Proc. Natl. Acad. Sci. USA
Vol. 88, pp. 522–526, January 1991
Evolution

Glutamine synthetase gene evolution: A good molecular clock

(phylogenetic trees/Markov process/organelle enzymes)

G. PESOLE*, M. P. BOZZETTI†, C. LANAVE*, G. PREPARATA‡, AND C. SACCONE*§

*Centro Studi sui Mitocondri e Metabolismo Energetico, Consiglio Nazionale delle Ricerche, presso Dipartimento di Biochimica e Biologia Molecolare, and †Istituto di Genetica, Universitá di Bari, Bari, Italy; and ‡Dipartimento di Fisica, Universitá di Milano, Milan, Italy

Communicated by Lynn Margulis, June 8, 1990 (received for review July 17, 1989)

ABSTRACT **Glutamine synthetase (EC 6.3.1.2) gene evolution in various animals, plants, and bacteria was evaluated by a general stationary Markov model. The evolutionary process proved to be unexpectedly regular even for a time span as long as that between the divergence of prokaryotes from eukaryotes. This enabled us to draw phylogenetic trees for species whose phylogeny cannot be easily reconstructed from the fossil record. Our calculation of the times of divergence of the various organelle-specific enzymes led us to hypothesize that the pea and bean chloroplast genes for these enzymes originated from the duplication of nuclear genes as a result of the different metabolic needs of the various species. Our data indicate that the duplication of plastid glutamine synthetase genes occurred long after the endosymbiotic events that produced the organelles themselves.**

Though the existence of a "molecular clock" is now widely accepted, the use of biological macromolecules to arrange our knowledge of biological events within a temporal framework remains a very serious problem. To perform quantitative estimates of evolutionary processes, well-defined methods for determining the precise correlation between the time of evolution and the divergence of the genetic material are needed. Methods based on deterministic or stochastic models of molecular evolution (see ref. 1 for a general survey) suffer from being based upon *a priori* assumptions, which in most cases have been found to be experimentally untenable, and have thus led to unreliable estimates.

We have proposed a stochastic model of gene evolution, the "stationary Markov clock" (2, 3), in which only the homologous genes that fulfill the "stationarity" condition—i.e., the condition of having the same base frequencies in sites (such as the first, second, or silent codon positions) that are presumably subjected to the same kind of dynamics—can behave as reliable clocks. Although the probability of maintaining the same base frequency in equivalent sites is theoretically extremely low, "DNA Markov clocks" appear in nature with a frequency that *a priori* would have seemed unreasonable (4–7).

The existence of a Markov clock is related, according to our hypothesis, to the genetic distance between homologous genes and thus depends on both the rate of the evolutionary process and the time of divergence of the sequences under comparison. This provides us with an important criterion with which to identify the types of clock (i.e., those based on first, second, or third silent codon positions) that are most appropriate for tracking the evolution of the species; the slower the gene evolutionary process, the broader the applicability of our model (2).

Although in general our method appears to apply better to species that do not have markedly long times of divergence, we here present evidence that genes encoding glutamine synthetase [GS; L-glutamate:ammonia ligase (ADP-forming), EC 6.3.1.2] behave as perfect Markov clocks for times of divergence as long as that between eukaryotes and prokaryotes. This allowed us to estimate the phylogenetic distance between species whose phylogeny cannot be easily reconstructed from fossil records, and provided important clues to the origin of organelle-specific enzymes.

MATERIALS AND METHODS

We analyzed 17 GS sequences: 3 mammalian (human, hamster, and rat); 2 *Drosophila melanogaster* (the cytosolic and mitochondrial isoforms; ref. 8); 7 plant [alfalfa (*Medicago sativa*), 1 isoform; pea (*Pisum sativum*), 3 isoforms; and bean (*Phaseolus vulgaris*), 3 isoforms]; and 5 prokaryotic (*Bradyrhizobium japonicum*, *Anabaena*, *Salmonella typhimurium*, *Thiobacillus ferrooxidans*, and *Escherichia coli*). All the GS nucleotide sequences were taken from Release 17 of the EMBL database (where the references for the sequences can be found) by using the retrieval program ACNUC (9). The multiple alignment of sequences was achieved on an iterative basis, starting from paired alignments, using the program BESTFIT from the package GLORIA (unpublished work). The aligned sequences were analyzed according to our stationary Markov model (2, 3).

RESULTS

Sequence Alignment. Fig. 1 shows the best alignment of the 17 GS amino acid sequences. The only case of alignment between eukaryotic and prokaryotic sequences of which we are aware is that reported by Tischer *et al.* (10). The alignment we have achieved produces an overall similarity that is consistently greater (18–20% sequence identity) than that reported by Tischer *et al.* It also reveals additional highly conserved regions that are probably involved in the active site or in regulatory mechanisms.

The protein structure of GS from *S. typhimurium* was extensively characterized (11, 12) and two principal domains, C and N, were identified. This enzyme is made up of 12 identical subunits, arranged in two layers of 6. The active site, marked by a pair of Mn^{2+} ions, appears to be formed by two antiparallel β structures, one in the C domain of one subunit and the other in the N domain of the neighboring subunit. From our multiple alignment, five regions that are highly conserved from prokaryotes to eukaryotes can be identified; the first [amino acids (aa) 116–159] is located in the active site of the N domain, while the second (aa 273–290), third (aa 328–342), fourth (aa 414–422), and fifth (aa 435–453) are in the active site of the C domain. It has been suggested that the fifth region (aa 435–453) contains an ATP binding site. Six amino acid residues were recognized as ligands to

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.

Abbreviations: GS, glutamine synthetase; Myr, million years.

§To whom reprint requests should be addressed at: Dipartimento di Biochimica e Biologia Molecolare, Universitá di Bari, via Amendola 165/A-70126, Bari, Italy.

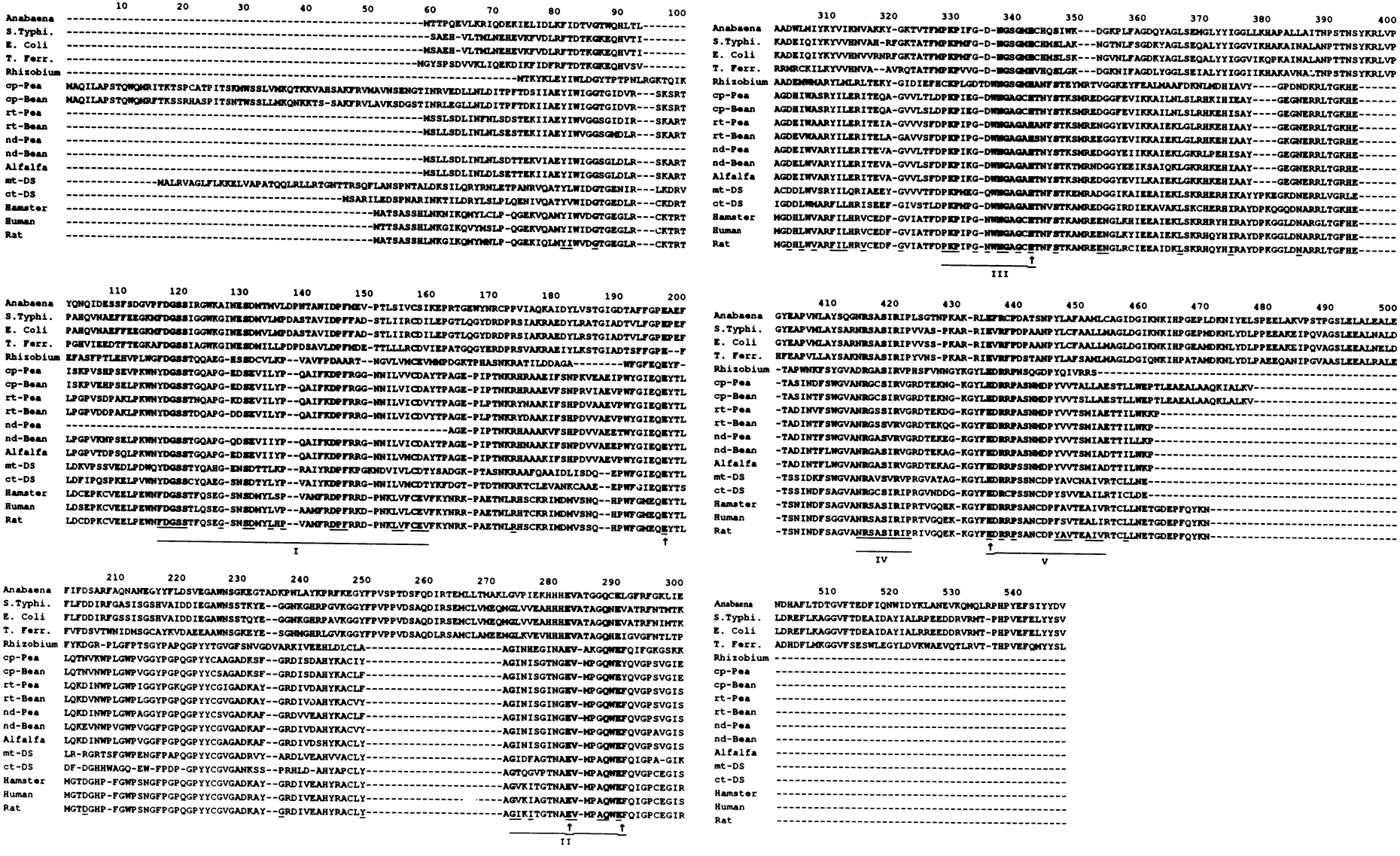

FIG. 1. Amino acid alignment of the 17 sequences considered in this paper. We have underlined those sites that present an amino acid similarity between prokaryotes and eukaryotes. Amino acids that are the same for all 17 are indicated in bold type. The five highly conserved regions are underlined and labeled with Roman numerals. Five of the six amino acid residues recognized as ligands to the Mn^{2+} ions are indicated by arrows. cp, Chloroplast; rt, root; nd, nodule; mt, mitochondrial; ct, cytoplasmic; DS, *Drosophila*.

the Mn^{2+} ion and five of these are conserved in all 17 sequences compared. This suggests a similar structure in the active site of both the prokaryotic and eukaryotic enzymes, the latter being formed by eight identical subunits. In total, 23 amino acid sites remain unchanged in all the sequences compared.

The GS proteins imported into organelles, be they mitochondria or chloroplasts, display an N-terminal extension (transit peptide) as do the majority of proteins destined to be imported. No significant similarity is found between chloroplast and mitochondrial transit peptides at either the amino acid or the nucleotide level.

Stationarity Check. According to our stationary Markov model, only those genes which fulfill the stationarity condition behave as reliable molecular clocks and can thus be used for quantitative measurements of genetic distances between species (2–7). The stationarity check was carried out with a χ^2 test as described (2), grouping together all first, second, or third (silent) codon positions. Surprisingly, the check revealed that stationarity was maintained in all the sequences considered at the level of the second codon positions.

At the other codon positions, stationarity was maintained only within each of the three groups as follows. Animal genes: first codon positions stationary in all species; third codon positions stationary in human and hamster (not in rat). Plant genes: first and third codon positions stationary in all species. Bacterial genes: first codon positions stationary in all species except *B. japonicum*; third codon positions nonstationary for all the species (only the most closely related organisms, *E. coli* and *S. typhimurium*, show a marginal stationarity).

Phylogenetic Distances and Base Substitution Rates. Our method (2, 3) allows us to estimate the time-of-divergence ratio (T/T') between any two pairs of sequences. This means that once the most reliable time of divergence between two species as determined from other sources (i.e., paleontological data) is fixed as input, all other T values can be calculated. In Fig. 2 the T/T' ratios are reported; they were calculated by analyzing the stationary codon positions (first, second, or third silent) for all the possible pairs of sequences. The phylogenetic trees shown in Fig. 3 were obtained by fixing the time of divergence between vertebrates and invertebrates at 600 million years (Myr) (13). The ultrametric nature of all the measurements of the T/T' ratios underlines the reliability of this tree topology.

The split between prokaryotes and eukaryotes comes out at 2500 ± 500 Myr, that between plants and animals at 900 ± 200 Myr (Fig. 3*A*) and that between human and rodents at 75 ± 25 Myr, estimates supported by paleontological and molecular evidence (13).

The mitochondrially imported *Drosophila* GS protein appears to have diverged from the cytosolic form at about the same time as the split between vertebrates and invertebrates took place. This implies that the organelle-specific enzyme originated from a single gene at a time around the vertebrate/invertebrate split (Fig. 3*B*).

The phylogenetic tree for plant GS genes is shown in Fig. 3*C*. For both pea and bean, three paralogous protein-coding genes, expressed in nodules, roots, and chloroplasts, were

Hamster	Human	mt-DS	ct-DS	Alfalfa	nd-BEAN	rt-BEAN	cp-BEAN	nd-PEA	rt-PEA	cp-PEA	
0.06±0.02 0.09±0.03 -	0.09±0.03 0.12±0.03 -	0.92±0.19 1.03±0.17	0.97±0.21 0.95±0.16 -	- 1.28±0.20 -	- 1.21±0.19 -	- 1.22±0.19 -	- 1.37±0.22 -	- 1.11±0.19 -	- 1.27±0.22 -	- 1.42±0.21 -	Rat
	0.14±0.07 0.12±0.09 -	1.00±0.26 1.03±0.31 -	1.01±0.26 0.99±0.29 -	- 1.31±0.33 -	- 1.24±0.31 -	- 1.24±0.32 -	- 1.38±0.35 -	- 1.16±0.32 -	- 1.27±0.32 -	- 1.44±0.36 -	Hamster
		1 1 -	1.01±0.26 1.00±0.28 -	- 1.31±0.33 -	- 1.24±0.31 -	- 1.25±0.32 -	- 1.40±0.35 -	- 1.16±0.32 -	- 1.27±0.32 -	- 1.46±0.36 -	Human
			1.02±0.25 1.07±0.31 -	- 1.37±0.34 -	- 1.35±0.34 -	- 1.42±0.36 -	- 1.72±0.42 -	- 1.23±0.34 -	- 1.39±0.35 -	- 1.65±0.40 -	mt-DS
S. Typ.	0.05±0.01 0.04±0.01 -			- 1.52±0.39 -	- 1.49±0.38 -	- 1.51±0.39 -	- 1.58±0.39 -	- 1.39±0.39 -	- 1.49±0.38 -	- 1.57±0.40 -	ct-DS
T. Fer.	0.80±0.12 0.80±0.14 -	0.79±0.11 0.74±0.14 -			0.17±0.07 0.12±0.08 0.15±0.05	0.20±0.08 0.20±0.10 0.26±0.08	0.39±0.12 0.46±0.17 0.48±0.15	0.14±0.05 0.24±0.12 0.09±0.04	0.23±0.08 0.29±0.13 0.27±0.08	0.42 0.42±0.15 0.42	Alfalfa
Anab.	1.22 1.22±0.20 -	1.30±0.17 1.17±0.19 -	1.20±0.17 1.20±0.20 -			0.23±0.08 0.20±0.10 0.31±0.09	0.40±0.08 0.46±0.17 0.46±0.14	0.24±0.10 0.23±0.12 0.17±0.06	0.26±0.09 0.27±0.12 0.30±0.09	0.43±0.13 0.45±0.16 0.42±0.14	nd-BEAN
Rhizob.	- 4.28±0.75 -	- 4.21±0.75 -	- 4.63±0.85 -	- 4.39±0.79 -			0.42±0.06 0.39±0.15 0.42±0.13	0.24±0.10 0.27±0.13 0.31±0.11	0.15±0.06 0.16±0.10 0.19±0.11	0.45±0.14 0.41±0.15 0.36±0.11	rt-BEAN
Eukaryot.	- 4.67±0.90 -	- 4.66±0.89 -	- 4.65±0.88 -	- 4.25±0.79 -	- 1.94±0.33 -			0.36±0.13 0.46±0.18 0.46±0.15	0.44±0.14 0.40±0.14 0.45±0.14	0.14±0.06 0.23±0.11 0.12±0.04	cp-BEAN
	E. Coli	S. Typ.	T. Fer.	Anab.	Rhizob.				0.24±0.10 0.26±0.14 0.30±0.10	0.37±0.13 0.48±0.19 0.50±0.17	nd-PEA
										0.46±0.14 0.42±0.15 0.39±0.12	rt-PEA

FIG. 2. Ratios of divergence times (T/T') between the 17 GS genes. (*Upper*) Eukaryotes. (*Lower*) Prokaryotes. For eukaryotes the average values are reported. Abbreviations are as in Fig. 1.

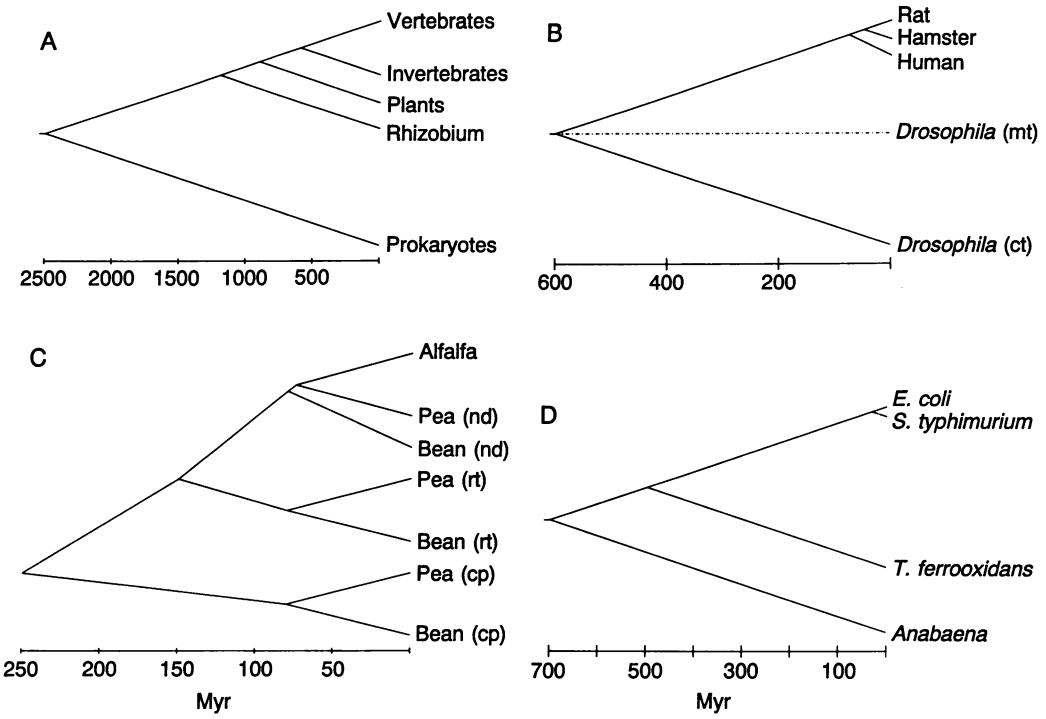

FIG. 3. Phylogenetic trees. (*A*) Animals, plants, and bacteria. (*B*) Animals (broken line refers to the *Drosophila* mitochondrial isoform of GS). (*C*) Plants. (*D*) Bacteria. Abbreviations are as in Fig. 1.

analyzed. All three comparisons between pea and bean organ-specific proteins show the same divergence time within their statistical fluctuations, and thus each of the three pairs of genes can be considered orthologous. The two cytoplasmic isoforms expressed in nodules and roots were found to be more closely related to each other than to the isoforms coding for enzymes imported into chloroplasts. The alfalfa gene is clearly more closely related to the nodule isoform.

The phylogenetic tree for bacteria (Fig. 3*D*) shows an approximate time of divergence of 30 ± 10 Myr for the two most closely related species, *S. typhimurium* and *E. coli*, which is rather different from that (120–160 Myr) indicated by Ochman and Wilson (14) but relatively close to that (37 ± 26 Myr) estimated by Hori and Osawa (15).

The average substitution rate calculated for the second codon position is 2.5×10^{-10} per site per year. The silent substitution rates for plants and animals are similar, on the average 2.1×10^{-9}, a value that agrees with the estimate of Martin *et al.* (16) based on the evolutionary analysis of glyceraldehyde-3-phosphate dehydrogenase genes.

The silent substitution rate between *S. typhimurium* and *E. coli* was also calculated for the third codon positions, even though the stationarity in this case is barely significant. However, we believe that the high value obtained (7.9 ± 2.2) can be explained only partially by the marginal stationarity since it is also much higher than that between human and rat (3.3 ± 1.0), which are clearly nonstationary. This suggests that silent substitution rates can be higher in prokaryotes than in eukaryotes (plants and animals). Other analyses are necessary, however, before we can draw definite conclusions.

Origin of *B. japonicum* GS II. We compared the *B. japonicum* gene for GS II with the homologous genes of animals, plants, and bacteria. The protein coded by this gene forms an octameric enzyme, like those found in animals and plants but different from the dodecameric form usually found in bacteria. The greater similarity with plant GS genes and the fact that to date all the bacteria known to contain the GS II gene (such as species of *Rhizobium* and *Agrobacterium*) are involved in symbiotic relationships with plants led to the conclusion that at one stage a gene transfer from plants to a symbiotic bacterium took place (17). However, in the light of our evolutionary analysis, carried out at the level of the second codon positions, the *B. japonicum* GS II gene was equally divergent from both plant and animal GS genes. A time of divergence of 1200 ± 200 Myr, which comfortably precedes the split between plants and animals, was calculated. This cannot be explained by a different replacement rate in bacteria with respect to the other species, on account of the pronounced ultrametric nature of all the calculated evolutionary distances (see Fig. 2), further confirming the excellent clock-like behavior of GS genes. Our data agree with the recent results reported by Shatters and Kahn (18).

Directional Mutation Pressure in GS Genes. In an attempt to understand the reasons for the exceptionally good clock-like behavior of GS genes, we measured the extent of directional mutation pressure on their evolution. We plotted the G+C content of third codon positions (P3) against that of the second codon position (P2), an analysis which, according to Sueoka (19), determines the extent of neutrality in gene evolution. Despite the high variability of the G+C content in P3 in all the compared genes, the G+C content of the P2 remained unchanged, indicating that all the asynonymous sites are under total selective constraint and, therefore, are not affected by the directional mutation pressure that determines their P3 value.

DISCUSSION

All analyses of molecular evolution are obviously model-dependent, but the really important point is the consistency of the analyses—i.e., the coherence between the actual observable data and the models that are used in analyzing them. The main feature of our model is that it does not impose any *a priori* condition on the stochastic structure of the stationary process of nucleotide substitution.

In the analysis of GS evolution all the second codon positions fulfill the stationarity condition. The coherence between the model and the data is supported by the observed ultrametric nature of all the estimated evolutionary distances.

rRNA genes are usually considered the most suitable molecular tools for tracing the evolution of species. Due to their basic function, they are well conserved in all living organisms. However, a number of parameters that characterize these molecules make them unreliable for quantitative

measurements of genetic distances (20). The best alignment is often difficult to achieve on account of the variable length of the molecules; the need to preserve stem and loop secondary structures implies simultaneous changes of two bases and thus a large number of convergent substitutions. Moreover, since the functions of the various domains are not yet completely understood, it is extremely difficult to isolate the different dynamics operating along the sequences. Indeed, spatial variations in base substitution rates are one of the main sources of irregularity in molecular clocks.

One of the most important prerequisites for a molecular clock is the grouping of sequence sites that are presumably subject to the same dynamics, although in theory each site can evolve separately (5). In this respect, protein-coding genes are much more manageable, since one can roughly distinguish three different dynamics, for the first, second, and silent codon positions respectively, each having a reasonable degree of homogeneity. In our method, each dynamic can be used in tracing evolution over a given time span, depending on the distance between species (2).

However, the evolutionary analysis of protein-coding genes may also become tricky, particularly for highly divergent sequences, because of difficulties in alignment, nonfulfillment of the stationarity condition, etc. The perfect clocklike behavior of GS genes in highly divergent organisms was thus a great surprise.

GS proteins are ubiquitous, well-conserved enzymes, but this property is not sufficient to explain their perfect clocklike behavior since the genes for other ubiquitous, well-conserved proteins do not behave as the genes for GS do. The base compositions of GS second codon positions do not follow those of the third codon positions. This, according to Sueoka (19), should indicate the absence of neutrality in the second codon positions. If this is the case, we must argue that the "clock" is to be found in the most constrained positions instead of in the neutral ones. This supports Zuckerkandl's idea that the "best clocks should be obtained with sufficiently large sets of second codon positions," which he defines as "the least affected by regional or evolutionary changes" (21).

We now want to take a closer look at certain specific relationships that this study has brought to light.

The analyses of GS evolution in plants appear to be extremely useful in establishing the phylogeny of dicots whose reconstruction is either difficult or highly debatable. Our conclusions about plants are based on several different orthologous genes and on the dynamics of all three codon positions. Regardless of which of the three codon positions we study, we arrive at similar estimates concerning the dating of plant evolutionary events. According to our data, the two cytoplasmic isoforms expressed in nodules and roots originated about 150 Myr ago. Our findings indicate that the chloroplast genes arose 250 Myr ago, long after the endosymbiotic event—dated between 800 and 2000 Myr—in which a cyanobacterium colonized an ancestral nonphotosynthetic eukaryotic cell. In the case of animals, a similar phenomenon is observed: the *Drosophila* mitochondrial gene split from the cytoplasmic isoenzyme at about the same time—600 Myr ago—as the split between vertebrates and invertebrates occurred.

These results indicate that organelle-specific enzymes may have originated from a duplication of nuclear genes. The endosymbiotic hypothesis suggests that a transfer of prokaryotic genes to nuclei occurred during the evolution of the primitive eukaryotic cell. In some cases, it is very likely that the old prokaryotic gene could not be active in the new nuclear genome environment or was totally lost because its function in the organelle could be dispensed with. Subsequently, a new organelle-specific enzyme could have been born to serve specialized metabolic functions.

The presence of GS enzymes in mitrochondria is linked to the nitrogen metabolism of the species (22), and in particular to the need for glutamine as a source of ammonia and for particular biochemical pathways for ammonia detoxification. In ureotelic vertebrates, such as mammals, intramitochondrially generated ammonia is converted to citrulline by the combined action of carbamoyl-phosphate synthetase I and ornithine carbamoyltransferase. In uricotelic organisms, such as insects and birds, the ammonia generated in the mitochondria is converted to glutamine by mitochondrial GS and then the glutamine exits to the cytoplasm, where it is converted into uric acid.

The *B. japonicum* GS II gene has a closer relationship with the eukaryotic genes than it has with prokaryotic genes; this can be explained by three hypotheses. (*i*) A lateral gene transfer from a protoeukaryote to a protobacterium occurred about 1200 Myr ago, before the animal/plant split 300 Myr later. This means that the protoeukaryote host probably had a nitrogen-fixing symbiont. That this transfer took place many millions of years ago is further confirmed by the identification of a GS II gene in the nitrogen-fixing symbiont *Frankia* (23), which is only remotely related to *Rhizobiaceae* and *Agrobacteriaceae*. All extant bacteria that are descendants of that protobacterium may retain the GS II gene as a result of their specific metabolic requirements. (*ii*) A lateral gene transfer from plants to prokaryotes (suggested in ref. 18) can only be explained if the protoeukaryote, from which both plants and animals originated, had two paralogous GS genes. During the evolutionary process that led to current plant and animal species, one of these two isoforms (GS II) was transferred from plants to a symbiotic bacterium; however, both plants and animals have now lost the GS II gene. (*iii*) The GS II gene, of prokaryotic origin, was transferred to a protoeukaryote from an endosymbiotic purple bacterium, the protomitochondrion (18). If this hypothesis were valid, the prokaryote, presumably protomitochondrion, would have had two GS isoforms. To clarify the situation, additional sequences and appropriate evolutionary analyses are necessary.

This work was partially financed by the Ministero Pubblica Istruzione and by the Progetto Strategico Genoma Umano, Consiglio Nazionale delle Ricerche, Italy.

1. Jukes, T. H., ed. (1987) *J. Mol. Evol.* **26**, 1–164.
2. Preparata, G. & Saccone, C. (1987) *J. Mol. Evol.* **26**, 7–15.
3. Saccone, C., Lanave, C., Pesole, G. & Preparata, G. (1990) *Methods Enzymol.* **183**, 570–583.
4. Holmes, E. C., Pesole, G. & Saccone, C. (1990) *J. Hum. Evol.* (1989) **18**, 775–794.
5. Lanave, C., Tommasi, S., Preparata, G. & Saccone, C. (1986) *BioSystems* **19**, 273–283.
6. Lanave, C., Preparata, G. & Saccone, C. (1985) *J. Mol. Evol.* **21**, 346–350.
7. Saccone, C., Pesole, G. & Preparata, G. (1989) *J. Mol. Evol.* **29**, 407–411.
8. Caizzi, R., Bozzetti, M. P., Caggese, C. & Ritossa, F. (1990) *J. Mol. Biol.* **212**, 17–26.
9. Gouy, M., Gautier, C., Attimonelli, M., Lanave, C. & Di Paola, G. (1985) *CABIOS* **1**, 167–172.
10. Tischer, E., DasSarma, S. & Goodman, H. M. (1986) *Mol. Gen. Genet.* **203**, 221–229.
11. Almassy, R. J., Janson, C. A., Hamlin, R., Xuong, N.-H. & Eisemberg, D. (1986) *Nature (London)* **323**, 304–309.
12. Yamashita, M. M., Almassy, R. J., Janson, C. A., Cascio, D. & Eisemberg, D. (1989) *J. Biol. Chem.* **264**, 17681–17690.
13. Nei, M. (1987) in *Molecular Evolutionary Genetics*, ed. Nei, M. (Columbia Univ. Press, New York), pp. 8–18.
14. Ochman, H. & Wilson, A. C. (1987) *J. Mol. Evol.* **26**, 74–86.
15. Hori, H. & Osawa, S. (1978) *J. Bacteriol.* **133**, 1089–1095.
16. Martin, W., Gierl, A. & Saedler, H. (1989) *Nature (London)* **339**, 46–48.
17. Carlson, T. A. & Chelm, B. K. (1986) *Nature (London)* **322**, 568–570.
18. Shatters, R. G. & Kahn, M. L. (1989) *J. Mol. Evol.* **29**, 422–428.
19. Sueoka, N. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 2653–2657.
20. Rothschild, L. J., Ragan, M. A., Coleman, A. W., Heywood, P. & Gerbi, S. A. (1986) *Cell* **47**, 640.
21. Zukerkandl, E. (1987) *J. Mol. Evol.* **26**, 34–46.
22. Mommsen, T. P. & Walsh, P. J. (1989) *Science* **243**, 72–74.
23. Edmands, J., Noridge, N. A. & Benson, D. N. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 6126–6130.