Proc. Natl. Acad. Sci. USA
Vol. 90, pp. 3009–3013, April 1993
Evolution

Evolution of the glutamine synthetase gene, one of the oldest existing and functioning genes

(gene duplication/molecular evolution/pre-prokaryote–eukaryote evolution)

Y. KUMADA*, D. R. BENSON†, D. HILLEMANN‡, T. J. HOSTED†, D. A. ROCHEFORT†, C. J. THOMPSON§, W. WOHLLEBEN‡, AND Y. TATENO¶‖

*Meiji Seika Kaisha Ltd., Chuoh-ku, Tokyo 104, Japan; †Department of Molecular and Cell Biology, University of Connecticut, Storrs, CT 06269-3044; ‡Biologie VI, Fakultät für Biologie, Universität Bielefeld, 4800 Bielefeld, Federal Republic of Germany; §Institute Pasteur, Paris Cedex 16, France; and ¶National Institute of Genetics, Mishima 411, Japan

Communicated by Motoo Kimura, December 14, 1992

ABSTRACT **We performed molecular phylogenetic analyses of glutamine synthetase (GS) genes in order to investigate their evolutionary history. The analyses were done on 30 DNA sequences of the GS gene which included both prokaryotes and eukaryotes. Two types of GS genes are known at present: the GSI gene found so far only in prokaryotes and the GSII gene found in both prokaryotes and eukaryotes. Our study has shown that the two types of GS gene were produced by a gene duplication which preceded, perhaps by >1000 million years, the divergence of eukaryotes and prokaryotes. The results are consistent with the facts that (*i*) GS is a key enzyme of nitrogen metabolism found in all extant life forms and (*ii*) the oldest biological fossils date back 3800 million years. Thus, we suggest that GS genes are one of the oldest existing and functioning genes in the history of gene evolution and that GSI genes should also exist in eukaryotes. Furthermore, our study may stimulate investigation on the evolution of "preprokaryotes," by which we mean the organisms that existed during the era between the origin of life and the divergence of prokaryotes and eukaryotes.**

Glutamine synthetase (GS) is a key enzyme in nitrogen metabolism; it has dual functions in two essential biochemical reactions, ammonia assimilation and glutamine biosynthesis (1, 2). It is also one of the few amide synthetases found in organisms. Prokaryotes and eukaryotes were once thought to synthesize different GSs: GSI for the former and GSII for the latter. It is now known, however, that GSII is also present in bacteria belonging to *Rhizobiaceae* (3–6), *Frankiaceae* (7), and *Streptomycetaceae* (8, 9). GSI, by contrast, has not been found in any eukaryote.

Glutamine produced by GS is essential for protein synthesis, and its amide nitrogen is donated to synthesize many essential metabolites. It is thus natural to consider GS as present in, and probably indispensable to, all organisms. In view of the central roles played by GS, it is reasonable to believe that the GS gene is extremely old. From the sequence alignment of GSI from *Salmonella typhimurium* and GSII from alfalfa (10), we could observe that the differences in amino acids between them was 0.75 per site. This value is quite large compared with those for other proteins, suggesting also that the GSI and GSII genes share a very old common ancestor.

The aforementioned discovery of the GSII gene in plant symbiotic bacteria led to the suggestion that the gene had originated from host plants through lateral gene transfer (3). This was later questioned by the further findings of the GSII gene in plant nonsymbiotic actinomycetes (8, 9). Shatters and Kahn (6) have suggested that the common ancestor of the GSII genes in *Rhizobiaceae* and in the host plant must be older than the plant itself, and have argued against the gene transfer.

In this paper we have traced the evolutionary history of the GS genes, using our own nucleotide sequence data and others' data from prokaryotic and eukaryotic species in order to estimate the age of the GS gene and resolve the controversy over the gene transfer. This attempt may stimulate evolutionary studies into pre-prokaryote–eukaryote divergence.

ACQUISITION AND ALIGNMENT OF GS SEQUENCES

We sequenced the GSI genes of *Streptomyces viridochromogenes* and *Frankia*. The sequence data on the GSII genes of these species and our data together formed the only case where both the GSI and GSII genes present in the same genome were completely sequenced. The GSI sequence of *Frankia* confirms that the GSI and GSII genes are separated by only about 500 base pairs in the genome (7). We also collected complete nucleotide sequence data of the GS genes from published papers and from the GenBank/EMBL/DDBJ database (Release 67.0). The total of 30 homologous sequences thus acquired are listed in Table 1.

By eye, we simultaneously aligned the 30 amino acid sequences that were deduced from the nucleotide sequences. In so doing we referred to *Salmonella* GSI, whose three-dimensional structure has been elucidated (10). We were unable to align the *Bacteroides fragilis* GS sequence (31), because it was atypical to the other sequences in several aspects.

RESULTS

Alignment of GS Sequences. The results of our alignment are in general agreement with previous studies that dealt with fewer GS sequences. To examine conservation of physicochemical characteristics within the sequences, we classified the amino acids in the aligned sequences as hydrophobic, ambivalent, basic hydrophilic, and acidic hydrophilic (32). The results show that, in most sites, hydrophilic amino acids align with other hydrophilic or ambivalent ones, and hydrophobic amino acids align with other hydrophobic or ambivalent ones. Few cases were observed where hydrophilic amino acids aligned with hydrophobic amino acids. (The aligned sequence data will be provided on floppy disk upon request to Y.T.)

From the results of the alignment, two features emerged. First, the four regions of the active center (10) were observed among all the sequences. This indicates that all the genes for

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.

Abbreviation: GS, glutamine synthetase.

‖To whom reprint requests should be addressed.

Table 1. Names and origins for 30 nucleotide sequences of GS genes

Gene*	Origin	Ref.
Mv I	*Methanococcus voltae*	11
Bs I	*Bacillus subtilis*	12
Ca I	*Clostridium acetobutylicum*	13
An I	*Anabaena* 7120	14
Tf I	*Thiobacillus ferrooxidans*	15
Ec I	*Escherichia coli*	16
St I	*Salmonella typhimurium*	17
Ab I	*Azospirillum brasilense*	18
Bj II	*Bradyrhizobium japonicum*	3
Rl I	*Rhizobium leguminosarum*	19
Rm II	*Rhizobium meliloti*	6
Sc I	*Streptomyces coelicolor*	20
Fs II	*Frankia* sp. (CpI1)	7
Fs I	*Frankia* sp. (CpI1)	This study†
Sv I	*Streptomyces viridochromogenes*	This study†
Sv II	*Streptomyces viridochromogenes*	8
Sh II	*Streptomyces hygroscopicus* SF-1293	9
AF II	Alfalfa	21
KBR1 II	Kidney bean (R1 chain)	22
KBR2 II	Kidney bean (R2 chain)	22
GPCH II	Garden pea (chloroplast)	23
GP341 II	Garden pea (pGS341)	24
OS II	Rice	25
OSCH II	Rice (chloroplast)	25
Np II	*Nicotiana plumbaginifolia*	26
MA II	Man	27
CH II	Chinese hamster	28
CHK II	Chicken	29
DM II	*Drosophila melanogaster*	30
DMMT II	*D. melanogaster* (mitochondria)	30

*I or II refers to GSI or GSII.
†These sequences have been deposited in the GenBank/EMBL/DDBJ database [accession nos. X70924 (Sv I) and L10631 (Fs I)].

the 30 proteins came from a single ancestral gene and that the GSI and GSII genes were duplicated from the ancestral gene. The arrangement of the GSI and GSII genes in the *Frankia* genome mentioned above gives a line of supporting evidence for gene duplication. Second, the four conserved regions were characterized by an abundance of positions having chemically similar amino acids in both GSI and GSII, indicating that GSII enzymes also have four functional regions corresponding to those identified in GSI enzymes.

Estimation of Evolutionary Distances. In estimating evolutionary distances, we discarded the common regions among the 30 sequences where deletions are observed even in one sequence, since there is no way to assess the rate and size of deletions during the course of evolution. Omission of the regions left us with 273 comparable amino acid sites that were used to estimate evolutionary distances.

To estimate evolutionary distances between the nucleotide sequences of the GS genes, we considered nonsynonymous substitutions only, and used Nei and Gojobori's (33) method for estimating the number of nucleotide substitutions between each pair among the 30 sequences.

Estimation of Phylogenetic Trees. We used three methods for estimating the phylogenetic tree of the GS genes: the unweighted pair-group method using arithmetic averages (UPGMA; refs. 34 and 35), the modified Farris (MF) method (36), and the neighbor-joining (NJ) method (37). The three methods gave similar tree topologies; they differed in a few branching orders. Unlike the UPGMA, the MF and NJ methods cannot place the root in an estimated phylogenetic tree without additional information that locates the root. In the present case, we have this information; the GSI and GSII genes are duplicated genes. Thus, the root can be placed at the point of gene duplication which led to diversion of the two genes. Among the three rooted trees thus obtained, we have chosen the NJ tree shown in Fig. 1, because the NJ method gives the correct tree topology at a higher probability than the other two methods (38).

Since we were also interested in estimating the times of major evolutionary divergences, we extracted and modified a part of the tree in Fig. 1. In so doing, we used only the GSII gene, because it existed in both eukaryotes and prokaryotes. The GSII gene topology was borrowed from the tree in Fig. 1, and branch lengths were recomputed by assuming that the evolutionary rate of the GSII gene is the same for all branches. This assumption is unavoidable when estimating divergence times from a phylogenetic tree. It is noted that GS genes have evolved mostly by neutral mutation (39, 40), as will be discussed below. Thus, the assumption of rate constancy was compatible with the GS genes except for the *Drosophila* GS gene. Recent data indicate that some 20 *Drosophila* genes have evolved 2–3 times faster than the corresponding genes in other animals (41). Thus, the rapid evolutionary rate of the GSII gene for *Drosophila* seems to be the rule rather than the exception. In any event, since including the *Drosophila* gene made other branch lengths negative in recomputation, we omitted this gene in the procedures which resulted in the tree shown in Fig. 2.

The phylogenetic tree thus obtained was calibrated by Dickerson's (42) estimate that animal and plant kingdoms diverged 1200 million years ago. The rate of nonsynonymous substitution was estimated to be 0.16×10^{-9} per site per year. This figure suggests that the GSII gene evolved slowly in comparison to other genes and may thus be a good chronometer for studying long-term evolution. The calibration indicates that man and Chinese hamster diverged 80 million years ago and that these two and chicken diverged 170 million years ago (Fig. 2). The former estimate is in accordance with the well-established mammalian divergence time, and the latter is within the time range into which the other end of the avian lineage merges (43). The calibration also shows that the divergence time between eukaryotes and prokaryotes is about 1800 million years, which agrees with previous results (44). Furthermore, it implies that the duplication of the GSI and GSII genes occurred about 3500 million years ago. Pesole *et al.* (45) gave an estimate of 2500 ± 500 million years for the time of gene duplication. Our estimate is based on the assumption that the GSI and GSII genes in *Rhizobium*, *Frankia*, and *Streptomyces* diverged simultaneously at the time of their speciation.

DISCUSSION

We appreciate the fact that the results presented here have a number of profound implications whose significance can be evaluated only after understanding potential errors in estimating evolutionary trees derived from nucleotide sequence data. (*i*) Although we were very careful when aligning the amino acid sequences used, it is unlikely that it is 100% accurate. (*ii*) A unique mutation at a given site may alter the constraints and thus rates of change in other regions of the protein. (*iii*) The nucleotide sequences may not necessarily show the correct evolutionary processes of the GS genes, because a nucleotide at a single site may have been changed more than once. We have minimized this effect by basing our calculations on nonsynonymous substitutions whose sites are not likely to have reached saturation and by using Nei and Gojobori's (33) method, which takes into account multiple substitutions at a site. (*iv*) No tree-estimation method always gives the correct tree, even in the simplest case where evolution is driven only by neutral mutations (36, 38, 46). These errors are unavoidable in any study of evolutionary relationships of species or genes using nucleotide or amino

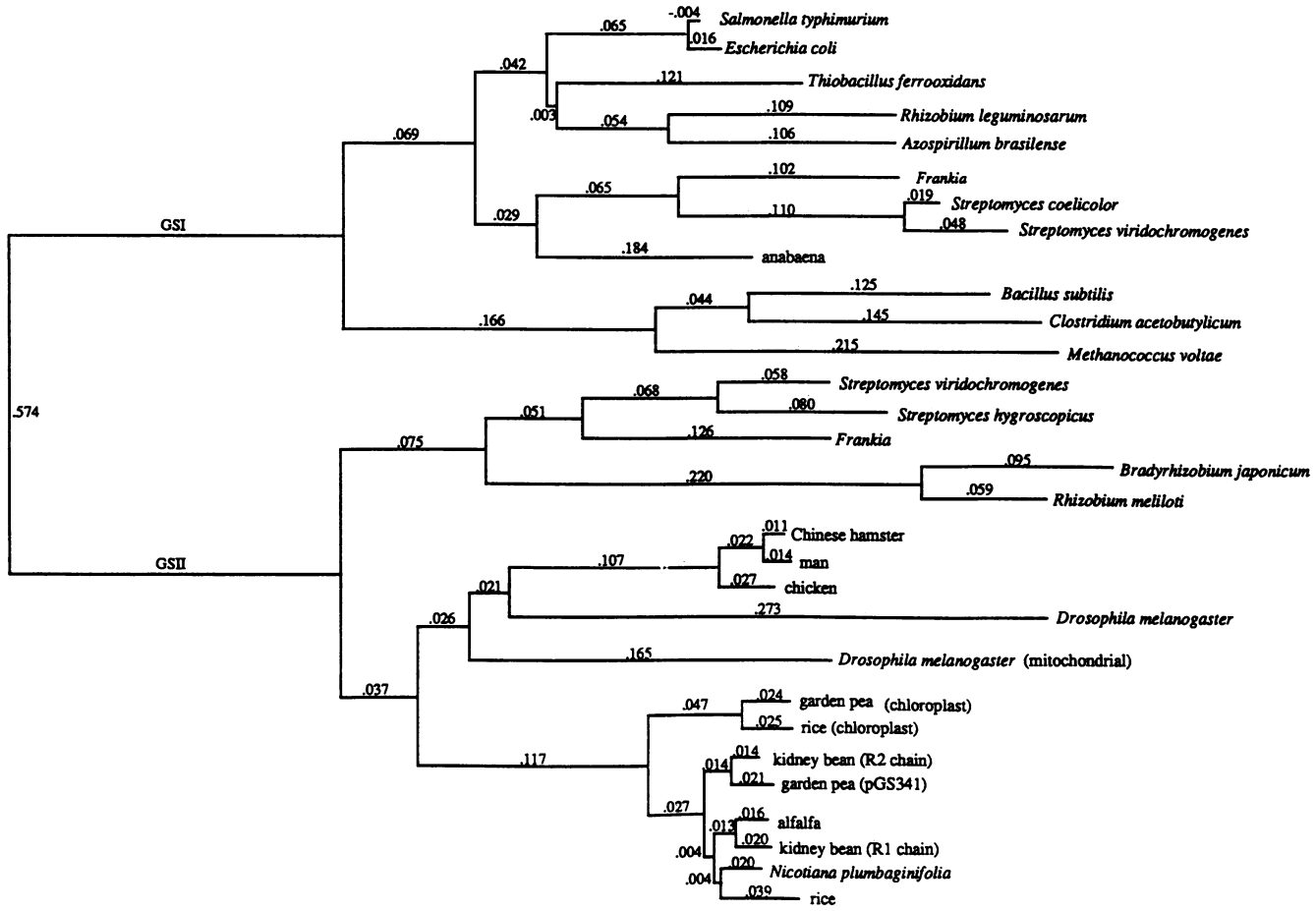

FIG. 1. Phylogenetic tree of 30 GS genes. The tree was estimated by the neighbor-joining method. Each gene at the terminal is presented by the name of species from which the gene originated. Branch length is given as the number of nonsynonymous substitutions per site. GSI and GSII are, respectively, ancestral GSI and GSII genes.

acid sequence data. It is thus important to regard an estimated phylogenetic tree as a best approximation rather than as a proven fact.

Nevertheless, with a few notable exceptions, our GS gene trees are in agreement with well-accepted evolutionary divergences, as indicated above. The tree in Fig. 1 also agrees with other results (20) showing that actinomycetes, which are high-G+C Gram-positive bacteria, are more closely related to Gram-negative bacteria than to low-G+C Gram-positive bacteria. Furthermore, our results are remarkably consistent with geological and fossil records of evolution. Our estimated time of the gene duplication, 3500 million years ago, is consistent with the observation that the oldest sedimentary rocks having evidence of prolific autotrophic microbial life are 3800 million years old (47). It is thus probable that both GSI and GSII genes existed 1700 million years before the divergence of prokaryotes and eukaryotes. Though no GSI gene has been reported for eukaryotes, the phylogenetic tree in Fig. 1 strongly suggests that the GSI genes also exist in eukaryotes in normal form or in abnormal form, like pseudogenes. Alternatively, it is possible, though less plausible, that the gene might have been lost in a common ancestor of the eukaryotes.

The clustering of the GS genes of four plant species—alfalfa, kidney bean (R1), *Nicotiana*, and rice—in Fig. 2, however, is not consistent with well-established taxonomic relationships. The clustering can in fact be shown as statistically insignificant by the method of Nei *et al.* (48). As Fig. 2 shows, the branch connecting the four species to the higher cluster is 0.002 long, and that connecting *Nicotiana* and rice to the higher cluster is 0.003 long. The standard error computed by their method is 0.0043 for the former branch and 0.0045 for the latter branch.

Pesole *et al.* (45) noted that the second codon positions of the GS genes they studied had evolved regularly. They reasoned that the regularity was due to a relatively low variation of second codon positions. Because the second position is most responsible for amino acid changes among the three positions, they suggested that the rate constancy has been maintained by natural selection. Our studies of the GS genes, however, do not support such a Darwinian interpretation. The four conserved regions that are of functional importance as determined by crystallography studies have fewer amino acid variations than the other parts. This is precisely what the neutral mutation theory predicts and is just the opposite of how natural selection operates (39, 40). Darwinian theory predicts that functionally important regions of GS should evolve rapidly. Our analysis gives an additional basis for the interpretation based on the neutral theory: the rate of synonymous substitution is much higher than that of nonsynonymous substitution. For example, between man and Chinese hamster, which constitute the closest pair in the tree in Fig. 1, the rate of synonymous substitution is much higher than the rate of nonsynonymous substitution.

Phylogenetic analysis can also be used to address the question of whether the GSI and GSII genes were transferred between organisms during their long history. In agreement with Shatters and Khan (6) and Pesole *et al.* (45), the evolutionary tree shows that the common ancestor of the GSII genes for plants and bacteria occurred before the divergence of animals and plants (Fig. 1). We also note that

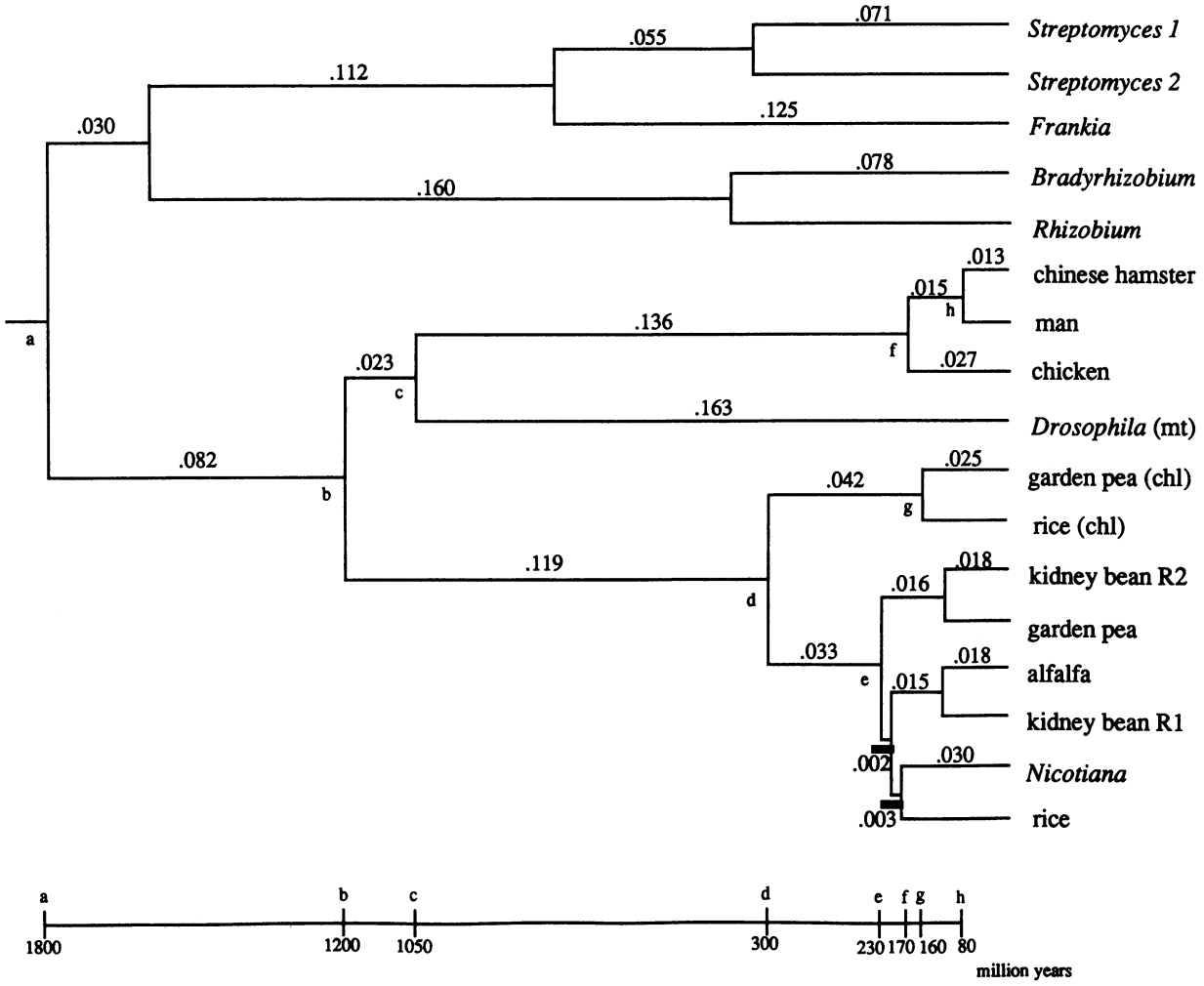

FIG. 2. Phylogenetic tree of GSII genes. Tree topology was taken from the tree in Fig. 1, and the branch lengths were recomputed on the basis of constant evolutionary rate. Branch length is given as the number of nonsynonymous substitutions per site. *Streptomyces 1* refers to *S. viridochromogenes* and *Streptomyces 2* to *S. hygroscopicus*. The letters a–g show branch points with corresponding divergence times in millions of years. Each of the two bold bars shows the standard error of the corresponding branch length. The upper bar is 0.0045 long and the lower bar 0.0043 long, which do not signify the two branch lengths.

the GSII gene of the plant symbiont *Frankia* is closer to those of the nonsymbiotic *Streptomyces* than to the plant genes or rhizobial genes in Fig. 1. These findings, in conjunction with a tandem arrangement of the GSI and GSII genes in *Frankia*, argue strongly against the proposed lateral gene transfer from host plant to bacterium (3).

It is intriguing that the two genes encoding the rice and garden pea chloroplast GSII enzymes are more closely related to each other than with their counterparts in the same species (Fig. 2), though each of the two genes is located in the nuclear genome (22–25). The common ancestor of the two genes diverged 300 million years ago by our estimate, which is earlier than the divergence of the monocot and the dicot as shown in Fig. 2. At least two explanations are possible for this observation. One is that one of the two genes was transferred through an interspecies barrier from one species to another about 160 million years ago (Fig. 2). Though the two species are completely separated at present, the interspecies barrier at that time might not have been as restrictive as it is at present. The other is that the two genes have undergone convergent evolution in the different species, because their products work in the similar environments (chloroplasts). The age of the gene encoding the mitochondrial GSII is similarly estimated as 1050 million years old. Our findings thus show that the genes encoding the organelle GSIIs diverged long after endosymbiosis, which is considered to have occurred about 2000 million years ago. In any event, our results do not support or deny endosymbiosis.

There is still an unfathomable gap between chemical evolution and biological evolution; the former deals with formation of primitive life in prebiotic environments and the latter at present is mostly concerned with evolution after prokaryote–eukaryote divergence. How did metabolic processes evolve in the period when primitive life became sufficiently vital (tentatively called the "preprokaryote") to evolve toward the prokaryotes and eukaryotes? The conditions under which evolution proceeded from the preprokaryote to a parent of the prokaryote and eukaryote remain a mystery. Our evidence indicating that duplication of the GSI and GSII genes preceded to a great extent the prokaryote–eukaryote divergence may trigger and stimulate studies on the evolution of the preprokaryotes, though our estimate of 1700 million years may not be quite accurate. During this period, it is probable that many other enzymes which could allow for nitrogen incorporation, including other forms of GS [perhaps represented by *Bacteroides fragilis* GS (31)], appeared and disappeared. The evolution of metabolic pathways in the era of the preprokaryote might have gradually refined the functions of GS so as to provide glutamine and glutamate and to disseminate nitrogen from glutamine into metabolic pathways leading to amino acids, nucleotides, and enzyme cofactors.

We thank Dr. K. Ikeo for helping us to use the software SINCAIDEN to compute the evolutionary distances and construct the phylogenetic trees.

1. Paolo, T.-B. & Cohen, G. N. (1973) *Annu. Rev. Biochem.* **42,** 113–134.
2. Tyler, B. (1978) *Annu. Rev. Biochem.* **47,** 1127–1162.
3. Carlson, T. A. & Chelm, B. K. (1986) *Nature (London)* **322,** 568–570.
4. Darrow, R. A. & Knotts, R. R. (1977) *Biochem. Biophys. Res. Commun.* **78,** 554–559.
5. Fuchs, R. L. & Keister, D. L. (1980) *J. Bacteriol.* **144,** 641–648.
6. Shatters, R. G. & Kahn, J. L. (1989) *J. Mol. Evol.* **2,** 422–428.
7. Rochefort, D. A. & Benson, D. R. (1990) *J. Bacteriol.* **172,** 5335–5342.
8. Behrmann, I., Hillemann, D., Pühler, A., Strauch, E. & Wohlleben, W. (1990) *J. Bacteriol.* **172,** 5326–5334.
9. Kumada, Y., Takano, E., Nagaoka, K. & Thompson, C. (1990) *J. Bacteriol.* **172,** 5343–5351.
10. Eisenberg, D., Almassy, R. J., Janson, C. A., Chapman, M. S., Suh, S. W., Cascio, D. & Smith, W. W. (1987) *Cold Spring Harbor Symp. Quant. Biol.* **52,** 483–490.
11. Possot, O., Sibold, L. & Aubert, J.-P. (1989) *Res. Microbiol.* **140,** 355–371.
12. Strauch, M. A., Aronson, A. I., Brown, S. W., Schreier, H. J. & Sonenshein, A. L. (1988) *Gene* **71,** 257–265.
13. Janssen, P. J., Jones, W. A., Jones, D. T. & Woods, D. R. (1988) *J. Bacteriol.* **170,** 400–408.
14. Tumer, N. E., Robinson, S. J. & Haselkorn, R. (1983) *Nature (London)* **306,** 337–343.
15. Rawlings, D. E., Jones, W. A., O'Neill, E. G. & Woods, D. R. (1987) *Gene* **53,** 211–217.
16. Colombo, G. & Villafranca, J. J. (1986) *J. Biol. Chem.* **261,** 10587–10591.
17. Janson, C. A., Kayne, P. S., Almassy, R. J., Grunstein, M. & Eisenberg, D. (1986) *Gene* **46,** 297–300.
18. Bozouklian, H. & Elmerich, C. (1986) *Biochimie* **68,** 1181–1187.
19. Colonna-Romano, S., Riccio, A., Guida, M., Defez, R., Lamberti, A., Iaccarino, M., Arnold, W., Priefer, U. & Pühler, A. (1987) *Nucleic Acids Res.* **15,** 1951–1964.
20. Wray, L. V., Jr., & Fisher, S. H. (1988) *Gene* **71,** 247–256.
21. Tischer, E., DasSarma, S. & Goodman, H. M. (1986) *Mol. Gen. Genet.* **203,** 221–229.
22. Gebhardt, C., Oliver, J. E., Forde, B. G., Saarelainen, R. & Miflin, B. J. (1986) *EMBO J.* **5,** 1429–1435.
23. Tingey, S. V., Tsai, F.-Y., Edwards, J. W., Walker, E. L. & Coruzzi, G. N. (1988) *J. Biol. Chem.* **263,** 9651–9657.
24. Tingey, S. V., Walker, E. L. & Coruzzi, G. M. (1987) *EMBO J.* **6,** 1–9.
25. Sakamoto, A., Ogawa, M., Masumura, T., Shibata, D., Takeba, G., Tanaka, K. & Fujii, S. (1989) *Plant Mol. Biol.* **13,** 611–614.
26. Tingey, S. V. & Coruzzi, G. M. (1987) *Plant Physiol.* **84,** 366–373.
27. Gibbs, C. S., Campbell, K. E. & Wilson, R. H. (1987) *Nucleic Acids Res.* **15,** 6293.
28. Hayward, B. E., Hussain, A., Wilson, R. H., Lyons, A., Woodcock, V., McIntosh, B. & Harris, T. J. R. (1985) *Nucleic Acids Res.* **14,** 999–1008.
29. Pu, H. & Young, A. P. (1989) *Gene* **81,** 169–175.
30. Caizzi, R., Bozzetti, M. P., Caggese, C. & Ritossa, F. (1990) *J. Mol. Biol.* **212,** 17–26.
31. Hill, R. T., Parker, J. R., Goodman, H. J. K., Jones, D. T. & Woods, D. R. (1989) *J. Gen. Microbiol.* **135,** 3271–3279.
32. Dickerson, E. R. & Geis, I. (1969) *The Structure and Action of Proteins* (Benjamin, Menlo Park, CA).
33. Nei, M. & Gojobori, T. (1986) *Mol. Biol. Evol.* **3,** 418–426.
34. Sneath, P. H. A. & Sokal, R. R. (1973) *Numerical Taxonomy* (Freeman, San Francisco).
35. Nei, M. (1975) *Molecular Population Genetics and Evolution* (North-Holland, Amsterdam).
36. Tateno, Y., Nei, M. & Tajima, F. (1982) *J. Mol. Evol.* **18,** 387–404.
37. Saitou, N. & Nei, M. (1987) *Mol. Biol. Evol.* **4,** 406–425.
38. Saitou, N. & Imanishi, T. (1989) *Mol. Biol. Evol.* **4,** 406–425.
39. Kimura, M. (1968) *Nature (London)* **217,** 624–626.
40. Kimura, M. (1983) *The Neutral Theory of Molecular Evolution* (Cambridge Univ. Press, Cambridge, U.K.).
41. Moriyama, N. E. & Gojobori, T. (1992) *Genetics* **130,** 855–864.
42. Dickerson, R. E. (1971) *J. Mol. Evol.* **1,** 26–45.
43. Sibley, C. G. & Alquist, J. E. (1990) *Phylogeny and Classification of Birds* (Yale Univ. Press, New Haven, CT).
44. Kimura, M. & Ohta, T. (1973) *Nature (London) New Biol.* **243,** 199–200.
45. Pesole, G., Bozzetti, P. M., Lanave, C., Preparata, G. & Saccone, C. (1991) *Proc. Natl. Acad. Sci. USA* **88,** 522–526.
46. Jin, L. & Nei, M. (1990) *Mol. Biol. Evol.* **7,** 82–102.
47. Schidlowski, M. (1988) *Nature (London)* **333,** 313–318.
48. Nei, M., Stephens, J. C. & Saitou, N. (1985) *Mol. Biol. Evol.* **2,** 66–85.