selective sweeps of rare beneficial mutations—in environments that are essentially constant in terms of selection for those mutations—might account for few or many cases of rapid evolutionary change in nature. But we also see no compelling reason to dismiss the role of a mechanism for punctuated evolution that requires only the two most basic evolutionary processes of mutation and natural selection.

Santiago F. Elena
Vaughn S. Cooper
Richard E. Lenski

Center for Microbial Ecology, Michigan State University, East Lansing, MI 48824, USA

REFERENCES

1. S. F. Elena, V. S. Cooper, R. E. Lenski, *Science* **272**, 1802 (1996).
2. K. C. Atwood, L. K. Schneider, F. J. Ryan, *Proc. Natl. Acad. Sci. U.S.A.* **37**, 146 (1951).
3. R. E. Lenski and M. Travisano, *ibid.* **91**, 6808 (1994).
4. S. J. Gould, *The Panda's Thumb* (Norton, New York, 1980), chap. 18.

29 July 1996; accepted 24 September 1996

Dating the Cenancester of Organisms

Russell F. Doolittle *et al.* (*1*) argue that their evidence indicates that the most recent common ancestor (cenancestor) of present-day organisms occurred about 2 billion years ago (Ga). This is considerably more recent than is commonly believed. Doolittle *et al.* recognize that the estimate of that time is related to a distance that in turn depends on the fraction of the amino acid positions of the sequence that are variable. They determined this as the fraction of the sites that have varied in at least one sequence. This is sufficiently large, 90 to 95%, so that a correction for the invariable positions would not materially affect their conclusions. But this assumes that any site that has varied once anywhere is forever variable. There is considerable evidence that the positions at which amino acid replacements are accepted are different in different taxonomic groups, or that there may be a sizable number of invariable positions, or both (*2*–*6*). This could not be correct unless positions in the sequences could sometimes be variable and at other times invariable, the point of the covarion model of protein evolution.

This description means that many of the varied sites may nevertheless have spent a considerable portion of their time in the invariable category. The estimates of the percent of the sequence that are covarions (percent of the sequence that is variable at any one time) include cytochrome c (*2*, *3*), α hemoglobin, β hemoglobin, ribonuclease, fibrinopeptide A, and insulin C peptide (*4*), as well as six mitochondrial proteins (*5*). In these proteins there are a total of over 2000 codons, only 614 (30%) of which are covarions, implying that 70% of the sites are not variable at any one time and for which many might rarely be variable. Doolittle *et al.* indicate that their dating of the cenancestor would be in line with current beliefs if 35% of the sites were invariable. If the figures for the proteins we just cited are representative of the proteins that Doolittle *et al.* used, the *effective* number of invariable sites may well be large enough to demand a correction which would make the current estimated 3.5-Ga date for the cenancestor within the error of their method. Because Doolittle *et al.* have not estimated the number of covarions nor how fast their composition changes for their sequences, one cannot be sure that the effective size of the invariable group is not as large as 35%. Their Poisson correction cannot address this problem.

This criticism does not depend on the correctness of the covarion model. As long as some positions are more variable than others, the distribution of mutations across sites is often much better matched with a gamma function. Had Doolittle *et al.*'s distances been corrected in this manner, their estimated time for the cenancestor would have been greater, possibly giving a date of 3.5 Ga.

Note added in proof: M. M. Miyamoto and I (W.M.F.) (*7*) have now shown that the data from the report by Doolittle *et al.* can be fit with a gamma distribution and, depending on the parameter chosen, obtain dates for the cenancestor ranging from the data proposed by Doolittle *et al.* of 2 Ga to values larger than 3.5 Ga.

Masami Hasegawa
Institute of Statistical Mathematics, 4-6-7 Minami-Azabu, Minato-ku, Tokyo 106, Japan
Walter M. Fitch
Department of Ecology and Evolutionary Biology, University of California, Irvine, CA 92717, USA

REFERENCES

1. R. F. Doolittle, D.-F. Feng, S. Tsang, G. Cho, E. Little, *Science* **271**, 470 (1996).
2. W. M. Fitch and E. Markowitz, *Biochem. Gen.* **4**, 579 (1970).
3. W. M. Fitch, *ibid.* **5**, 231 (1971).
4. ———, *J. Mol. Evol.* **8**, 13 (1976).
5. J. Adachi and M. Hasegawa, *ibid.* **40**, 622 (1995).
6. J. H. Reeves, *ibid.* **35**, 17 (1992); A. Sidow, T. Nguyen, T. P. Speed, *ibid.*, p. 253.
7. M. M. Miyamoto and W. F. Fitch, *Syst. Biol.* **45**, 566 (1996).

19 April 1996; accepted 28 August 1996

Doolittle *et al.* (*1*) calculate dates for two major events in the early evolution of life: the last common ancestor of all now-living organisms and the ancestor of eukaryotes and prokaryotes. These dates are derived from the molecular records stored in the various proteins of extant organisms. To calibrate the molecular clock, Doolittle *et al.* rely on data for metazoan evolution derived from the fossil record. They date the prokaryote-eukaryote ancestor and the last common ancestor to about 2000 Ma. These calibrations are wrong because: (i) they do not properly consider horizontal gene transfer, and (ii) they significantly underestimate the effect of differing substitution rates among amino acid positions on their calibration of the molecular clock.

1) Doolittle *et al.* mention (*1*) horizontal gene transfer as a possible complication; however, in most cases the reported analyses include too few bacterial species to allow for the detection of horizontal gene transfer events. For example: if more bacterial triose phosphate isomerase (TPI) sequences are added to the data set used by Doolittle *et al.*, one recognizes that the eukaryotic TPIs group with their counterparts from purple bacteria (*2*). P. J. Keeling and W. F. Doolittle have interpreted this relation to indicate that the eukaryotic TPI was contributed to the eukaryotic cell by the endosymbiont that evolved into the mitochondria (*3*). If this interpretation is correct, then the prokaryote-eukaryote split determined for this enzyme does not reflect the origin of the eukaryotes, but rather the uptake of mitochondria into the eukaryotic cell.

The same argument can be made for those enzymes found in archaea that indicate a close relation between archaea and eubacteria. Several molecular phylogenies have been reported that indicate such closeness—in particular, between gram positive bacteria and archaea—whereas other enzymes and especially functions involved in DNA duplication, transcription, and translation indicate a close association between the archaea and the nucleocytoplasmic component of the eukaryotic cell (*4*). Including more prokaryotic species in the analyses (*5*), we found the representatives of the archaea grouping within the eubacteria in the several cases (*6*). If the enzymes that group the archaea within the eubacterial domain were obtained by the archaea through horizontal transfer from eubacteria, then this event (or these events) of hori-

zontal gene transfer or lineage fusion (*4*) is reflected in the dating and not the age of the last common ancestor.

2) As Doolittle *et al.* recognize (*1*), proteins are conserved over billions of years because natural selection eliminates amino acid substitutions that significantly reduce functionality. Sites involved in catalysis or substrate binding often do not vary at all, while others may vary considerably without affecting protein functionality, leading to variation among sites in substitution rates. Doolittle *et al.* consider the influence of among-site rate variation on their molecular clock calibration, but significantly underestimate its effect. Rate variation among sites appears to be much greater than can be accounted for by irreplaceable residues alone and much greater than the "extreme distribution of probabilities" that Doolittle *et al.* assumed in their simulations.

The gamma distribution has been widely used in analyses of nucleotide sequence data to examine the effects of among-site rate variation on phylogenetic analysis and molecular distance estimates. The shape parameter, α, is inversely related to the amount of rate variation among sites. Low values of α correspond to large differences among sites in underlying rates of substitution, while high values correspond to small differences. Estimates of the shape parameter obtained from aligned amino acid sequences for the 70-kD heat shock protein family, the triose phosphate isomerases, and V-ATPase catalytic subunits range from 0.57 to 1.21 (*7*). A model assuming an invariant class of sites and a gamma distribution of rates among remaining sites provides a significantly better fit to the data and gives estimates of the shape parameter ranging from 2.42 to 3.29 with between 18 and 21% of the sites invariant (Fig. 1).

At low levels of difference, where Doolittle *et al.* calibrated the curve, there is little difference among the curves. For levels of difference that reflect the time of the prokaryote-eukaryote split, the curves diverge significantly. Using the estimate derived from parsimony analysis without invariant sites (α = 1.21) and 70% difference, we found that an uncorrected formula underestimates the amount of divergence by 45%. Using the estimate of α = 3.29 and 18% invariant sites, an uncorrected formula underestimates divergence by almost 60%. Using the estimate of α = 0.7 derived from maximum-likelihood analysis, an uncorrected formula underestimates the number of substitutions by more than 66%. Correcting the data that Doolittle *et al.* present according to these estimates would date the prokaryote-eukaryote split at between 3.5 and 6 Ga. Even if we accept the contention by Doolittle *et al.* that at most 5% of sites are invariant, and if we use the largest estimate of α we obtained (the most generous combination of parameters possible), the date of the prokaryote-eukaryote split would be no more recent than 2.8 Ga.

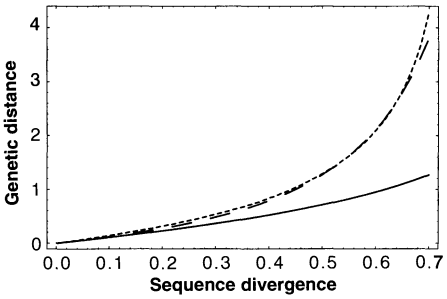

Fig. 1. Relations between the average number of substitution events per site (*y*-axis) and the observed differences (*x*-axis). Observed distance = (1 − identity) × 100. Solid line depicts the relation without correction for variation among sites; dashed line depicts the relation with an among-site variation described by a gamma distribution with α = 0.7; dotted line depicts the relation assuming α = 2.4 and 20% invariant sites.

J. Peter Gogarten
Lorraine Olendzenski
Elena Hilario
Department of Molecular and Cell Biology, University of Connecticut, Storrs, CT 06269–3044, USA
Chris Simon
Kent E. Holsinger
Department of Ecology and Evolutionary Biology, University of Connecticut

REFERENCES AND NOTES

1. R. F. Doolittle, D.-F. Feng, S. Tsang, G. Cho, E. Little, *Science* **271**, 470 (1996).
2. We added TPI sequences from *Thermotoga maritima*, *Mycoplasma flocculare*, and a *Vibrio* species. For further description, see (*5*).
3. P. J. Keeling and W. F. Doolittle, personal communication; W. F. Doolittle, in *Evolution of Microbial Life*, D. McL. Roberts *et al.*, Eds. (Cambridge Univ. Press, New York, 1996), pp. 1–21.
4. Some of the data were reviewed by J. P. Gogarten [*Trends Ecol. Evol.* **10**, 147 (1996)]. For a more recent summary, see J. P. Gogarten, E. Hilario, L. Olendzenski, in *Evolution of Microbial Life*, D. McL. Roberts *et al.*, Eds. (Cambridge Univ. Press, New York, 1996), pp. 267–292.
5. Sequences were from R. F. Doolittle *et al.* (*1*) and downloaded from the NCBI (entrez@ncbi.nlm.nih.gov) using NENTREZ and aligned using clustalw [J. D. Thompson, D. G. Higgins, T. J. Gibson, *Nucleic Acids Res.* **22**, 4673 (1994)]. Phylogenies were calculated using the distance matrix analysis implemented in clustalw (ignoring positions with gaps) and the programs PROTPARS, PROTDIST, and FITCH from PHYLIP [J. Felsenstein, Phylogeny Inference Package 3.5c (1993), Univ. of Washington, Seattle, WA]. In the latter two cases, gaps were encoded as missing data.
6. In none of the phylogenies calculated for dihydrofolate reductases, dihydrolipoamide dehydrogenase, phosphoglycerate kinase, and enolase did the representatives of the archaea form the deepest (that is, closest to the eukaryotes) branch within the prokaryotes. Only in the cases of argininosuccinate synthase and porphobilinogen synthase did the enlarged data sets still support the topology given by Doolittle *et al.* in figure 3 in their article (*1*).
7. We used two methods to arrive at these estimates: (i) the parsimony-based method of Sullivan *et al.* [*Mol. Biol. Evol.* **12**, 988 (1995)] used for analysis of nucleotide sequences, and (ii) a maximum-likelihood method developed by Z. Yang [Phylogenetic analysis by maximum likelihood (PAML), Version 1.1. (1995) Institute of Molecular Evolutionary Genetics, Pennsylvania State University, University Park, PA]. The former method (i) does not employ a particular model of amino acid substitution, and, as expected, estimates of α derived from it (0.98 to 1.21) are substantially higher than those derived from the maximum-likelihood method (0.57 to 0.77). Maximum-likelihood estimates derived from a Dayhoff amino acid substitution model differed from those derived from a Jones model [D. T. Jones, W. R. Taylor, J. M. Thornton, *Comput. Applic. Biosci.* **8**, 275 (1992)] by 4% or less for adenosine triphosphatase (0.68 and 0.66) and hsp70 (0.76 and 0.76). The estimate of α from the Dayhoff model was 0.66, but only 0.57 for the Jones model with TPI. We accept α = 0.7 as a reasonable estimate for the effect of among-site rate variation from the maximum-likelihood analysis.

25 April 1996; accepted 28 August 1996

Response: Hasagawa and Fitch and Gogarten *et al.* raise the question of whether our calculated values for the most distantly related sequences (especially those involved in measuring the divergence of prokaryotes and eukaryotes) were underestimated because we did not take sufficient account of differential rates of change at individual amino acid positions ("sites"). Both comments suggest that a gamma distribution could have been employed to estimate the magnitude of a correction needed over and beyond a simple Poisson expectation. The method we used (*1*), however, is not an uncorrected Poisson, but employs a specified amino acid substitution table and has virtues not unlike the empirical process introduced by Dayhoff (*2*) that correct the other weakness in a simple Poisson: namely, the assumption that all amino acid interchanges are equally likely. The results of this correction are not unlike those obtained from the use of a gamma distribution calculated with reasonable parameters (*3*).

The point can best be made by comparing the relationship between distance (*d*) and fraction of residues changed (*p*) for a number of procedures, including the plot provided by Gogarten *et al.* which uses a gamma distribution. As pointed out in both comments, results obtained from a gamma distribution are sensitive to the choice of the *a* parameter and to the fraction of irreplaceable residues. The *a* parameter can be determined empirically by counting (phyletically) the number of changes at each individual site of a nucleic acid or protein. Other things being equal, a low value for *a* results in a greater deviation from the simple Poisson expectation

In figure 1 of their comment, Gogarten

et al. use a low value of a in one case and a high fraction of irreplaceable residues in the other to emphasize how different can be the observed and actual substitutions in the cases of a gamma distribution compared with a simple uncorrected Poisson. In contrast, attempts by others to make general corrections for a gamma distribution (*3*) have resulted in much more modest corrections (Fig. 1A). In this regard, Ota and Nei (*3*) used $a = 2$, a value which resulted in curves that closely overlay an empirical curve generated by Dayhoff *et al.* (*2*), even though the former corrects for "sites" and the latter for constraints on substitution. The value of $a = 2$ was also in close correspondence to an empirical finding in the initial study by Uzzell and Corbin (*4*), which suggested that a negative binomial distribution with values drawn from a gamma distribution has a better fit to rates of change in cytochrome c than does a simple Poisson.

Plots of distances calculated by our method (*5*) have a unique quality: namely, the distances in the nearer range are smaller than expected relative to a simple Poisson (Fig. 1B). The reason is that the most probable ("early") changes in a protein sequence tend to be of the sort that yield high substitution scores. Consider, for example, a sequence of 100 residues in which ten changes occur and for which p will be close to 0.1. The S values, from which the distances will be calculated, will only change modestly, and the calculated d will tend to have a value near 0.05. To illustrate with the PAM-250 scoring table, a change of a valine to an isoleucine has virtually no impact on the score. As sequences become increasingly different, of course, this tendency for inconsequential change is eroded. As a consequence, the more distant sequences exhibit lower S values and higher distances (Fig. 1B).

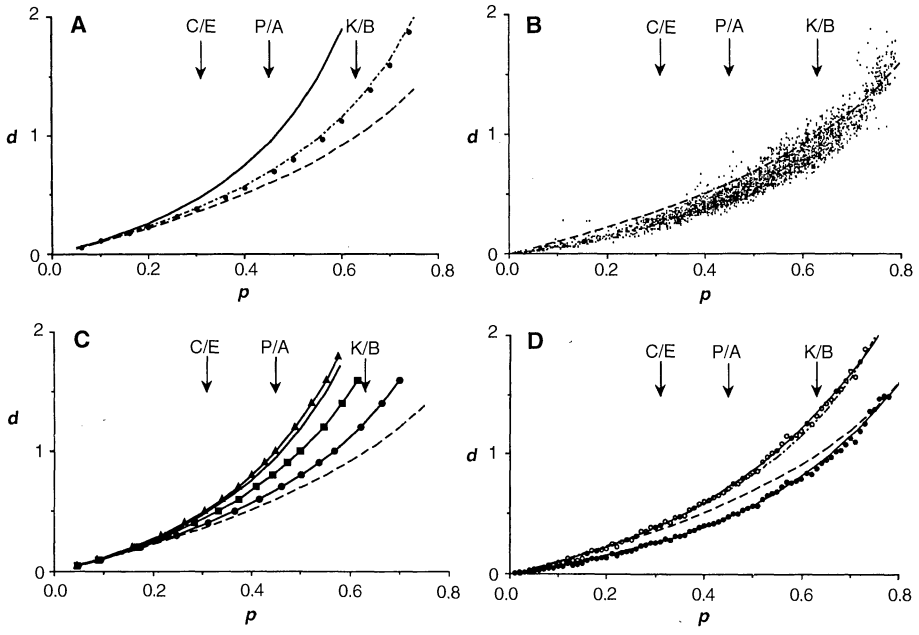

Fig. 1. Plots of calculated distance values (d) plotted against fraction of changed residues (p). Key divergence points are designated by the vertical arrows: E/C, echinoderm/chordate; P/A, plant/animal; K/B, eukaryote/eubacteria. (**A**) Comparison of d/p curves obtained by Ota and Nei (*3*) (·–) using a gamma distribution ($a = 2$) with that of Gogarten *et al.*, who use $a = 0.7$ (dots). Uncorrected Poisson (––) and Dayhoff PAM procedure (solid line) are shown for comparison. (**B**) The 2,767 distances used in our study (*1*) plotted against fraction of changed residues; uncorrected Poisson (––). (**C**) Curves plotted from Grishin equations (*6*). Uncorrected Poisson (––); correction for amino acid interchange (○); correction for sites (□); and general correction for sites and amino acids (△). Gogarten curve from panel A is shown for comparison. (**D**) Average distance curve from our study (solid line); uncorrected Poisson (––); gamma distribution curve of Ota and Nei (*3*) (·–); and our average curve multiplied by a scaling factor (1.5) that superimposes the nearer region of the curve on the uncorrected Poisson (○).

Table 1. Distances at some key P values determined by different methods.*

Method	C/E $P =$ 0.31	P/A $P =$ 0.45	K/B $P =$ 0.63	Ratio 0.63/ 0.31	Time (My)† K/B div.	Ratio 0.63/ 0.45	Time (My)‡ K/B div.	Ratio 0.45/ 0.31	Time (My)† P/A div.
Grishin (*6*)	0.51	1.00	2.35	4.61	2536	2.35	2350	1.96	1078
Gogarten *et al.*	0.49	0.94	2.20	4.49	2470	2.34	2340	1.92	1056
Doolittle *et al.* (*1*)	0.27	0.48	0.88	3.26	1793	1.83	1830	1.78	979
Ota and Nei (*3*)	0.41	0.70	1.30	3.17	1744	1.86	1860	1.71	941
Dayhoff *et al.* (*2*)	0.40	0.67	1.24	3.10	1705	1.85	1850	1.68	924
Uncorrected Poisson	0.37	0.60	0.99	2.68	1474	1.65	1650	1.62	891

*Distances have been taken from the data plotted in Fig. 1. †Time (My) obtained by multiplying the ratio by the (fossil record) divergence time of echinoderms (E) and chordates (C), which is taken as 550 My. ‡Time (My) obtained by multiplying ratio by 1000 My, a proposed divergence time for plants (P) and animals (A).

Until recently, we would have dismissed Gogarten *et al.*'s choice of parameters for a gamma distribution as overzealous in the extreme. However, during the course of preparing this response, we plotted the several versions of Poisson corrections recently described by Grishin (*6*). Unexpectedly, his theoretical formulation correcting both for rate variation at sites and for the nature of amino acid replacement fell remarkably close to the curves submitted by Gogarten *et al.* (Fig. 1C). The Grishin equation uses an exponential distribution of rates for different sites that is equivalent to a gamma distribution with an a factor = 1.

What impact do all these corrections have on distance-time considerations? If d values are read off the various curves (Fig. 1) at specified values of p, relative divergence times can be obtained directly. For example, $p = 0.31$ corresponds to the deepest divergence point (echinoderm-chordate = E/C) used in our calibration series, $p = 0.45$ to the plant-animal divergence (P/A) and the third point, and $p = 0.63$ to the controversial divergence point for eukaryotes and eubacteria (K/B). The ratios of distances for these points reveal how effective each approach is in correcting distant points over and beyond the simple Poisson (Table 1). The longest time obtained for the prokaryote-eukaryote divergence by this procedure is slightly more than 2500 My, by the Grishin equation (*6*), and slightly less than 2500 My, with the curve from Gogarten *et al.* These values stand in stark contrast to the exaggerated times suggested by Gogarten *et al.* in their comment. Their results may be a result, at least in part, to their assuming that our formulation is a simple Poisson, on the one hand, and, perhaps, because they did not calibrate their own curve to known fossil record divergences.

Beyond that, our method does as well as the Ota and Nei procedure (*3*)—which uses a gamma distribution, but which does not take account of individual amino acid variation—and also as well, or better, as the Dayhoff procedure (*2*), which takes account of amino acid substitution preferences, but

does not take into account the rate differences at different sites. The point is better illustrated when our distances are multiplied by an empirical constant (1.5) that makes the early values ($p < 0.3$) coincident with a simple Poisson (Fig. 1D). The range from $p = 0$ to $p = 0.31$ corresponds to the fossil record divergence times that were used to calibrate our time scale.

Because our method does not provide a correction for site variation [a limitation we acknowledged (*1*)], we applied a 10 to 15% correction based on a simulation experiment. Now the theoretical equation of Grishin (6) suggests that a more realistic correction would have been 25 to 30%. In no case, however—not even the "worst case" parameter selection of Gogarten *et al.*—do these corrections lead to prokaryote-eukaryote divergence times greater than 2500 Ma (Table 1).

On a slightly different tack, one of the virtues of restricting our study to enzymes has been that the sequences are relatively slow changing, and the overall pairwise similarities remain in a range that has usually been thought to need minimal correction (*7*). This property served as the fulcrum for an internal measure that we felt corroborated our general interpretation; the determination of longest distances using the 27 fastest-changing enzyme sequences gave values that were not radically different from those found with the 27 slowest changing. Because deviations are a function of dissimilarity (Fig. 1), the 27 slowest changing proteins should have given the more reliable divergence time; in fact, it put the divergence nearer to the present.

Turning to the matter of possible horizontal transfers resulting from sequences being imported during the acquisition of organelles or otherwise, in our article (*1*) we noted that, inadvertantly, some of the sequences used may have had such a history, but unless their numbers were excessive, the results would not be greatly affected. We also implied that some of the sequences used may have been paralogs and not orthologs, noting particularly in our comparison of the slowest changing half of the set with the fastest changing half (as judged by prokaryote-eukaryote differences) that the former would be more likely to have some horizontal imports among them and the latter more likely to have paralogs. Newly available sequences make it clear that at least three of the 57 enzymes used contained bacterial paralogs (glyceraldehyde 3-phosphate dehydrogenase, glutamine synthetase, and argininosuccinate synthase). These have now been corrected, and, not unexpectedly, the corresponding distances are somewhat smaller. These changes would tend to be offset by the inclusion of any entries that turn out to be horizontal imports, which would have artifactually smaller distances reflecting a short circuit.

The relation between the archaebacteria and eubacteria needs special comment. As Gogarten *et al.* correctly state, our original study (*1*) did not have sufficient archaebacterial representation to make a firm statement about their phylogenetic position (we had called attention to the fact in footnote 25 of our article). During the interval since the submission of the manuscript for that article (April 1995), many more archaebacterial sequences have become available (8), and we have been able to increase the number of enzymes that have archaebacterial representation from nine to 16. Of the 16, we now find that in ten cases the archaebacteria cluster strongly with the eubacteria. In three others the archaebacteria sequences are clearly more similar to eukaryotes, and in the remaining three the eubacteria and eukaryote sequences are significantly more similar to each other than either is to the archaebacteria.

Given these observations, we agree, as noted in our article, that there may have been some kind of extensive horizontal transfer among the bacteria or, possibly, a chimeric fusion involving archaebacteria on the way to the eukaryotic cell. With regard to the latter, we tend to favor the suggestions of Hartman (9) and Sogin (*10*) that the host cell was an early diverging entity, over the notion of Gupta and Golding (*11*) that a Gram-negative bacterium was involved. Whatever the case, our major findings are not much affected by the omission of any seqeunces that may have been involved in that hypothetical acquisition.

Russell F. Doolittle
Da-Fei Feng
Simon Tsang
Glen Cho
Elizabeth Little
Center for Molecular Genetics,
University of California, San Diego,
La Jolla, CA 92093–0634, USA
E-mail: rdoolittle@ucsd.edu

REFERENCES AND NOTES

1. R. F. Doolittle, D.-F. Feng, S. Tsang, G. Cho, E. Little, *Science* **271**, 470 (1996).
2. M. O. Dayhoff, R. M. Schwartz, B. C. Orcutt, in *Atlas of Protein Sequence and Structure* (National Biomedical Research Foundation, Washington, DC, 1978), vol. 5, suppl. 3, pp. 345–352.
3. T. Ota and M. Nei, *J. Mol. Evol.* **38**, 642 (1994).
4. T. Uzzell and K. W. Corbin, *Science* **172**, 1089 (1971).
5. The distance is equal to the negative log of similarity score corrected for the nature of the amino acid composition, according to the formula $d = -1n(S_{obs} - S_{rand}/S_{ident} - S_{rand})$, where S is obtained from an amino acid substitution table—for example, the PAM-250 or BLOSUM-62 tables. The effectiveness of this formula to relate distance with evolutionary time has been tested by simulation with a computer model that follows the divergence of mutated protein sequences under various circumstances of constraint.
6. N. V. Grishin, *J. Mol. Evol.* **41**, 675 (1995).
7. M. Nei, *Molecular Evolutionary Genetics* (Columbia Univ. Press, New York, 1987).
8. The overall sequence count from the 57 selected enzymes has increased from 531 to 650, mostly with new sequences from microorganisms, but also with more "bridge organisms" and fossil record representatives (D.-F. Feng, G. Cho, R. F. Doolittle, in preparation).
9. H. Hartman, *Speculations Sci. Technol.* **7**, 77 (1984).
10. M. Sogin, *Curr. Opin. Genet. Devel.* **1**, 457 (1991).
11. R. S. Gupta and G. B. Golding, *J. Mol. Evol.* **37**, 573 (1993); G. B. Golding and R. S. Gupta, *Mol. Biol. Evol.* **12**, 1 (1995).

12 June 1996; accepted 28 August 1996