

Determining divergence times with a protein clock: Update and reevaluation

(animals/plants/fungi/eukaryotes/eubacteria/archaeobacteria)

DA-FEI FENG, GLEN CHO, AND RUSSELL F. DOOLITTLE*

Center for Molecular Genetics, University of California at San Diego, La Jolla, CA 92093-0634

Contributed by Russell F. Doolittle, September 23, 1997

ABSTRACT A recent study of the divergence times of the major groups of organisms as gauged by amino acid sequence comparison has been expanded and the data have been reanalyzed with a distance measure that corrects for both constraints on amino acid interchange and variation in substitution rate at different sites. Beyond that, the availability of complete genome sequences for several eubacteria and an archaeobacterium has had a great impact on the interpretation of certain aspects of the data. Thus, the majority of the archaeobacterial sequences are not consistent with currently accepted views of the Tree of Life which cluster the archaeobacteria with eukaryotes. Instead, they are either outliers or mixed in with eubacterial orthologs. The simplest resolution of the problem is to postulate that many of these sequences were carried into eukaryotes by early eubacterial endosymbionts about 2 billion years ago, only very shortly after or even coincident with the divergence of eukaryotes and archaeobacteria. The strong resemblances of these same enzymes among the major eubacterial groups suggest that the cyanobacteria and Gram-positive and Gram-negative eubacteria also diverged at about this same time, whereas the much greater differences between archaeobacterial and eubacterial sequences indicate these two groups may have diverged between 3 and 4 billion years ago.

In theory, past biological events can be reconstructed on the basis of present-day macromolecular sequences. Certainly, the assignment of organisms to various biological groups on the basis of ribosomal or protein sequences has been largely successful. Attaching absolute time scales to phylogenetic trees has proved more troublesome, however. As a case in point, we recently reported the results of an analysis of 531 amino acid sequences from 57 different sets of enzymes drawn from 15 different biological groups (1). We had aligned the 57 sets of sequences, determined how similar they were from group to group, and calculated evolutionary distances based on those similarities. The distance data were calibrated on the basis of divergence times drawn from the fossil record, and extrapolations were made to estimate the divergence times of more distantly related groups. The data were also used to compute an overall phylogeny for these same groups. The most provocative finding was that the divergence time between eukaryotes and eubacteria was, after various corrections, only slightly more than 2 billion years ago.

The results were harshly criticized by others on a number of counts. Some felt that it was unreasonable to extrapolate so far backwards in time on the basis of, mostly, the vertebrate fossil record (G. Olsen, quoted in ref. 2). Others felt that the distance value calculations did not take sufficient account of variations

in the rate of change at different amino acid locations (3–5). Concern was also expressed that the data were corrupted by the presence of sequences imported during the endosymbiotic acquisition of organelles (4). A pervading thought in all the criticism was that the divergence time between eukaryotes and eubacteria had to be greater than 3.5 billion years because of the occurrence of microfossils that have been undisputedly dated to 3.45 billion years ago and that reportedly resemble modern cyanobacteria (6).

In a recent response to these criticisms, we showed that even with the use of rigorous methods for correcting for site variation, the data yielded divergence times for eubacteria and eukaryotes of 2.5 billion years or less, if it was assumed that the plant–animal divergence was 1 billion years (7). In support of this finding, Adkins and Li (8), using a subset of the same data and an algorithm that takes account of site variation, found a value of 2.2 billion years, again presuming the plant–animal divergence was 1 billion years ago.

We have now completed a comprehensive updating of the enzyme sequence data set, the emphasis being on increasing the number of sequences from the more under-represented biological groups, including those used in the fossil record-based calibration. At the same time we employed a method of calculation that rigorously corrects for site variation (9). Indeed, the additional data and the improved methods of calculation have led to some significant changes. The greatest impact was the result of having access to the total genomic sequences of several eubacteria and particularly an archaeobacterium (10). In our initial effort (1), archaeobacteria were represented in only 9 of the enzyme sets, and at the time we noted there were anomalies among them. In the updated data set archaeobacteria are represented in 34 of the 64 enzyme sets. Remarkably, the majority of phylogenies calculated with these newly available sequences are not consistent with current notions about the Tree of Life in that the archaeobacterial sequences are either outliers or mixed in with eubacterial orthologs. Clearly, the matter of finding divergence times for eukaryotes, archaeobacteria, and eubacteria is dependent on the assumed branching order of these groups.

We have addressed the problem by considering the data in various sets and subsets, assignment being strictly dependent on the phylogenetic trees generated for each enzyme. In this regard, the enzyme sequence sets were categorized according to the presence or absence of an archaeobacterial representative, and, if present, on its position in the sequence-based phylogeny. These groups were employed judiciously according to the particular task at hand. Thus, all 64 sets were used for determining divergence times of the various groups of eukaryotes, but only certain sets were useful for determining the divergence times of eukaryotes and archaeobacteria.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

© 1997 by The National Academy of Sciences 0027-8424/97/9413028-6\$2.00/0
PNAS is available online at <http://www.pnas.org>.

Abbreviations: My, millions of years; Mya, millions of years ago.

A commentary on this article begins on page 12751.

*To whom reprint requests should be addressed at: Center for Molecular Genetics, Room 206, University of California at San Diego, La Jolla, CA 92093-0634. e-mail: rdoolittle@ucsd.edu.

The following three assumptions formed the framework for the analysis. First, it was assumed that the standard model for the Tree of Life, which has eukaryotes as a sister group to or emerging from the archaeobacteria, is correct (11–17). Second, it was assumed that a major importation event involving the transfer of genes from a eubacterium into the eukaryotic lineage occurred either shortly after or as part of the invention of eukaryotes (14, 18). Third, it was accepted that genes have been exchanged between eubacteria and archaeobacteria on numerous occasions (19, 20). Given these conditions, all the data could be applied in a proper setting.

Thus, the divergence times of the principal groups of eukaryotes were determined with all 64 enzymes sets (481 sequences from eukaryotes) without regard for the divergence of eukaryotes from archaeobacteria or eubacteria. These divergence times were obtained both by simple extrapolation, on the one hand, and by construction of phylogenetic trees, on the other, the latter having the advantage of accommodating different rates of change for different lineages. In this regard, the full set of eubacterial sequences was used as an outlier for purposes of rooting the eukaryote tree, but not for determining the eukaryote–prokaryote divergence time.

Similarly, the divergence time of eubacteria and archaeobacteria was determined without regard to eukaryotic sequences, except for the presumption that the sequences were changing at the same rate as their eukaryotic homologs. Additionally, those sets of sequences that showed clear signs of intergroup transfer between archaeobacteria and eubacteria were omitted from the calculation.

The divergence time of eukaryotes from archaeobacteria was determined from that group of enzymes that exhibited phylogenies reflecting the standard model of the Tree of Life. In contrast, the timing of the alleged endosymbiotic event that transferred eubacterial genes to eukaryotes was determined with the group of enzymes in which the archaeobacterial representatives were the outliers.

METHODS

Sequences. The sequence database used in this study is an expanded form of the 57-enzyme data set used in our earlier work (1). Seven additional enzyme sets were added that met the minimal criteria of being represented by animals, plants or fungi, and eubacteria, and many more sequences were added to the original set; the total set comprised 823 sequences. The 64 sets of sequences were each aligned by the progressive method (21) in conjunction with the BLOSUM-62 matrix (22);

phylogenetic trees were constructed as previously described (21). After alignment, the 64-enzyme set spanned a total of approximately 25,000 amino acid residue positions.

The 34 enzyme sets that had archaeobacterial representation were sorted into three classes dependent on the phylogenetic position of the archaeobacterial entry. Group A included those enzymes in which the archaeobacterial contribution appeared orthodox in that they were positioned nearer the eukaryotes than the eubacteria (Table 1). Group B was made up of those enzyme sets in which the archaeobacterial representatives were the outliers. Group C was composed of those sequences that had archaeobacterial representatives mixed in with the eubacteria. In four cases the sequences from halobacteria and methanogens, both archaeobacteria, occurred at markedly different positions in the trees. In three cases, the methanogen was the outlier and the halobacterium was mixed in with eubacteria. In another case, the methanogen was close to the eukaryotes but the halobacterium was the outlier. The four anomalies increased the number of comparisons involving archaeobacteria to 38.

Calculation of Evolutionary Distances. In our first paper on this subject, we used a simple Poisson relationship in conjunction with similarity measures to calculate the evolutionary distance (23). Although this treatment corrected for a number of factors, including the nature of amino acid distributions and the likelihood of individual changes, it did not correct for site-to-site variation in the rate of amino acid replacement. We were aware of this shortcoming and had used an after-the-fact correction for the longest distances.

Recently, however, Grishin (9) published an elegant analysis of distance–time relationships and described several equations for correcting various problematic factors, including one that corrects both for the nature of amino acid interchange and for site-to-site variations in rate. In the present study all distances were calculated according to Grishin's formula:

$$q = \frac{\ln(1 + 2D)}{2D},$$

where q is the fraction of unchanged residues and D is the evolutionary distance. We have tested the Grishin relationship in a simulation exercise (24) and found that the calculated distance varies proportionately with the number of evolutionary hits. If anything, the relationship tends to overcorrect; as a result, distributions of distances tend to have long tails on the high side representing the least similar sequence pairs. Accordingly, the longest divergence times were calculated from

Table 1. Assignment of enzymes sets to groups depending on archaeobacterial phylogeny

Group A	Group B	Group C
Hydroxymethylglutaryl-CoA reductase	Lactate dehydrogenase	IMP dehydrogenase
DNA-dependent RNA polymerase	Glyceraldehyde-3-phosphate dehydrogenase B*	Glyceraldehyde-3-phosphate dehydrogenase B†
Enolase*	Glycine hydroxymethyltransferase	Dihydrofolate reductase
Isoleucyl-tRNA synthetase	Pyruvate kinase	Ornithine transcarbamoylase
Argininosuccinate synthase	Phosphoglycerate kinase	Dimethylallylamine transferase
Isocitrate dehydrogenase	Dihydroliipoamide dehydrogenase*	Dihydroliipoamide dehydrogenase†
Aspartate transcarbamoylase	Ribose-phosphate pyrophosphokinase	Nucleoside diphosphate kinase
Histidyl-tRNA synthetase	Argininosuccinate lyase	Porphobilinogen synthase
	UDP-glucose epimerase	DNA topoisomerase
	Triose-phosphate isomerase	Seryl-tRNA synthetase†
	Enolase†	Glutamate–ammonia ligase
	Valyl-tRNA synthetase	Carbamoyl-phosphate synthase N
	Threonyl-tRNA synthetase	Carbamoyl-phosphate synthase C
	Adenylosuccinate synthase	
	Aldehyde dehydrogenase	
	Seryl-tRNA synthetase*	
	Dihydroorotate oxidase	

*Methanogen.

†Halobacterium.

Table 2. Average sequence resemblances, distances, and divergence times of vertebrate groups

Comparison	<i>n</i> *	% identical	<i>D</i> †	LCA,‡ Mya
Mammal/mammal	48	91	11.2	100
Eutheria/marsupial	3	92	9.5	130
Mammal/bird-reptile	16	82	24.6	300
Amniote/amphibian	11	80	27.9	365
Tetrapod/fish	15	75	37.1	405
Gnathostome/lamprey	1	76	35.5	450

*Number of enzyme sets compared.

†Distances calculated from $q = [\ln(1 + 2D)]/2D$.

‡LCA, last common ancestor, given as millions of years ago, based on fossil record.

both the mean and median distances. Although for purposes of illustration we frequently use average resemblances, all distances were calculated for the individual *q* values and then averaged.

Calibration and Extrapolation. Six reasonably well established divergence times were used to calibrate the system, all based on the vertebrate fossil record (Table 2). The number of enzyme sets available for each point varied, ranging from 45 for the mammalian radiation to a lone entry for the agnatha vs. other vertebrates. In the current calibration we omitted a point based on two echinoderm sequences that had been used in our earlier study (1), determining that divergence by extrapolation instead.

Because of the importance of the calibration, we examined the data in several ways. In each case, however, distances were calculated individually for each of the enzyme sets and then averaged before plotting against the divergence time based on the fossil record (Fig. 1). The slopes were then used for the direct calculation of the greater divergence times from distances. All plots were constrained to pass through the origin.

Initially, the six calibration points were used as simple averages, without regard for the number of enzymes represented at each time point; in this case the slope was 0.0891 *D*/Mya. In a second set of analyses, the points were weighted according to their representation (i.e., the better-represented points counted more), and the slope was 0.0876. Finally, an "omit test" was performed, in which, one at a time each of the time points was omitted from the calculation. None of the omissions resulted in more than a 5% change in slope. In the end we used a slope of 0.088 for all extrapolations.

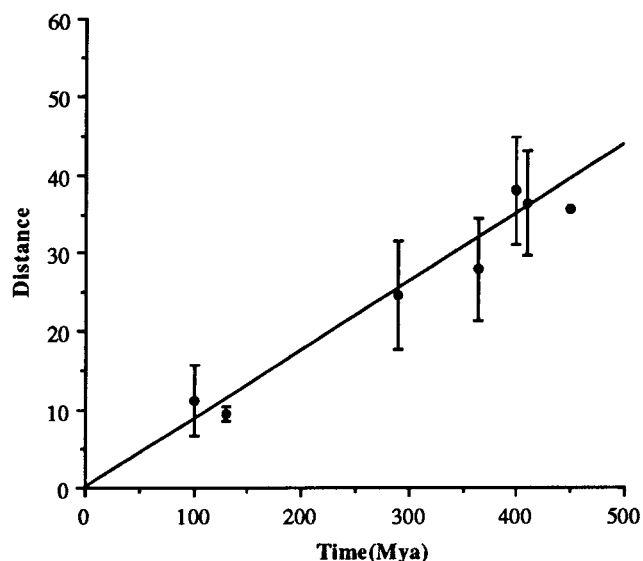


FIG. 1. Evolutionary distances calculated according to Grishin (9) plotted against divergence times from the vertebrate fossil record. Error bars are standard deviations.

RESULTS

Eukaryotic Divergences. Eukaryotic divergences were first determined by direct extrapolation using the calculated slope (Table 3). In the case of echinoderms, only two of the enzymes were represented, and the divergence time of 590 My is based on slopes calculated from these sequences alone. The deuterostome-protostome divergence, which separates the vertebrate-echinoderm lineage from other animals, was measured at 850 Mya, and the schizocoelome-pseudocoelome divergence, which separates early emerging groups such as nematodes from arthropods, molluscs, and others, is 1050 Mya. Phylogenetic analysis, which adjusts for relative rates of change, shortened the latter two times to 730 and 815 My, respectively (Table 3).

Also by direct extrapolation, plants and animals last shared a common ancestor about 1,200 Mya, and fungi diverged from either of those groups at about 1,275 Mya. If we assume the protist lineage was monophyletic, then extrapolation puts the protist divergence from other eukaryotes at about 1,550 Mya. The bulk of the 38 protist sequences used in this study were

Table 3. Divergence times from extrapolation and phylogeny

Group	<i>n</i> *	% identical	<i>D</i> †	TimeX,‡ My	TimeP,§ Mya
Echinoderm-chordate	2	69	52	590	—
Deuterostome-protostome	24	64	75	850	730
Schizocoelome-pseudocoelome	15	61	92	1045	815
Fungi-animal	57	54	112	1272	1130
Plant-animal	37	55	107	1215	1200
Protist-plant-animal-fungi	22	51	136	1545	1540
Archaeobacteria-eukaryotes¶	8	42	212	2409	—
			(199)	(2261)	—
Eubacteria-eukaryotes	17	42	193	2188	—
			(183)	(2080)	—
Archaeobacteria-eubacteria**	25	33	333	3784	—
			(275)	(3125)	—

*Number of intergroup comparisons.

†See text for method of calculating distances (*D*). Values in parentheses are from medians.

‡TimeX = extrapolation in millions of years. Values in parentheses are from medians.

§TimeP from phylogenetic analysis with plant-animal = 1200 Mya.

¶Data from group A, Table 1, only.

||Data from group B, Table 1, only.

**From groups A and B, Table 1.

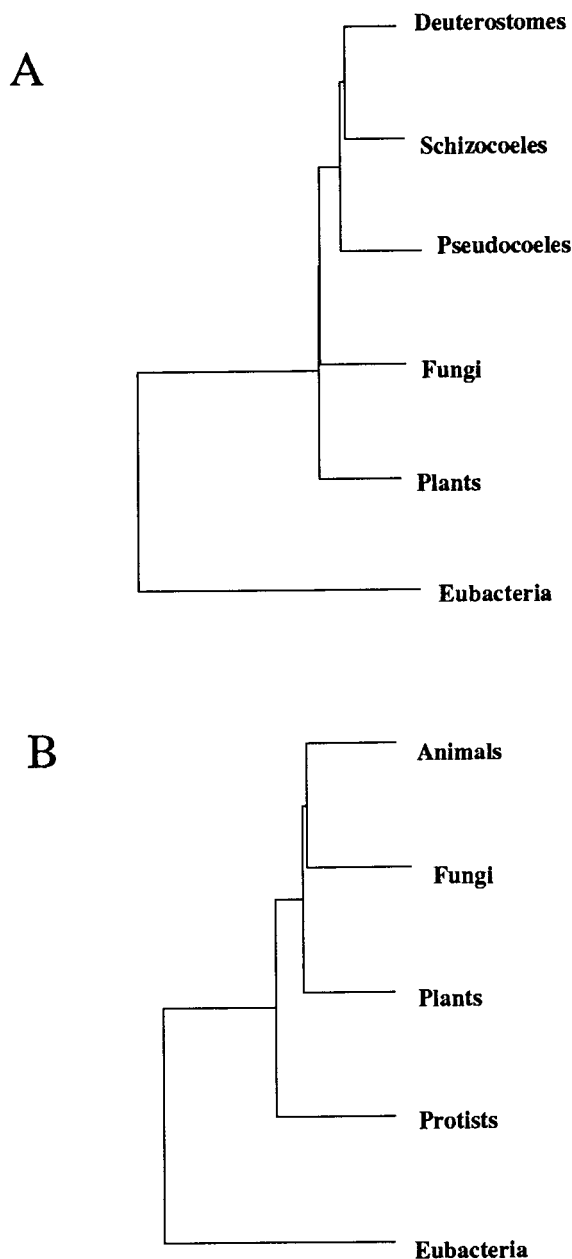


FIG. 2. Phylogenetic trees of major groups of eukaryotes based on mean intergroup distances; the eubacteria were used as outlier for obtaining the topology of other groups. (A) Based on 11 fully represented enzyme sets. (B) Based on average intergroup distances for all enzyme sets. In both cases the plant divergence was taken to be 1200 Mya.

from three genera: *Plasmodia* ($n = 10$), *Trypanosoma* ($n = 8$), and *Leishmania* ($n = 6$).

A phylogenetic tree was constructed from the raw distance data used in the direct extrapolation, eubacterial sequences being used as an outlier group to establish a proper branching order. As reported by others (25–27) and in our previous study (1), the phylogenetic analysis shows animals and fungi to be sister groups (Fig. 2). If the extrapolated plant-animal time of 1,200 Mya is used to set the clock (Fig. 2B), then the phylogenetically determined time for the fungus–animal divergence is 1,130 Mya, the difference between this value and the 1,275 measured by direct extrapolation reflecting an increased rate of amino acid replacement along the fungal lineage relative to plants and animals. The protist divergence time determined from the phylogeny was approximately 1,540 Mya, in accord with the extrapolated value (Table 3).

Table 4. Resemblances between major groups of organisms categorized by branching order*

Group	%		
	identical	D^\dagger	Time, ‡ Mya
Archaeobacteria–Eukaryotes			
All Pairs ($n = 38$)	34.2	319 (275)	3628 (3131)
Group A ($n = 8$)	40.9	213 (190)	2424 (2261)
Group B ($n = 17$)	30.7	396 (307)	4498 (3489)
Group C ($n = 13$)	36.2	285 (223)	3234 (2534)
Eubacteria–Eukaryotes			
All Pairs ($n = 34$)	37.4	259 (223)	2947 (2534)
Group A ($n = 8$)	32.1	352 (298)	3994 (3386)
Group B ($n = 14$)	42.1	193 (183)	2188 (2080)
Group C ($n = 12$)	35.6	274 (223)	3110 (2534)
Archaeobacteria–Eubacteria			
All pairs ($n = 38$)	37.8	273 (247)	3105 (2807)
Group A ($n = 8$)	36.1	290 (261)	3292 (2966)
Group B ($n = 17$)	31.6	353 (290)	4015 (3295)
Group C ($n = 13$)	46.8	158 (120)	1796 (1364)

*See Table 1 for listing of group members. Boldface corresponds to sets actually used for determining divergence times.

† See text for method of distance calculation. Values in parentheses are medians.

‡ Divergence times. Values in parentheses are from medians.

Archaeobacteria–Eukaryotes. If current views of the Tree of Life are correct, then only eight of the enzyme sets (group A) could be used for determining the last common ancestor of eukaryotes and archaeobacteria. On the average, these eight sets are 42% identical (i.e., $q = 0.42$) for eukaryotes and archaeobacteria, and the extrapolated divergence time is 2,100 My (Table 3). In those other cases where the archaeobacterial sequences were the outliers, the low resemblance between eukaryote and archaeobacterial sequences (31% identical, Table 4) reflects the anomaly attributable to the postulated import of eubacterial sequences into eukaryotes.

Eubacterial Imports. There were 17 enzyme sets in which at least one archaeobacterial representative was the outlier (group B). We have interpreted this arrangement as being due to the importation of eubacterial sequences into eukaryotes, most likely as a part of an endosymbiotic capture. Such a short-circuit would move eubacterial and eukaryote sequences closer together (Fig. 3). The resemblances of archaeobacterial and eubacterial sequences should be unaffected by the import event, however.

The distance between the eukaryote and eubacterial sequences in group B indicates that these sequences are about as similar as are the archaeobacterial and eukaryote sequences in group A, implying that the acquisition by eukaryotes of these eubacterial sequences occurred at about the same time as the divergence of archaeobacteria and eukaryotes (Fig. 3). Similarly, the average distances between eubacterial and archaeobacterial sequences in groups A and B, as determined from the 8 enzymes sets in group A and the 17 sets in group B, were close to each other (36% and 32% identical, respectively; Table 4). The extrapolated divergence time for these levels of resemblance would be between 3,100 and 3,800 Mya (Table 3).

Divergence of Eubacterial Groups. Taken all together, the sequences from cyanobacteria are slightly more similar to those from Gram-positive bacteria than they are to those from Gram-negative bacteria, although some of the Gram-positive sequences, especially those from group A, were more similar to Gram-negative ones. The average degrees of resemblance (45% identity) equate to a divergence time of about 2,000 Mya, if we presume the same rate of change as exhibited by eukaryotes. On the other hand, ribosomal RNA sequences have pinpointed the endosymbiotic origins of mitochondria to the rickettsia branch of purple bacteria, and

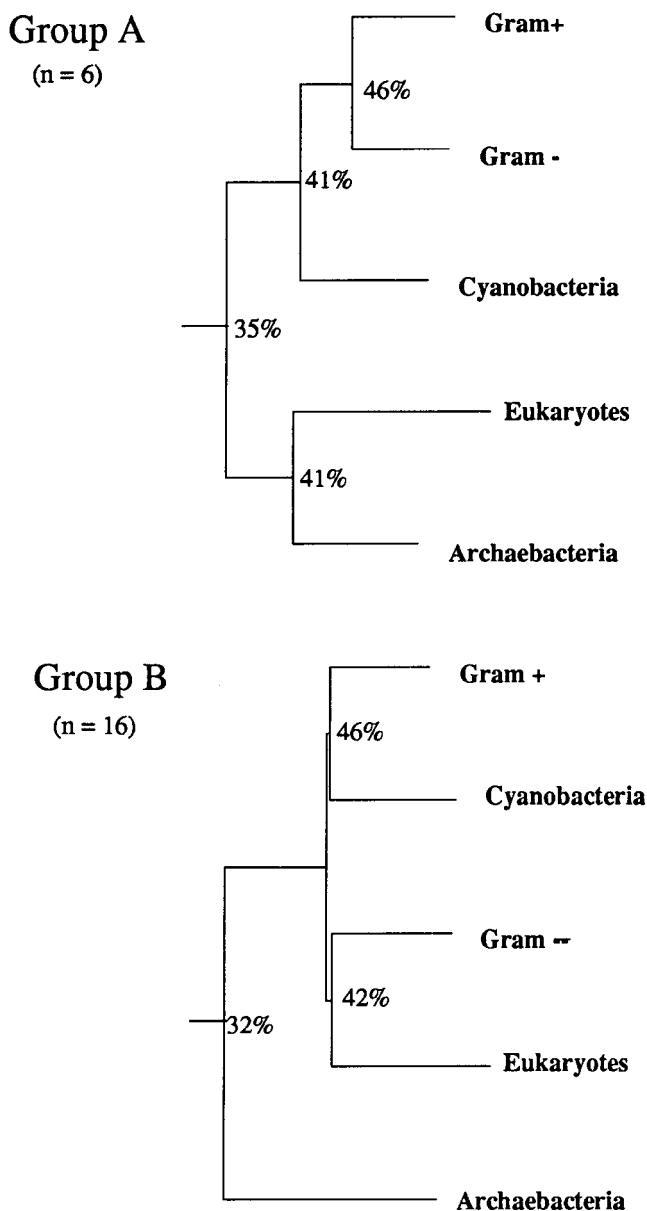


FIG. 3. (A) Phylogeny of eukaryotes, archaeobacteria, and principal eubacterial groups as determined with six of the eight enzyme sets in which the archaeobacteria are nearest neighbors with eukaryotes (group A in Table 1; two of the enzyme sets were not included because sequences were not available for all bacterial groups). (B) Phylogeny of the same groups determined with 16 of the 17 enzymes sets in which the archaeobacteria are outliers (group B in Table 1; in the case of lactate dehydrogenase, sequences were not available for all bacterial groups). Average sequence resemblances (% identical) are shown at key divergence points.

these divisions must have preceded that occurrence. One explanation is that the eubacterial lineages have been changing more slowly than the eukaryotes. The large differences between eubacterial and archaeobacterial sequences puts a limit on how much slower that rate could be. Alternatively, the similarity between the major groups of eubacteria may have been enriched by horizontal transfers between their lineages during the time since they diverged. In either case, the most reasonable interpretation is that the divergences leading to the cyanobacteria, purple bacteria (Gram-negative), and Gram-positive lineages occurred only slightly before the endosymbiotic event, which direct extrapolation puts at about 2,200 Mya (Table 3).

DISCUSSION

Eukaryotic Divergences. Of all the measurements reported in our earlier paper (1), the divergences of the various eukaryote groups should have been the least controversial. Not only are the extrapolations the shortest, but the average sequence resemblances between the groups are all greater than 50% identical, a level of similarity not much affected by site-to-site variation. So the question arises, why do our results differ from those of other workers who have attempted similar extrapolations recently? In particular, Wray *et al.* (28) undertook to find the divergence times of principal animal groups based on a blend of amino acid sequences from four mitochondrial-encoded proteins (ATPase 6, cytochrome oxidases I and II, and NADH dehydrogenase subunit 1), three nuclear-encoded proteins (α - and β -globins and cytochrome *c*), and ribosomal RNA. Their divergence time of about a billion years for echinoderms and chordates is considerably longer ago than our time of 590 My, and significantly longer ago than our divergence times for other animal divergences that must have antedated the echinoderm–chordate split. In our view, the difference is mainly attributable to the choice of proteins by Wray and Levinton (28). Globins, for example, are small, fast-changing proteins with limited usefulness for determining early divergences. Beyond that, the increased rate of change along the invertebrate lineages tends to inflate divergence times found by extrapolation.

Recently, Nikoh *et al.* (29) measured the divergence times of some animal groups by using two protein sequences (fructose-bisphosphate aldolase and triose-phosphate isomerase). On the average, they found that the protochordate amphioxus diverged from other chordates about 700 Mya, longer ago than our divergence time for echinoderms, a group whose divergence must certainly have occurred earlier. Their value could be regarded as support for the conclusions of Wray and Levinton (28), but we would emphasize that our divergence times (Table 3) for deuterostomes from protostomes, on the one hand, and schizocoelomes and pseudocoelomes, on the other, are based on 21 and 9 sets of sequences, respectively, and involve many more amino acid positions than used by either of these groups (28, 29).

Archaeobacteria–Eukaryotes. In our previous report (1) we did not categorize the various enzyme groups with regard to their individual phylogenies. Instead, all archaeobacterial distances were averaged, as were all eubacterial distances. If the current standard model of the Tree of Life is accepted, and this is an important if, then only eight of the enzyme sets in our database meet the phylogenetic condition for determining the last common ancestor of eukaryotes and archaeobacteria (Table 1). A number of other proteins not included in our database have been reported to exhibit the “standard phylogeny,” however, including elongation factors (15, 16) and RecA protein (17). Also, a survey of aminoacyl-tRNA synthetase sequences has shown that 7 of 11 of those enzymes not included in our 64-enzyme database also support this phylogeny (R.F.D. and J. Handy, unpublished results). In any event, the average resemblance for the eight archaeobacterial and eukaryote enzymes in this group amounts to 41% identity and an extrapolated divergence time of about 2,300 Mya (Table 4).

Importation. In our last report (1), while allowing that one or two anomalous sequences reflecting organellar import might have crept into our collection, we doubted that the numbers could be significant. This position was based mainly on the fact that the distribution of eukaryote–eubacteria distances was quite uniform, and that analyses in which random sets of sequence were omitted did not affect the outcome. It never occurred to us that the *majority* of these comparisons might involve imports, as now seems to be the case, although we did specifically comment that in the event the eukaryotic cell was a chimera of a eubacterium and an

archaeobacterium, as has been proposed by some (30–33), then the time we had measured would reflect that event.

In the interim, several reports have appeared that suggest the importation of eubacterial enzymes into the eukaryotic lineage was indeed a part of the endosymbiotic event leading to mitochondria (14, 18), and that this must have occurred very early in the evolution of eukaryotes, doubt now being cast on the notion that certain extant mitochondrial eukaryotes diverged before that event (34, 35).

We have now identified 17 enzymes for which phylogenetic arrangements show the eubacteria and eukaryotes to be sister groups. Many of these enzymes are associated with major metabolic pathways such as glycolysis. As it happens, the sequence resemblances between eukaryotes and eubacteria for these enzymes are about the same as the resemblances observed between eukaryotes and archaeobacteria for the group A enzymes (42% and 41% identity, respectively), suggesting that the alleged importation event took place at about the same time or very shortly after the divergence of eukaryotes and archaeobacteria (Fig. 3). Further support for this interpretation is offered by the fact that the eubacterial and archaeobacterial sequences from both groups A and B are similarly distant (36% and 32% identical, respectively). In contrast, the sequences of Gram-positive bacteria, purple bacteria, and cyanobacteria are considerably more similar (45% identical on the average) and must reflect a much more recent divergence.

Horizontal Exchanges. With regard to the anomalously mixed phylogenetic positionings of archaeobacteria and eubacteria (group C), most appear to have the archaeobacteria acquiring genes from eubacteria. In 7 of the 13 cases the data clearly show that an archaeobacterial sequence is significantly more similar to some eubacterial sequences than are other eubacterial orthologs. For example, in the case of dihydrolipoamide dehydrogenase, the halobacterium entry is 50% identical with three Gram-positive bacterial sequences, whereas three orthologous sequences from Gram-negative bacteria are only 43% identical to any of the four others.

Additional support for the idea that exchanges have occurred between eubacteria and archaeobacteria was afforded by the observation that in four of the enzyme sets the halobacterial sequences differed radically in their phylogenetic positions compared with those from methanogens. A similar mixed situation for these two archaeobacteria has been reported for certain heat shock proteins (36).

In summary, protein sequence comparisons strongly indicate that fungi and animals diverged only a little more than 1 billion years ago, and plants diverged from the animal–fungus common ancestor only a little before that. Comparable protist sequences are, on the average, more than 50% identical with those from later-diverging eukaryotes and should have diverged about 1,450 Mya.

Just over 2,000 Mya, a conjunction of events gave rise to eukaryotes, as reflected in protein sequences yielding two different phylogenetic arrangements. First, the divergence from archaeobacteria is chronicled in a set of eight enzymes averaging 41% identity with eukaryotes. Second, the acquisition by eukaryotes of a set of 17 enzymes from eubacteria is supported by phylogenetic analysis and, coincidentally, the same degree of resemblance (42% identity).

Comparison of sequences from 25 appropriate enzyme sets with eubacterial and archaeobacterial representation shows them to average 33% identity, a level of resemblance corresponding to a divergence time of 3,200–3,800 Mya, presuming a constant rate of change throughout. This would be the time of the common ancestor of all extant organisms, a much longer time than we postulated in our earlier article (1). On the other hand, analysis of an equivalent set of enzyme sequences shows that the sequences of cyanobacteria and Gram-positive and

Gram-negative bacteria are much more similar (45% identical) to each other than they are to archaeobacterial orthologs, these divergences likely occurring 2,100–2,500 Mya, about the same time as the divergence of eukaryotes from archaeobacteria and the eukaryotic acquisition of a host of eubacterial enzymes. These observations are not incompatible with reports of the oldest eukaryotes in the microfossil record (37), and may even accommodate a 2,100 My-old megascopic fossil (38). They cast serious doubt, however, on whether 3,450 My-old microfossils (6) truly represent modern cyanobacteria.

We are grateful to W. Ford Doolittle and his colleagues for comments on the manuscript. This work was supported in part by a National Aeronautics and Space Administration Specialized Center for Research and Training Grant in Exobiology at the University of California, San Diego.

1. Doolittle, R. F., Feng, D.-F., Tsang, S., Cho, G. & Little, E. (1996) *Science* **271**, 470–477.
2. Morrell, V. (1996) *Science* **271**, 448.
3. Hasegawa, M. & Fitch, W. M. (1996) *Science* **274**, 1750.
4. Gogarten, J. P., Olenzinski, L., Hilario, E., Simon, C. & Holsinger, K. E. (1996) *Science* **274**, 1750–1751.
5. Miyamoto, M. & Fitch, W. M. (1996) *Syst. Biol.* **45**, 566–573.
6. Schopf, J. W. (1993) *Science* **260**, 640–646.
7. Doolittle, R. F., Feng, D.-F., Tsang, S., Cho, G. & Little, E. (1996) *Science* **274**, 1751–1753.
8. Adkins, R. M. & Li, W.-H. (1997) in *Horizontal Gene Transfers*, eds. Syvanen, M. & Kado, C. (Chapman-Hall, London), in press.
9. Grishin, N. V. (1995) *J. Mol. Evol.* **41**, 675–679.
10. Bult, C. J., White, O., Olsen, G. J., Zhou, L., Fleischmann, R. D., *et al.* (1996) *Science* **273**, 1058–1073.
11. Gogarten, J. P., Kibak, H., Dittrich, P., Taiz, L., Bowman, E. J., Bowman, B. J., Manolson, M. F., Poole, R. J., Date, T., Oshima, T., Konishi, J., Denda, K. & Yoshida, M. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 6661–6665.
12. Iwabe, N., Kuma, K.-I., Hasegawa, M., Osawa, S. & Miyata, T. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 9355–9359.
13. Rivera, M. C. & Lake, J. A. (1992) *Science* **257**, 74–76.
14. Brown, J. R. & Doolittle, W. F. (1995) *Proc. Natl. Acad. Sci. USA* **92**, 2441–2445.
15. Keeling, P. J. & Doolittle, W. F. (1995) *Proc. Natl. Acad. Sci. USA* **92**, 5761–5764.
16. Baldauf, S. L., Palmer, J. D. & Doolittle, W. F. (1996) *Proc. Natl. Acad. Sci. USA* **93**, 7749–7754.
17. Brendel, V., Brocchieri, L., Sandler, S. J., Clark, A. J. & Karlin, S. (1997) *J. Mol. Evol.* **44**, 528–541.
18. Keeling, P. J. & Doolittle, W. F. (1995) *Proc. Natl. Acad. Sci. USA* **94**, 1270–1275.
19. Smith, M. W., Feng, D. F. & Doolittle, R. F. (1992) *Trends Biochem. Sci.* **17**, 489–493.
20. Brown, J. R., Masuchi, Y., Robb, F. T. & Doolittle, W. F. (1994) *J. Mol. Evol.* **38**, 566–576.
21. Feng, D. F. & Doolittle, R. F. (1996) *Methods Enzymol.* **266**, 368–382.
22. Henikoff, S. & Henikoff, J. G. (1993) *Proteins Struct. Funct. Genet.* **17**, 49–61.
23. Feng, D. F., Johnson, M. J. & Doolittle, R. F. (1985) *J. Mol. Evol.* **21**, 112–125.
24. Feng, D.-F. & Doolittle, R. F. (1997) *J. Mol. Evol.* **44**, 361–370.
25. Hasegawa, M., Hashimoto, T., Adachi, J., Iwabe, N. & Miyata, T. (1993) *J. Mol. Evol.* **36**, 380–388.
26. Wainright, P. O., Hinkle, G., Sogin, M. L. & Stickel, S. K. (1993) *Science* **260**, 340–342.
27. Baldauf, S. L. & Palmer, J. D. (1993) *Proc. Natl. Acad. Sci., USA* **90**, 11558–11562.
28. Wray, G. A., Levinton, J. S., Shapiro, L. H., (1996) *Science* **274**, 568–573.
29. Nikoh, N., Iwabe, N., Kuma, K., Ohno, M., Sugiyama, T., Watanabe, Y., Yasui, K., Shi-cui, Z., Hori, K., Shimura, Y. & Miyata, T. (1997) *J. Mol. Evol.* **45**, 97–106.
30. Hartman, H. (1984) *Speculations Sci. Technol.* **7**, 77–81.
31. Puhler, G., Leffers, H., Gropp, F., Palm, P., Klenk, H.-P., Lottspeich, F., Garrett, R. A. & Zillig, W. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 4569–4573.
32. Sogin, M. (1991) *Curr. Opin. Genet. Dev.* **1**, 457–463.
33. Golding, G. B. & Gupta, R. S. (1995) *Mol. Biol. Evol.* **12**, 1–6.
34. Roger, A. J., Clark, C. G. & Doolittle, W. F. (1996) *Proc. Natl. Acad. Sci. USA* **93**, 14618–14622.
35. Germot, A., Philippe, H. & Le Guyader, H. (1996) *Proc. Natl. Acad. Sci. USA* **93**, 14614–14617.
36. Bustard, K. & Gupta, R. S. (1997) *J. Mol. Evol.* **45**, 193–205.
37. Knoll, A. H. (1992) *Science* **256**, 622–627.
38. Han, T. H. & Runnegar, B. (1992) *Science* **257**, 232–235.