase	best	85%	78%
pyruvate dehydrogenase beta	best	90%	69%
acetyl CoA acyltransferase	best	91%	61%
glyceraldehyde 3-phosphate dehydrogenase	best	82%	54%
inosine monophosphate dehydrogenase	best	21%	85%
nucleoside diphosphate kinase	best	20%	65%
DNA topoisomerase 1	best	62%	12%
pyruvate dehydrogenase alpha	best	62%	9%
S-adenosyl-L-homocysteine hydrolase	best	3%	55%
DNA topoisomerase 2	best	37%	19%
glutamine synthetase	best	14%	2%
ornithine aminotransferase	>99%	best	>99%
PCNA	>99%	best	>99%
ER60	98%	best	>99%
ribonucleotide reductase small subunit	83%	best	93%
urate oxidase	87%	best	77%
aspartate aminotransferase	90%	best	59%
enolase	47%	best	97%
triosephosphate isomerase	48%	best	84%
HMG-CoA reductase	41%	best	55%
amidophosphoribosyltransferase	22%	best	69%
aldehyde dehydrogenase	32%	best	58%
thymidylate synthase	95%	96%	best
asparagine synthase	91%	>99%	best
glycine hydroxymethyl transferase	86%	94%	best
hexokinase	84%	2%	best
glucose 6-phosphate dehydrogenase	12%	20%	best
fumarase	2%	16%	best
all genes combined	best	53%	99%
rate-constant genes combined	68%	best	99%

clustering of chordates and arthropods is not the result of rate bias in the nematode lineage.

Divergence times among plants, animals and fungi are quite similar (table 2). In fact, the divergence times for animal–plant and animal–fungus are almost identical (1547 and 1538 Ma, respectively). Attempts to resolve these relationships using four-cluster analysis also did not produce unequivocal support for a closer association of animals and fungi. An animal–fungi clade has been suggested in analysis of the 18S rRNA gene (Kumar & Rzhetsky 1996; Wainright *et al*. 1993). Even in a combined analysis of all 34 genes in which there was at least one plant, animal, fungus and outgroup sequence available (12 729 amino acids), there was no significant support for any of the three alternative phylogenies using four-cluster analysis (table 4) or bootstrap analysis of a neighbour-joining tree. This finding was unchanged in a combined analysis of the 20 rate-constant genes (6328 amino acids). Our results are similar to a maximum-likelihood analysis of a smaller data set (19 genes) in which the relationships of the three kingdoms could not be established statistically (Kuma *et al*. 1995). For these reasons, the divergence of plants, animals and fungi is not resolved in figure 3.

4. DISCUSSION

The results show that divergences among chordates, arthropods and nematodes average about 400 million years earlier than predicted by the fossil record (figures 1 and 3). Also, the divergence between the two protostomes (arthropod and nematode) is at least as old as the split between a protostome (arthropod) and a deuterostome (chordate). The three-way split of plants, animals and fungi, 1576 ± 88 Ma (figure 3), establishes the earliest time for divergences within animals. By inference, the primitive animal phyla (e.g. Porifera, Cnidaria, Ctenophora) diverged between that time and 1200 Ma. In an analysis of two genes (Nikoh *et al*. 1997), a more recent divergence time (940 Ma) was obtained for the divergence between

sponges (Porifera) and higher phyla (Eumetazoa). However, a protostome–deuterostome divergence of 700 Ma was used as a calibration point in that study; a 1300 Ma divergence between Porifera and Eumetazoa would be obtained with the protostome–deuterostome divergence time estimated here.

Several lines of evidence suggest that these new time estimates are more robust than earlier estimates. First, a much larger number of genes was used (figure 1), which will reduce the statistical error in the time estimate (Kumar & Hedges 1998). Second, the median (999 Ma) of the distribution of arthropod–chordate time estimates (figure 2) is similar to the mean. Although we used a gamma distance to account for rate variation among sites in calculating evolutionary distances for each gene (table 1), a minimal (Poisson) correction yielding a date of 906 ± 40 Ma for the arthropod–chordate divergence also rejects the hypothesis that this divergence occurred during the Cambrian explosion. Additional sequences will be needed to determine whether other metazoan phyla not examined here also arose at an earlier time than predicted by the fossil record.

The recognition that animal phyla diverged much earlier than when they appear in the fossil record has a direct impact on evolutionary models assuming a rapid phylogenetic diversification in the Cambrian (Gould 1989; Ohno 1996). The Cambrian explosion of fossils may simply record the appearance of hard parts in animals that already had a long prior history (Conway-Morris 1993). Besides being soft-bodied, the earliest animals are believed to have been very small or microscopic (Davidson *et al*. 1995; Runnegar 1982). This combination may explain the absence of fossil evidence for the early history of animals. Whether the increase in body size and acquisition of hard parts was directly related to the rise in atmospheric oxygen (Canfield & Teske 1996; Knoll 1992) has yet to be determined. An alternative argument is that macroscopic animals existed during this long gap in the Proterozoic. Support for that argument comes from 1100–1300 Ma trace fossils from Texas (Breyer *et al*. 1995) and India (Seilacher *et al*. 1998) thought to be the tunnels of triploblastic metazoans (worms). However, concerns have been raised as to the age of the Indian rocks and whether the organisms responsible for those burrows were unicellular or multicellular (Brasier 1998).

In many ways, this uncoupling of phylogenetic and phenotypic evolution (Fortey *et al*. 1997) is similar to the diversification of placental mammals, where fossil evidence records an explosive radiation in the early Tertiary following a much older phylogenetic history recorded by molecular clocks (Hedges *et al*. 1996; Kumar & Hedges 1998). The premise that these long gaps exist in the fossil record presents a challenge to palaeontologists, and the exquisite preservation potential of some late Proterozoic rock formations (Li *et al*. 1998; Xiao *et al*. 1998) provides a means of testing these hypotheses in the future.

We thank X. Gu and J. Zhang for their gz-aa program. S.K. was supported by a grant from NIH (GM20293) to M. Nei, and by funds from Arizona State University. S.B.H. was supported by grants from NSF (DEB9615643) and NASA (NCC2-1057).

REFERENCES

Aguinaldo, A. M., Turbeville, J. M., Linford, L. S., Rivera, M. C., Garey, J. R., Raff, R. A. & Lake, J. A. 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.

Ayala, F. J., Rzhetsky, A. & Ayala, F. J. 1998 Origin of the metazoan phyla: molecular clocks confirm paleontological estimates. *Proc. Natn. Acad. Sci. USA* **95**, 606–611.

Benton, M. J. 1993 *The fossil record 2*. London: Chapman & Hall.

Benton, M. J. 1997 *Vertebrate paleontology*. New York: Chapman & Hall.

Brasier, M. 1998 From deep time to late arrivals. *Nature* **395**, 547–548.

Breyer, J. A., Busbey, A. B., Hanson, R. E. & Roy III, E. C. 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.

Brown, R. H., Richardson, M., Boulter, D., Ramshaw, J. A. M. & Jeffries, R. P. S. 1972 The amino acid sequence of cytochrome *c* from *Helix aspera* Müeller (garden snail). *Biochem. J.* **128**, 971–974.

Canfield, D. E. & Teske, A. 1996 Late Proterozoic rise in atmospheric oxygen concentration inferred from phylogenetic and sulphur-isotope studies. *Nature* **382**, 127–132.

Conway-Morris, S. 1993 The fossil record and the early evolution of the Metazoa. *Nature* **361**, 219–225.

Davidson, E. H., Peterson, K. J. & Cameron, R. A. 1995 Origin of bilaterian body plans: evolution of developmental regulatory mechanisms. *Science* **270**, 1319–1325.

Dayhoff, M. O., Schwarts, R. M. & Orcutt, B. C. 1978 A model of evolutionary change in proteins. In *Atlas of protein sequence and structure* (ed. M. O. Dayhoff), pp. 345–352. Washington, DC: National Biomedical Research Foundation.

Feng, D.-F., Cho, G. & Doolittle, R. F. 1997 Determining divergence times with a protein clock: update and reevaluation. *Proc. Natn. Acad. Sci. USA* **94**, 13 028–13 033.

Fortey, R. A., Briggs, D. E. G. & Wills, M. A. 1997 The Cambrian evolutionary 'explosion' recalibrated. *BioEssays* **19**, 429–434.

Gould, S. J. 1989 *Wonderful life*. New York: W. W. Norton.

Gu, X. 1997 The age of the common ancestor of eukaryotes and prokaryotes: statistical inferences. *Molec. Biol. Evol.* **14**, 861–866.

Gu, X. 1998 Early metazoan divergence was about 830 million years ago. *J. Molec. Evol.* **47**, 369–371.

Gu, X. & Zhang, J. 1997 A simple method for estimating the parameter of substitution rate variation among sites. *Molec. Biol. Evol.* **14**, 1106–1113.

Hedges, S. B., Parker, P. H., Sibley, C. G. & Kumar, S. 1996 Continental breakup and the ordinal diversification of birds and mammals. *Nature* **381**, 226–229.

Higgins, D. G., Bleasby, A. J. & Fuchs, R. 1992 CLUSTAL V: improved software for multiple sequence alignment. *Comput. Appl. Biosci.* **8**, 189–191.

Knoll, A. H. 1992 The early evolution of eukaryotes: a geological perspective. *Science* **256**, 622–627.

Kuma, K., Nikoh, N., Iwabe, N. & Miyata, T. 1995 Phylogenetic position of *Dictostelium* inferred from multiple protein data sets. *J. Molec. Evol.* **41**, 238–246.

Kumar, S. 1996 *Phyltest: a program for testing phylogenetic hypotheses*. University Park, PA: Institute of Molecular Evolutionary Genetics, Pennsylvania State University.

Kumar, S. & Hedges, S. B. 1998 A molecular timescale for vertebrate evolution. *Nature* **392**, 917–920.

Kumar, S. & Rzhetsky, A. 1996 Evolutionary relationships of eukaryotic kingdoms. *J. Molec. Evol.* **42**, 183–193.

Kumar, S., Tamura, K. & Nei, M. 1993 *MEGA: molecular evolutionary genetic analysis*. University Park, PA: Pennsylvania State University.

Li, C.-W., Chen, J.-Y. & Hua, T.-E. 1998 Precambrian sponges with cellular structures. *Science* **279**, 879–882.

Nikoh, N. (and 10 others) 1997 An estimate of divergence time of Parazoa and Eumetazoa and that of Cephalochordata and Vertebrata by aldolase and triose phosphate isomerase clocks. *J. Molec. Evol.* **45**, 97–106.

Ohno, S. 1996 The notion of the Cambrian pananimalia genome. *Proc. Natn. Acad. Sci. USA* **93**, 8475–8478.

Ota, T. & Nei, M. 1994 Estimation of the number of amino acid substitutions per site when the substitution rate varies among sites. *J. Molec. Evol.* **38**, 642–643.

Runnegar, B. 1982 A molecular-clock date for the origin of the animal phyla. *Lethaia* **15**, 199–205.

Runnegar, B. 1986 Molecular palaeontology. *Palaeontology* **29**, 1–24.

Rzhetsky, A., Kumar, S. & Nei, M. 1995 Four-cluster analysis: a simple method to test phylogenetic hypotheses. *Molec. Biol. Evol.* **12**, 163–167.

Seilacher, A., Bose, P. K. & Pfluger, F. 1998 Triploblastic animals more than 1 billion years ago: trace fossil evidence from India. *Science* **282**, 80–83.

Takezaki, N., Rzhetsky, A. & Nei, M. 1995 Phylogenetic test of the molecular clock and linearized tree. *Molec. Biol. Evol.* **12**, 823–833.

Vanfleteren, J. R., Peer, Y. V. d., Blaxter, M. L., Tweedie, S. A. R., Trotman, C., Liu, L., Hauwaert, M.-L. V. & Moens, L. 1994 Molecular genealogy of some nematode taxa as based on cytochrome *c* and globin amino acid sequences. *Molec. Phyl. Evol.* **3**, 92–101.

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

Wray, G. A., Levinton, J. S. & Shapiro, L. H. 1996 Molecular evidence for deep Precambrian divergences. *Science* **274**, 568–573.

Xiao, S., Zhang, Y. & Knoll, A. H. 1998 Three dimensional preservation of algae and animal embryos in a Neoproterozoic phosphorite. *Nature* **391**, 553–558.

As this paper exceeds the maximum length normally permitted, the authors have agreed to contribute to production costs.