8. L. Machta, in *The Changing Chemistry of the Oceans*, D. Dyrssen and D. Jagner, Eds. (Almqvist & Wiksell, Stockholm, 1972), pp. 121–145.
9. K. E. Zimen and F. K. Altenhein, *Z. Naturforsch. Teil A* **28**, 1747 (1973).
10. R. Bacastow and C. D. Keeling, in (*11*), pp. 86–135.
11. G. M. Woodwell and E. V. Pecan, Eds., *Carbon and the Biosphere* (Atomic Energy Commission, Technical Information Service, Washington, D.C., 1973).
12. C. D. Keeling, in *Energy and Climate: Outer Limits to Growth* (National Academy of Sciences, Washington, D.C., in press).
13. Data and models on the global carbon cycle have been summarized by Baes *et al.* (*4*), and by Bolin (*14*).
14. B. Bolin, *Sci. Am.* **223**, 131 (September 1970); in *The Physical Basis of Climate and Climate Modelling* (GARP Publ. Ser. No. 16, World Meteorological Organization, Geneva, 1975), pp. 225–235.
15. R. M. Garrels and E. A. Perry, in *The Sea*, E. G. Goldberg, Ed. (Wiley, New York, 1974), vol. 5, pp. 303–306.
16. For a summary of work on ^{14}C in the oceans see M. Stuiver, in (*11*), pp. 6–20.
17. C. D. Keeling, in *Chemistry of the Lower Atmosphere*, S. I. Rasool, Ed. (Plenum, New York, 1973), pp. 251–329.
18. K. Wyrtki, *Deep-Sea Res.* **9**, 11 (1962); W. H. Munk, *ibid.* **13**, 707 (1966); J. Craig, *J. Geophys. Res.* **74**, 5491 (1969).
19. H. Oeschger, U. Siegenthaler, U. Schotterer, A. Gugelmann, *Tellus* **27**, 168 (1975).
20. The natural ^{14}C variations, as observed in tree rings, seem to be larger than the ^{14}C variations calculated from models, including the BD model, under the assumption that they are due to varying ^{14}C production due to changing solar activity as reflected in sunspot numbers. However, this is not necessarily due to wrong model results for the attenuation factors, because the relationship between sunspot numbers and changes in ^{14}C production is not firmly established. Furthermore, there may be other reasons for ^{14}C variations, such as fluctuations in ocean mixing.
21. W. S. Broecker and T.-H. Peng, *Tellus* **26**, 21 (1974).
22. A. Rebello and K. Wagener, paper presented at the 4th International Symposium on the Chemistry of the Mediterranean, Rovinj, Yugoslavia, 1976.
23. W. A. Reiners, in (*11*), p. 303. The high carbon content of the terrestrial biosphere is obtained from an estimate of total nitrogen in organic detritus and a C/N ratio of 12.
24. H. Lieth, *J. Geophys. Res.* **68**, 3887 (1963).
25. G. W. Woodwell and R. A. Houghton, in *Dahlem Workshop on Global Chemical Cycles and Their Alternation by Man*, W. Stumm, Ed. (Abakon, Berlin, 1977), pp. 61–72.
26. B. Bolin, *Science* **196**, 613 (1977).
27. H. D. Freyer and L. Wiesberg, in *Isotope Ratios as Pollutant Source and Behaviour Indicators* (International Atomic Energy Agency, Vienna, 1975), p. 49; M. Stuiver, personal communication.
28. W. S. Broecker, *Chemical Oceanography* (Harcourt Brace Jovanovich, New York, 1974).
29. W. Stumm, in *The Changing Chemistry of the Ocean*, D. Dyrssen and D. Jagner, Eds. (Almqvist & Wiksell, Stockholm, 1972), pp. 329–346.
30. C. D. Keeling, R. B. Bacastow, A. E. Bainbridge, C. A. Ekdahl, P. R. Guenther, L. S. Waterman, J. F. S. Chin, *Tellus* **28**, 538 (1976); C. D. Keeling, J. A. Adams, C. A. Ekdahl, P. R. Guenther, *ibid.*, p. 552; C. D. Keeling, private communication. Note that the corrected scale ("1974 manometric mole fraction scale") used in these publications and also in this article is slightly different from the scale used formerly by Keeling and co-workers. The value for the end of 1971 is 326.5 ppm (new scale) compared to 322.5 ppm (old scale).
31. Data up to 1970 are from C. D. Keeling [*Tellus* **25**, 174 (1973)]; data for 1971 to 1974 are from R. M. Rotty (*34*); and the 1975 value (2.40 ppm/year) was estimated by Keeling (*12*).
32. G. S. Callendar, *Tellus* **10**, 243 (1958); J. R. Bray, *ibid.* **11**, 220 (1959).
33. The production functions leading to the prescribed CO_2 excess were computed by leaving the production rate constant for 1 year, then comparing the actual atmospheric concentration with the concentration aimed at (given in Fig. 6), correcting the production rate accordingly, continuing the procedure for one more year, and so on.
34. R. M. Rotty, *Global Carbon Dioxide Production from Fossil Fuel and Cement, A.D. 1950–A.D. 2000* (Institute for Energy Analysis Research Memorandum [IEA(M)-76-4], Oak Ridge Associated Universities, Oak Ridge, Tenn., 1976).
35. W. S. Broecker, in (*11*), pp. 32–50.
36. P. Möller and P. P. Parekh, *Sci. Total Environ.* **4**, 177 (1975).
37. R. M. Garrels and F. T. Mackenzie, *Evolution of Sedimentary Rocks* (Norton, New York, 1971), chaps. 5 and 6.
38. E. Eriksson, *J. Geophys. Res.* **68**, 3871 (1963).
39. R. H. Whittaker and G. E. Likens, in (*11*), pp. 281–302.
40. Thanks are due to M. Stuiver for stimulating discussions, and to W. S. Broecker and another reviewer, unknown to us, for valuable comments on this article. This work is financially supported by the Swiss National Science Foundation.

Origins of Prokaryotes, Eukaryotes, Mitochondria, and Chloroplasts

A perspective is derived from protein and nucleic acid sequence data.

Robert M. Schwartz and Margaret O. Dayhoff

Many proteins and nucleic acids are "living fossils" in the sense that their structures have been dynamically conserved by the evolutionary process over billions of years (*1*). Their amino acid and nucleotide sequences occur today as recognizably related forms in eukaryotes and prokaryotes, having evolved from common ancestral sequences by a great number of small changes (*2*). These sequences may still carry sufficient information for us to unravel the early evolution of extant biological species and their biochemical processes. There are two principal computer methods that can be used to treat sequence data in order to elucidate evolutionary history. These were first described more than 10 years ago (*3, 4*) and have been used to construct a vertebrate phylogeny from each of a number of proteins (*5–7*). This phylogeny is generally consistent with both the fossil record and morphological data. Only recently has enough sequence information become available from diverse types of bacteria and blue-green algae, and from the cytoplasm and organelles of eukaryotes, for us to attempt the construction of a biologically comprehensive evolutionary tree. These sequences include ferredoxins, 5*S* ribosomal RNA's, and c-type cytochromes. We describe here an evolutionary tree derived from sequence data that extends back close to the time of the earliest divergences of the present-day bacterial groups.

Knowledge of the evolutionary relationships between all species would have great predictive advantage in many areas of biology, because most systems within the organisms would show a high degree of correlation with the phylogeny. Knowing the relative order of the divergences of prokaryote types and their protein constituents is important to understanding the evolution of metabolic pathways. With such information the long-standing question of how eukaryote organelles originated might be resolved.

Before we consider the phylogenies based on sequence data, we will briefly review the fossil record and describe the time scale during which the various prokaryote groups diverged (*8, 9*). The early fossil record is sparse and subject to some uncertainty of interpretation. The oldest known bacterium-like structures that could possibly be biogenic are preserved in the Swaziland sediments and are more than 3.1 billion years old (*8,*

Dr. Schwartz is a senior research scientist at the National Biomedical Research Foundation. Dr. Dayhoff is an associate professor in the Department of Physiology and Biophysics at Georgetown University Medical School and Associate Director of Research at the National Biomedical Research Foundation at Georgetown University Medical Center, Washington, D.C. 20007.

0036-8075/78/0127-0395$02.00/0

10). Stromatolitic structures have been dated at nearly 3.0 billion years old (*11*). These structures are widely assumed to be evidence for the antiquity of blue-green algae but might equally well be interpreted as products of communities of photosynthetic bacteria, gliding flexibacteria, or non-oxygen-releasing ancestors of the blue-green algae (*8*). Both coccoid and filamentous microfossils from 2.3 billion years ago have been identified. The fossil record of microorganisms from about 2 billion years ago clearly shows a great abundance and diversity of morphological types, many resembling present-day bacteria and blue-green algae. Geological evidence suggests that free oxygen began to accumulate in the atmosphere about 2.0 billion years ago (*12*). Eukaryote cells may have originated as early as 1.4 billion years ago, because there is an abrupt increase in cell size and diversity in microfossils of this age (*13*). The multicellular eukaryote kingdoms, plants, animals, and fungi, are thought to have diverged between 1 billion and 700 million years ago. Fossils of metazoans nearly 700 million years old have been found.

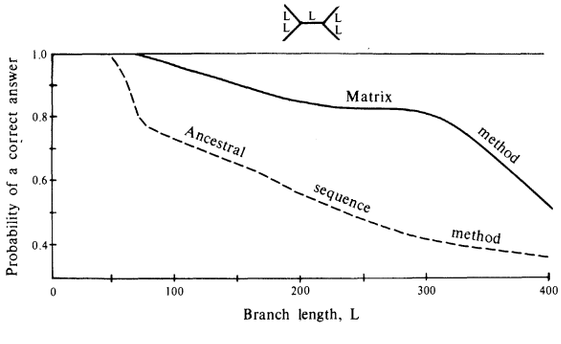
There is good evidence that, in many bacterial sequences of basic metabolic importance, the rate of accumulation of point mutations accepted in the wild-type population is even slower than it is in higher plants and animals (*14*). Thus, even though the time span for the evolution of prokaryotes is more than two times as long, we hope to infer correct evolutionary trees for them as well as we do for eukaryotes. Because the morphological evidence of biologists has proved to be inadequate to the task of organizing the major groups of bacteria and because the fossil record is difficult to interpret, sequence data may prove to be essential.

In this article, we assume that the major types of bacteria have conserved the integrity of the groups of sequences performing basic metabolic functions; we also assume that the substitution of a new sequence for one already functioning in a group through genetic transfer is sufficiently rare to be discounted. Frequent transfer between closely related species should not impair our ability to deduce the course of evolution of the major bacterial types. Only sequences that were transferred will lead to conflicting evolutionary histories for the species involved; sequences from any of the close species would be equally useful in deducing the evolutionary position of the bacterial type.

Reconstructing Evolution on the Basis of Sequence Data

Evolutionary history can be conveniently represented by a tree on which each point corresponds to a time, a macromolecular sequence, and a species within which the sequence occurred. Although we may not yet be able to infer its exact location, there is one point that corresponds to the earliest time and the original ancestral sequence and species. Time advances on all branches of the tree emanating from this point. During evolution, sequences in different species have gradually and independently accumulated changes, yielding the sequences found today in the extant species represented at the ends of the branches. The absolute chronology, of course, cannot be inferred from sequences; the topology of the branches gives the relative order of events. On the evolutionary trees in this article the lengths of the branches are proportional to the inferred amount of change in the sequences. It is not usually possible to infer the position of the point of earliest time on a tree from the sequence data. An exception occurs when a gene doubling produces a reiterated sequence that is ancestral to all the sequences on the tree and when this duplication has been well preserved. For some biological groups, such as the vertebrates, the fossil record can be used to fix the location of the point of earliest time as well as its approximate chronological date.

In constructing an evolutionary tree on the basis of sequence data, we treat each amino acid or nucleotide residue in a sequence as an inherited biological trait. This assumption implies our ability



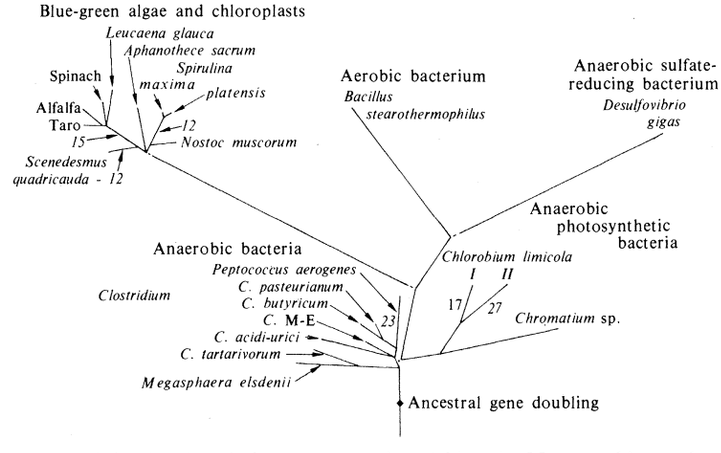


Fig. 1 (left). Probability of inferring the correct topology from sequences of simulated evolutionary connection. A history of five equal intervals of mutational distance was used, as shown at the top of the figure. Ten sets of sequences of 100 residues were generated for each mutational distance (L = 25, 50, 75, 100, 125, 150, 175, 200, 225, 250, 300, 400). Random events were assigned according to the average mutability and mutation pattern of each amino acid (*40*). In totaling the number of correct topologies inferred, a single correct answer counted 1, a wrong answer 0, a two-way tie 0.5, and a three-way tie 0.33. Smoothed curves through the data points are shown. The matrix method is clearly superior. This figure is adapted from (*2*). Fig. 2 (right). Evolutionary tree derived from ferredoxin sequences. Two subtrees were constructed separately on the basis of matrices of percentage differences between sequences from bacteria and from plants and the blue-green algae. A matrix based on an alignment omitting multiple-residue insertions and deletions was used to estimate the evolutionary distance and the connection between the subtrees. Mutiple-residue insertions in the sequences were omitted in constructing the bacterial subtree. The numbers on the tree are evolutionary distances in units of accepted point mutations per 100 residues. The order of divergence of the *Bacillus* and *Desulfovibrio* lines is unclear, as indicated by the dashed connection of the *Desulfovibrio* branch. An insertion of three residues in only these two sequences is consistent with the topology shown, but the topology with the *Desulfovibrio* branch coming directly off the ancestral line to the anaerobic bacteria does have nearly as short an overall branch length. For this alternative topology, the calculated branch lengths in the neighborhood of the *Desulfovibrio* connection would be slightly different.

to align sequences so that changes reflect substitutions of one residue for another during the course of evolution. Insertions and deletions of genetic material affect our ability to align sequences. Superimposed and parallel point mutations limit the accuracy with which we can infer the amount of evolutionary change. (Superimposed mutations are multiple changes at the same site in a sequence; parallel mutations are independent changes in two or more sequences resulting in the same amino acid or nucleotide at corresponding positions.) Experimental and methodological errors in sequencing present further difficulties. These problems affect the overall perspective less as more data become available.

One of the computer methods used for constructing phylogenetic trees proceeds by generating ancestral sequences (*15*); the other produces a least-squares fit to a matrix of evolutionary distances between the sequences (*2, 4*). The ancestral sequence method is a problem in double-minimization of inferred changes. For each possible configuration of the evolutionary tree, a set of ancestral sequences is determined that minimizes the number of inferred changes. Of these configurations, the one that minimizes the total number of changes between ancestral and known sequences is selected as the best representation of the evolutionary tree. This method also yields a set of ancestral sequences corresponding to the branch points of the tree.

In the matrix method, used to construct the trees described here, we begin by calculating a matrix of percentage differences between sequences in an overall alignment. Large unmatched regions, either internal or at the ends of sequences, do not correspond to point mutations and therefore are omitted from our calculations. The matrix elements are corrected for inferred parallel and superimposed mutations according to a scale based on the average way amino acids change during evolution (*16*); this gives evolutionary distances in accepted point mutations per 100 residues. For a given matrix, the determination of the best tree is a problem in double-minimization. For each possible configuration, a set of branch lengths is determined that provides a weighted least-squares fit between the distances given by the reconstructed matrix and those of the original matrix. The configuration that has the minimal total branch length is selected as the best solution.

To test the accuracy of both of these methods, we produced a number of families of sequences that were related through simulated evolutionary change that included amino acid replacements but no insertions or deletions (*2*). The results (Fig. 1) show that for sparse trees of distantly related sequences the matrix method is clearly superior to the ancestral sequence method. Because this is precisely the type of tree with which we are concerned, all of the individual trees shown here were constructed by the matrix method.

The matrix method is much more accurate for distantly related sequences because the information utilized is degraded more slowly by superimposed and parallel mutations than that utilized in the ancestral sequence method. If genetic material has been deleted or inserted, there is an additional factor favoring the use of the matrix method: the results obtained are less affected by variations in gap placement. The number of matching residues is not critically dependent on the exact alignment of two sequences. Typically, several alignments varying in the placement of gaps give the same number, and there are usually many more alignments corresponding to slightly smaller numbers of matches.

Our alignments are based on a computer program that determines the best alignment of two sequences (*17, 18*). The residue-by-residue comparisons match amino acids according to a model of the point mutation process that takes into account amino acid mutabilities and replacement probabilities (*16*). These pairwise sequence comparisons are adjusted to produce a comprehensive alignment. In the alignments used here, the relative magnitudes of terms in the matrices of percentage differences closely reflect the order of similarity of the pairs of sequences. Other criteria can be used in determining gap placement. For example, because overall conformation appears to be well conserved by evolution, Dickerson *et al.* (*19*) propose an alignment of c-type cytochromes that matches residues according to their positions in the three-dimensional structure of the molecules. Alignments based on mutations are more appropriate here because the programs for reconstructing phylogeny seek a minimum number of genetic events.

The topological configuration obtained is not very sensitive to the correction method used for the matrix elements. In the simulated problems (Fig. 1), the curve of correctly inferred topologies obtained directly with the matrix of percentage differences is almost identical with that obtained with the values corrected for presumed superimposed and parallel mutations. The reconstructed branch lengths approached an asymptote when the matrix of percentage differences was used, whereas they were correct on the average (but only very approximately in any particular case) when the corrected matrix was used.

Four protein superfamilies include sequences from several prokaryotes and eukaryotes: ferredoxin, 5*S* ribosomal RNA, c-type cytochromes, and azurin-plastocyanin. Two other superfamilies, flavodoxin and rubredoxin, have sequences that are known from at least four types of bacteria in common with the first superfamilies. Each of these groups of sequences can be used to construct an evolutionary tree; generally, the information that an individual tree provides corresponds closely to the evolution of the biochemical system within which the molecule functions. Plastocyanin, for example, functions in oxygen-releasing photosynthesis, and these sequences provide information about the evolution of photosynthesis in the blue-green algae and in the chloroplasts of higher plants. Together with azurin sequences they also depict the divergence of the blue-green algae from the bacteria.

None of these individual trees, however, gives an overall picture of the course of evolution and the development of new biochemical adaptations from the appearance of the earliest living forms to the divergence of the eukaryote kingdoms. For example, no one tree contains sequences from both cytoplasm and chloroplasts of higher plants; thus, individual trees leave unresolved the question of whether or not the eukaryote organelles had symbiotic origins. Understanding the development of new biochemical pathways requires information that cannot come from a tree derived from data on a single type of molecule. Fortunately, the groups of organisms and eukaryote organelles from which sequences are available overlap in such a way that the trees can be correlated and a composite tree can be constructed. This composite tree depicts more fully the relationships between the major developments in early biological evolution.

Ferredoxins

The ferredoxins are small, iron-containing proteins that are found in a broad spectrum of organisms and that participate in such fundamental biochemical processes as photosynthesis, oxidation-reduction respiratory reactions, nitrogen fixation, and sulfate reduction. The amino acid sequences of these proteins have been elucidated by a number of

workers, including particularly K. T. Yasunobu, H. Matsubara, and their coworkers [see (*20, 21*)]. The tree (Fig. 2) derived from these sequences provides a framework for the events outlined in the other evolutionary trees presented here; moreover, a gene-doubling shared by all the ferredoxin sequences makes it possible to deduce the point of earliest time in these trees. The clostridial-type ferredoxins, in particular, are still very similar in sequence to the extremely ancient protein that duplicated. Most of these ferredoxins are composed of fewer than the 20 coded amino acids, and they lack those amino acids that are thermodynamically least stable, such as tryptophan and histidine (*22*).

The clostridial-type ferredoxins show the strongest evidence of gene-doubling. Using our computer program RELATE (*17*), we compared all pairs of different segments that were 15 residues long within each of these proteins and calculated a probability of less than 10^{-9} that the repetitive character of the two halves occurred by chance. From an alignment of the first and second halves of these sequences, we inferred an ancestral half-chain sequence. This sequence, doubled, was included in the computations of the evolutionary tree. Because all of the ferredoxin sequences show some evidence of gene-doubling, this event must have occurred prior to the species divergences shown here. The doubled sequence is therefore located at the base of the tree. All organisms near the base of the tree (*Clostridium*, *Megasphaera*, and *Peptococcus*) are anaerobic, heterotrophic bacteria (*23*). Most species of these groups lack heme-containing proteins, such as the cytochromes and catalase. It has long been thought that, of the extant bacteria, these species most closely reflect the metabolic capacities of the earliest species (*24*). In Fig. 2, *Chlorobium* and *Chromatium* are pictured as having diverged very early from the ancestral heterotrophic bacteria, although the exact point of this divergence is not clearly resolved. *Chromatium* and *Chlorobium* are anaerobic bacteria capable of photosynthesis using H_2S as an exogenous electron donor. Of the anaerobic bacteria shown, only *Chlorobium limicola* cannot live fermentatively, and it is reasonable to suppose that this ability is primitive and was lost by this bacterium. The two ferredoxins in *Chlorobium* are the result of a gene duplication within this line.

In Fig. 2, the *Bacillus* and *Desulfovibrio* lines diverge next from the line leading to the blue-green algae. Members of the genus *Bacillus* are either strictly aerobic or facultatively aerobic, capable of respiring aerobically but living fermentatively under anaerobic conditions. *Desulfovibrio* is a sulfate-reducing bacterium; it respires anaerobically using sulfate as the terminal electron acceptor. The use of sulfate by *Desulfovibrio* contrasts sharply with the use of oxygen as the terminal electron acceptor of respiration by *Bacillus*. This difference suggests that the divergence of these bacteria occurred after some components in the respiratory chain had developed. The topology pictured here indicates that the final components in the chain evolved separately in the *Bacillus*, *Desulfovibrio*, and blue-green algal lines.

The plant-type ferredoxins are all very closely related. Those from the green alga *Scenedesmus* and the higher plants are found in the chloroplasts of these organisms. *Spirulina* and *Nostoc* belong to one major division of the blue-green algae, the filamentous type; *Aphanothece* represents the other major division, the coccoid type (*25*).

The tree in Fig. 2 indicates that the structure of ferredoxin has changed much more in some lines than in others; this, at least in part, reflects changes in ferredoxin function. The clostridial sequences are little changed over the entire time represented, possibly more than 3 billion years. The adjustment to bacterial photosynthesis required somewhat more change, the adjustment to an aerobic metabolism required still more, and the most change occurred in the adaptation to oxygen-releasing photosynthesis. Unlike the situation in eukaryotes, the rate of acceptance of point mutations in prokaryote sequences is very uneven and cannot provide a useful evolutionary clock.

Origins of Eukaryote Organelles

There are two schools of thought concerning the origin of the eukaryote mitochondria and chloroplasts: one is that they arose by the compartmentalization of the DNA within the cytoplasm of an evolving protoeukaryote (*26*); the other is that they arose from free-living forms that established symbiotic relationships with host cells (*27*). According to the first theory, all genes arose within a single ancestral line, and homologs found in both the nucleus and the organelles arose by gene duplication. Thus, in the evolutionary tree for ferredoxin, the animals and fungi would appear together with the higher plants in the upper portion of the tree, after the divergence of the blue-green algae.

According to the symbiotic theory, the chloroplasts descended from free-living blue-green algae; other symbionts would include the mitochondrion, which was originally a free-living aerobic bacterium, and the flagellum and mitotic apparatus, which may have descended from spirochetes. It is proposed that these prokaryotes separately invaded protoeukaryote host cells with which they became symbiotic and continued to evolve to their current status as organelles. If the symbiotic theory is accurate, mitochondrial and chloroplast genes should show evidence of recent common ancestry with the separate types of contemporary free-living prokaryote forms. The host and organelles should occur on different branches that also contain free-living forms. The tree of plant-type ferredoxins would thus depict the radiation of blue-green algae followed by the development of symbiosis between one of these organisms and an ancestor of *Scenedesmus* and the higher plants. Although the appearance of the ferredoxin tree, particularly its branch lengths, is more consistent with this theory than with the first explanation, the tree by itself does not enable us to distinguish between the two theories because no ferredoxin sequences are available from the eukaryote cytoplasm or mitochondria.

5*S* Ribosomal RNA

The 5*S* ribosomal RNA molecule has a low molecular weight and is about 120 nucleotides in length. It is associated with the larger ribosomal subunit and is thought to function in the nonspecific binding of transfer RNA to the ribosome during protein synthesis (*28*). Because this function is independent of the kind of amino acid, this type of molecule could be extremely ancient, predating the contemporary form of the genetic code. Sequences of 5*S* ribosomal RNA have been determined by a number of workers, including B. J. Forget, S. M. Weissman, and C. R. Woese [see (*29, 30*)]; they have been taken from a wide variety of sources, including aerobic and anaerobic bacteria, blue-green algae, and the cytoplasm of several eukaryotes. The cytoplasmic sequences, in particular, present the possibility that an evolutionary tree based on this molecule will provide further insight into the origin of the eukaryotes.

Aligning nucleotide sequences in a way that reflects their evolution is more difficult than aligning amino acid sequences because there are only four kinds of bases. However, knowledge of

the secondary structure alleviates the alignment problem somewhat because we can assume that positions involved in the base-paired regions of the molecule were highly conserved during evolution. We have aligned the known sequences to reflect their natural division into groups from prokaryotes and eukaryotes and to match a model of their secondary structure (*29*) adapted from Nishikawa and Takemura (*31*).

We derived an evolutionary tree (Fig. 3) on the basis of this alignment and placed its base on the branch to the anaerobic bacterium *Clostridium* in conformance with the ferredoxin tree. All the eukaryote 5*S* ribosomal RNA's were isolated from cytoplasmic ribosomes, one of the three ribosomal systems found in eukaryotes.

The branch leading to these cytoplasmic sequences diverges from the prokaryotes at a point that is close to the origin of the *Bacillus* branch. Like some members of the genus *Bacillus*, *Escherichia* is also facultatively aerobic and has both fermentative and respiratory metabolisms. Unless aerobic respiration arose separately in the *Bacillus* and *Escherichia* lines, their most recent common ancestor had this bimodal metabolic capacity also.

The two lines leading to organisms that are capable of oxygen-releasing photosynthesis appear on opposite sides of this tree. One leads to rye and *Chlorella*, a eukaryote green alga, the other to *Anacystis*, a blue-green alga. *Anacystis* is grouped in the same family with *Aphanothece* (*25*), which appears on the ferredoxin tree. These coccoid blue-green algae are certainly more closely related than the blue-green algal orders represented by *Aphanothece* and *Spirulina*. Thus, we predict that *Aphanothece* would be found to diverge near the end of the *Anacystis* branch, preceded slightly by the divergence of the chloroplast branches. The very separate history of the cytoplasmic sequences points to a symbiotic origin of the chloroplasts, with the cytoplasmic sequences representing the evolution of the organism that was invaded by a blue-green alga.

C-Type Cytochrome

The evolutionary tree based on c-type cytochromes (Fig. 4) is important to an understanding of the origin and evolution of the mitochondrion. R. Ambler, E. Margoliash, D. Boulter, M. Kamen, E. Smith, and G. Pettigrew, in particular, are responsible for sequencing many of these proteins [see (*32*)]. Cytochrome c is coded in the nucleus but functions in the mitochondrion. This is usually explained in the symbiotic theory by transfer of genetic information, including the gene for cytochrome c, from the invading aerobic bacterium to the protoeukaryote host during the development of their current relationship. In the nonsymbiotic theory, genetic rearrangement is also an essential feature.

The eukaryote mitochondrion is placed in the portion of the tree (Fig. 4) that includes the aerobic bacteria after their divergence from the blue-green algae and chloroplasts. The cytochrome c sequences are on a branch that most recently diverged from cytochrome c_2 of the nonsulfur, purple, photosynthetic bacteria; together these diverged from the branch leading to cytochrome c_{551} from strict aerobes such as *Pseudomonas*. This contrasts with the evolution of the eukaryote cytoplasmic constituents depicted by the tree derived from 5*S* ribosomal RNA (Fig. 3). There the eukaryotes diverged with the facultatively aerobic bacteria, such as *Bacillus*, from the line leading toward the blue-green algae. *Pseudomonas* diverged from this line somewhat later. This contrasting

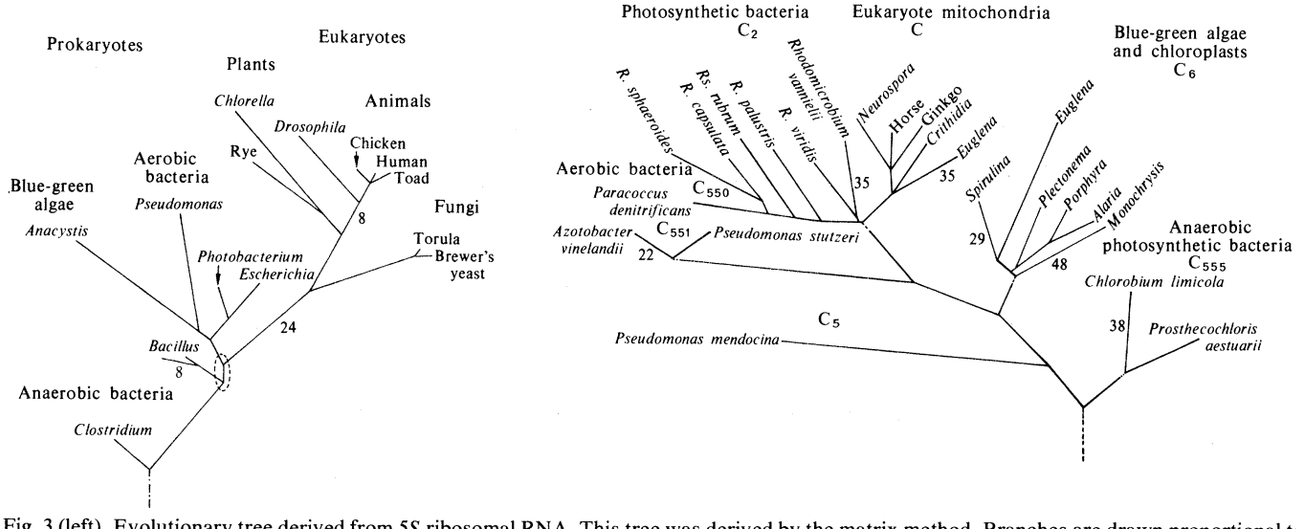


Fig. 3 (left). Evolutionary tree derived from 5*S* ribosomal RNA. This tree was derived by the matrix method. Branches are drawn proportional to the amount of evolutionary change they represent; selected branch lengths are indicated in numbers of accepted point mutations per 100 residues. The order of divergence for the branches leading to the eukaryotes and to *Bacillus* is not clearly resolved; the tree whose topology reverses the order of these branches has nearly as short an overall length. The prokaryote species shown are *Clostridium pasteurianum*, *Bacillus megaterium*, *B. licheniformis*, *Escherichia coli*, *Pseudomonas fluorescens*, and *Anacystis nidulans*. The *Chlorella* species is *C. pyrenoidosa*. Fig. 4 (right). Evolutionary tree derived from c-type cytochromes. The subtrees pictured here, such as the cytochrome c_6 tree, were derived separately from matrices of percentage differences between the complete sequences. Branches connecting the subtrees were estimated from a matrix calculated from an alignment that omitted multiple-residue insertions and deletions. Branch lengths are given in accepted point mutations per 100 residues. The points of earliest divergence in the individual subtrees cannot be precisely determined and, therefore, the bases of the subtrees are represented by dashed lines. Cytochrome c_{550} and all of the c_2's differ from the other c-type cytochromes in having a deletion close to the heme-binding cysteine in their sequences and on this basis were placed on a branch separate from the cytochrome c sequences. *Euglena* and *Crithidia* have been placed on a single branch together, a configuration slightly less than optimal, because they share a unique mutation of the active cysteine. The connections of the c_{551} and c_{555} sequences have been centered. The genera *Rhodospirillum* and *Rhodopseudomonas* are abbreviated *Rs.* and *R.*, respectively. Cytochrome c_6 sequences were taken from the following species: *Spirulina maxima*, *Monochrysis lutheri*, *Porphyra tenera*, *Euglena gracilis*, *Alaria esculenta*, and *Plectonema boryanum*; cytochrome c sequences were taken from protists *Euglena gracilis* and *Crithidia oncopelti*. This tree is adapted from (*2*).

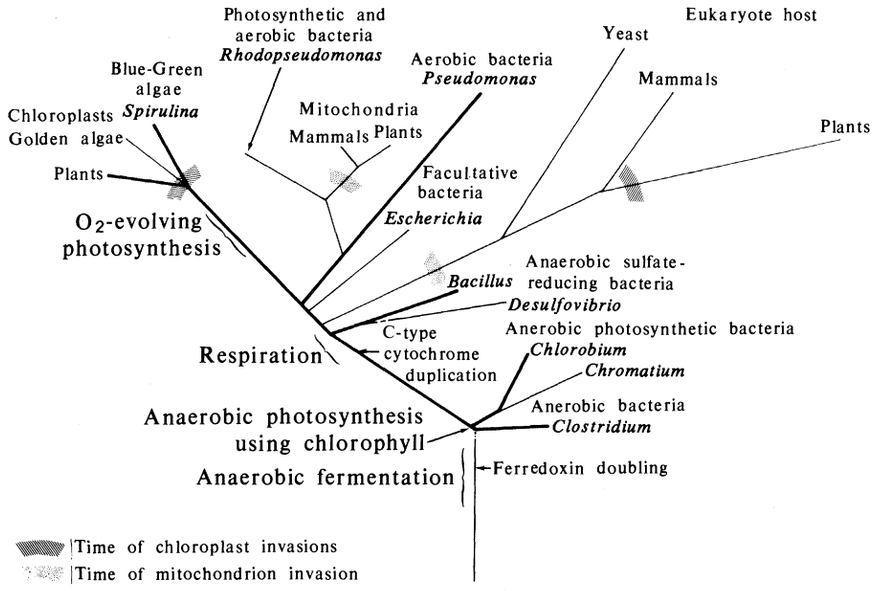


Fig. 5. Composite evolutionary tree. This tree presents an overview of early evolution based on ferredoxin, c-type cytochromes, and 5*S* ribosomal RNA sequences. The heavy lines represent a tree calculated from a matrix of evolutionary distances combining two or more of the individual trees. The lighter lines represent branches scaled from a single tree and added to the combined tree. The point in evolution at which the mitochondrial symbiosis occurred is stippled; the chloroplast symbiosis is shaded.

picture reinforces arguments favoring a symbiotic origin for the mitochondrion. The cytoplasmic 5*S* ribosomal RNA sequences appear to describe the evolution of the protoeukaryote host. The subtree for animals that was derived from cytochrome c (*5*) is consistent with that derived from 5*S* ribosomal RNA sequences. According to the symbiotic theory, this would be because these divergences were subsequent to the mitochondrial invasion of host cells.

Cytochrome c_2 is found in bacteria such as *Rhodomicrobium* and *Rhodopseudomonas*. These bacteria photosynthesize anaerobically but respire aerobically. *Paracoccus denitrificans* possesses cytochrome c_{550}, which is very similar to cytochrome c_2 along its entire length. The position of *Paracoccus* in the tree suggests that it arose from a nonsulfur photosynthetic bacterium by the loss of its photosynthetic ability (*19*).

Two different c-type cytochromes, c_5 and c_{551}, are found in *Pseudomonas*, a nonphotosynthetic bacterium. These may be the result of a gene duplication early in the tree with subsequent loss of the c_5 gene in some lines. For species in which the gene was not lost, we would expect that the topology of the main tree subsequent to the duplication would be reiterated on the branch that included the c_5 sequence. A transfer of genetic material from a bacterium that contained the c_5 gene to *Pseudomonas* could also explain this branch.

The presumed duplication of the c-type cytochrome gene in *Pseudomonas* brings up a problem in interpreting evolutionary trees that is especially acute here. If we are treating two products of an unsuspected gene duplication, we will assume that the two lines of protein evolution correspond to a single pattern of species evolution. This can lead us to believe that two closely related organisms are quite distant from one another. For example, if cytochrome c_{551} had gone undetected in species of *Pseudomonas*, the tree in Fig. 4 would suggest that *Azotobacter* and *Pseudomonas* were very distantly related. At present, metabolic function guides our selection of homologous proteins and helps avoid this difficulty. In the tree of c-type cytochromes, protein function has changed, and this presents an added difficulty in interpretation. As a more complete picture of the protein complement of each species becomes known, this problem will become unimportant.

Cytochrome c_6 is found in the photosynthetic lamellae of blue-green algae and in the chloroplasts of eukaryotes where it functions in the electron transport chain between photosystems I and II. As in the ferredoxin tree, there is a close similarity between the sequences from the blue-green alga *Spirulina* and the various eukaryote algal chloroplasts; as in the 5*S* ribosomal RNA tree, the blue-green algae are most closely related to strictly aerobic bacteria, such as *Pseudomonas*. It is not possible to locate precisely the point at which the main tree connects to the subtree of cytochrome c_6. However, the topology and branch lengths reflect a symbiotic origin for photosynthesis in eukaryotes. There are separate branches leading to the two filamentous blue-green algae, *Spirulina* and *Plectonema*, intermixed with the eukaryote algae branches. The most direct explanation for this is that the cytochrome c_6 subtree reflects, at least in part, the evolutionary relationships among the blue-green algae that became symbionts rather than the speciation of eukaryotes. Some of the eukaryote algal chloroplasts appear to be derived from different symbiotic associations, as Raven has suggested (*33*).

The point of earliest time on the tree for c-type cytochromes was placed near the divergence of *Chlorobium* and *Prosthecochloris*; both of these are anaerobic, obligate, photosynthetic bacteria. This is consistent with the position of the sequence from *Chlorobium* in the ferredoxin tree. As in the tree of 5*S* ribosomal RNA's, prokaryote branch lengths are strikingly unequal, probably reflecting changes in protein function, and cannot be used to estimate time reliably.

Composite Evolutionary Tree

Each of the individual trees we have presented contains information about the early course of biological evolution. We have used the topologies and evolutionary distances derived for these trees to construct a composite tree (Fig. 5) which, although it is based on sparse data from a few species, begins to provide a coherent evolutionary framework that can be expanded as new sequence data become available.

We constructed the composite evolutionary tree from a composite matrix according to the same methods that we used to construct the individual trees. The matrix included the six species that appear on at least two of the three individual trees (their branches are represented by the heavy lines in Fig. 5). First, the trees were scaled so that distances were comparable. To do this, we compared the overall length of the distances each tree had in common with each other tree. These ratios were adjusted slightly to give a consistent set of scale factors. A combined matrix of distances between the six species was calculated by averaging the scaled contributions from each of the individual trees.

As previously, for each possible configuration of the combined trees, a set of branch lengths was determined that provided a weighted least-squares fit between the matrix of distances between species and a matrix reconstructed from the tree. The configuration with the shortest overall branch length was chosen. This configuration is the one that is also consistent with all of the individual trees. Finally, branches found in only one tree were scaled in length and added to the composite tree, maintaining the relative internodal distances (these are represented by light lines in Fig. 5).

The genetic doubling of the clostridial-type ferredoxins allows us to locate the base on this tree. Moreover, because the species whose sequences have changed least since this doubling event were all anaerobic, heterotrophic bacteria, it is likely that the ability to live fermentatively is primitive.

The composite tree describes the evolution of photosynthesis using chlorophyll, starting with the development of the ability to synthesize this class of compounds. Three families of photosynthetic bacteria are represented: Chromatiaceae, Chlorobiaceae, and Rhodospirillaceae. The divergence of the Chromatiaceae and the Chlorobiaceae from the other anaerobic bacteria was quite early, and it is clear that this type of photosynthesis arose at a very early stage in evolution and has not changed much. The Rhodospirillaceae provide an example of the confusion that morphological criteria can cause. As Stanier *et al.* (*23*) point out, these bacteria are indistinguishable in structure and pigments from members of the Chromatiaceae under anaerobic conditions; however, when grown under strictly aerobic conditions, they appear to be the same as nonphotosynthetic bacteria of similar form, such as the pseudomonads. The c-type cytochrome sequences clearly place them in a portion of the tree surrounded by strictly aerobic forms on a branch leading to *Pseudomonas*.

If we assume that it is very hard to achieve a photosynthetic metabolism with its many coordinated macromolecules, but relatively easy to lose one through any one of many genetic changes, then it is reasonable to suppose that the main trunk of this tree represents a continuum of photosynthetic forms. We would not be surprised to find photosynthetic forms branching off at any point. Except for the early anaerobes, all nonphotosynthetic bacteria would be descended from a few ancestral forms that have independently lost their photosynthetic ability. This is very different from biological classifications where all photosynthetic forms are grouped together, separate from the nonphotosynthetic forms.

The next major event shown on the main trunk of the tree is the development of aerobic respiration. As we noted in discussing the ferredoxin evolutionary tree, the divergence of *Bacillus* and *Desulfovibrio* probably marks the appearance of some components of this adaptation. The final elements in this adaptation were evolved separately because these groups differ in their terminal electron acceptor in respiration. Additionally, the divergence of cytochrome c_5 occurred just prior to this time. In the tree of the c-type cytochrome, it is unclear whether this divergence represents a gene duplication or a recent genetic transfer; in the context of the evolution of a respiratory metabolism, a duplication of the ancestral gene could have provided the genetic material and relaxed evolutionary constraints necessary for the development of aerobic respiration.

The clearest and most direct interpretation of the sequence data is provided by the symbiotic theory for the origin of the eukaryotes. The branch we identify as the eukaryote host diverged at about the same time as *Bacillus* and *Escherichia*. Both of these bacteria are facultative aerobes. This suggests that the ancestral eukaryote host was also facultatively aerobic at the time of its divergence.

The bacterium that became the mitochondrion was most closely related to the third family of photosynthetic bacteria, the Rhodospirillaceae. The topology of the tree invites the speculation that this ancestral bacterium was photosynthetic until shortly before or just after it invaded the host. Because a single cytochrome c is found in most eukaryotes, it seems reasonable to suppose that the aerobic respiratory metabolism of this invading protomitochondrion was more effective than that of the host and that the host lost any primitive system that it might have had.

The final biochemical adaptation depicted here is the development of photosystem II. The blue-green algae and chloroplasts of the eukaryotes are capable of oxygen-releasing photosynthesis, and the available sequence data point to this capacity having evolved only once. It appears to have combined the new biochemical adaptation, photosystem II, with proteins modified from two earlier adaptations, bacterial photosynthesis and respiration. The chloroplasts, like the mitochondria, are grouped together with free-living prokaryotes, a result consistent with the symbiotic theory.

It is frequently suggested that aerobic respiration developed after oxygen-releasing photosynthesis as a protective mechanism in response to atmospheric oxygen produced by blue-green algae (*24, 34*). The composite tree clearly suggests that many components of the respiratory chain predate oxygen-releasing photosynthesis; all of the organisms on the tree above *Desulfovibrio* are aerobic, whereas the earlier branches all lead to anaerobic forms. Although possible, it seems unlikely that the use of oxygen as the terminal electron acceptor in respiration evolved separately in the lines leading to the blue-green algae, the facultative aerobic bacteria, and the strictly aerobic bacteria after the development of photosystem II in the blue-green algal line. As Schopf (*8*) has pointed out, it is difficult to imagine the development of oxygen-releasing photosynthesis prior to the development of a rudimentary mechanism for coping with oxygen. Oxygen is produced intracellularly in photosynthesis, and it is intracellular components that must be protected from oxidation. A rudimentary form of aerobic respiration most probably arose at a time near the divergence of the *Bacillus* line from that leading to the blue-green algae.

The composite tree makes it particularly clear that the three branches that contribute to the eukaryote host and organelles are distinctly separate; each is closely related to free-living prokaryotes. The chloroplasts share a recent ancestry with the blue-green algae; the mitochondrion shares a recent ancestry with certain respiring and photosynthetic bacteria, the Rhodospirillaceae; whereas the eukaryote host diverged from the other groups at a considerably earlier time along with *Bacillus* and *Escherichia*.

Corroborative Sequence Data

A limited amount of sequence data from four or more species is available from other proteins. Azurin and plastocyanin are blue, copper-containing proteins whose sequences show statistical evidence of their common evolutionary origin (*20*). Moreover, they appear to be parts of homologous electron transport systems because each of them exchanges electrons with c-type cytochromes. Azurin is thought to exchange

an electron with cytochrome c_{551} in bacterial respiration (*35*); plastocyanin exchanges an electron with cytochrome c_6 in the electron transport chain between photosystems I and II.

Plastocyanin sequences from both eukaryote chloroplasts and a blue-green alga have been determined by a number of workers (*20, 36*) including D. Boulter and A. Aitken, in particular; azurin sequences from aerobic bacteria have been determined by R. Ambler (*20*). We constructed an evolutionary tree (Fig. 6) based on these sequences. The relationships depicted are consistent with those presented in the other trees. The aerobic bacteria, on the right side of the tree, are very closely related, and the extents of their divergences are comparable to those of the higher plants, shown on the left side of the tree. *Anabaena*, a filamentous blue-green alga, is closely related to chloroplasts of the eukaryote green alga *Chlorella* and of the higher plants. The topology of this tree again is consistent with the symbiotic origins for chloroplasts. The divergences of the higher plant chloroplasts are close enough to classical phylogenies to be explained by a single symbiotic association that predated plant divergences. Lewin (*37*) has recently proposed a new division, the Prochlorophyta, for a group of bright green, generally spherical, prokaryote algae. These had previously been classified with the blue-green algae. However, like the higher plant chloroplasts and unlike the blue-green algae they contain both chlorophylls a and b and no detectable bilin pigments. Sequences from these organisms might be especially helpful in extending our understanding of the evolution of higher plant photosynthesis.

Flavodoxins are low-molecular weight proteins that have been isolated so far from only bacteria and algae. They have one flavin mononucleotide prosthetic group per molecule and substitute for ferredoxins in a variety of reactions, including the phosphoroclastic splitting of pyruvate and the photosynthetic reduction of nicotinamide-adenine dinucleotide phosphate. Flavodoxin sequences are known from three anaerobic bacteria, *Clostridium, Megasphaera,* and *Desulfovibrio,* and an aerobic bacterium, *Azotobacter*; most of these sequences were determined by K. T. Yasunobu, J. L. Fox, and their co-workers [see (*38*)]. The matrix of the percentage differences between these flavodoxin sequences (Table 1) reveals the close relationship of the highly conserved *Megasphaera* and *Clostridium* sequences and of the more highly evolved *Desulfovibrio* and *Azotobacter* sequences, supporting the phylogeny we have drawn.

Rubredoxin, another protein that participates in electron transport, has also been used as a basis for constructing an evolutionary tree (*20, 39*). This tree includes sequences from *Megasphaera elsdenii, Peptococcus aerogenes, Clostridium pasteurianum, Pseudomonas oleovorans, Desulfovibrio gigas,* and *D. vulgaris,* and is also consistent with the ones we have constructed except that the closely related *Desulfovibrio* sequences diverge close to the earliest point on the tree and are very highly conserved. That the placement of the *Desulfovibrio* branch is inconsistent with that depicted in the trees for ferredoxin and flavodoxin suggests that there might be an unsuspected gene duplication in the rubredoxins, a rare occurrence of an accepted gene transfer, or a misleading concatenation of evolutionary events in these short sequences.

Summary

If current estimates of the antiquity of life on earth are correct, bacteria very much like *Clostridium* lived more than 3.1 billion years ago. Bacterial photosynthesis evolved nearly that long ago, and it seems reasonable, in view of our composite tree, to attribute the most ancient stromatolites, formed nearly 3.0 billion years ago, to early photosynthetic bacteria. Blue-green algae appear to have evolved later. The tree shows that by the time oxygen-releasing photosynthesis originated in the blue-green algal line, there must have been a great diversity of morphological types, including bacteria that are ancestral to most of the major groups pictured on the composite tree. This time probably corresponded to the great increase in complexity of the fossil record about 2 billion years ago. Our composite tree suggests that aerobic respiration preceded oxygen-releasing photosynthesis. This may mean that the formation of oxygen from water in the upper atmosphere was important to evolving prokaryotes prior to 2 billion years ago. Oxygen-releasing photosynthesis arose later and was, in large measure, responsible for the final transition to the present-day oxygen level in the atmosphere. Judging from the relative branch lengths on the tree, the mitochon-

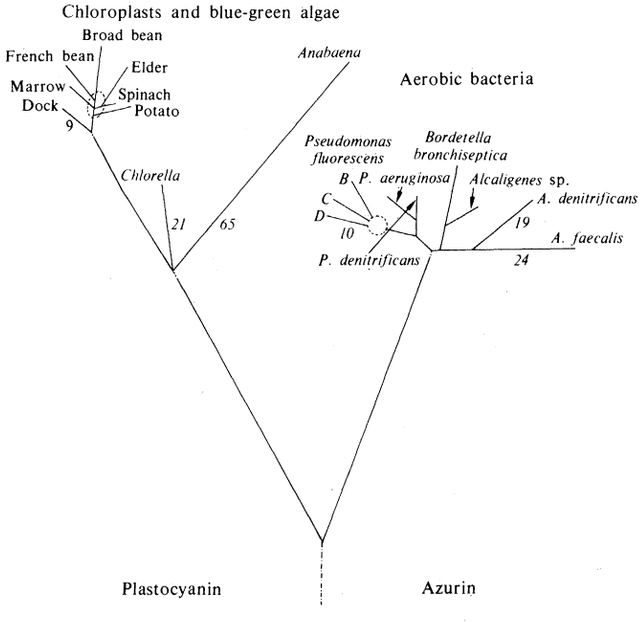


Table 1. Flavodoxins

	Number of differences			
Megasphaera elsdenii		76	107	155
Clostridium MP	55		104	152
Desulfovibrio vulgaris	71	69		151
Azotobacter vinelandii	85	83	82	

Percentage difference

Fig. 6. Evolutionary tree derived from azurin and plastocyanin. Subtrees for the two proteins were derived separately from matrices calculated from complete sequences. In order to estimate the evolutionary distance between the two families, we used a matrix calculated from an alignment that omitted multiple-residue insertions and deletions between the families. Branch lengths are drawn proportional to amounts of evolutionary change in the units of accepted point mutations per 100 residues. The positions of the connecting branch within the subtrees cannot be precisely determined from the data, and therefore the connections are represented by dashed lines. The order of divergence of the three biotypes of *Pseudomonas fluorescens* could not be resolved. The order of divergence of the chloroplasts of higher plants is also unclear; the topology with the minimal overall length is shown. The algal species shown are *Chlorella fusca* and *Anabaena variabilis*.

drial invasion occurred during this transition. Finally, perhaps 1.1 billion years ago, several independent symbioses between protoeukaryotes and various blue-green algae gave rise to photosynthetic eukaryotes; some of these developed into modern eukaryote algae, whereas a single line, possibly from an ancestral green alga, appears to have evolved into the higher plants.

By combining the information from evolutionary trees based on several types of sequences, we have developed a broad outline of early events in the emergence of life that can be refined as new sequence information becomes available. The schema presents a working hypothesis for relating the many observations from metabolic and morphological studies of bacteria and from paleogeology. Eventually all of the biochemical components of intermediary metabolism may be correlated with the development of the prokaryote types and their metabolic capacities.

References and Notes

1. V. Bryson and H. J. Vogel, Eds., *Evolving Genes and Proteins* (Academic Press, New York, 1965); M. O. Dayhoff and R. V. Eck, Eds., *Atlas of Protein Sequence and Structure* (National Biomedical Research Foundation, Washington, D.C., 1968), vol. 3.
2. M. O. Dayhoff, *Fed. Proc. Fed. Am. Soc. Exp. Biol.* **35**, 2132 (1976).
3. R. V. Eck and M. O. Dayhoff, Eds., *Atlas of Protein Sequence and Structure* (National Biomedical Research Foundation, Washington, D.C., 1966), vol. 2.
4. W. M. Fitch and E. Margoliash, *Science* **155**, 279 (1967).
5. P. J. McLaughlin and M. O. Dayhoff, *J. Mol. Evol.* **2**, 99 (1973).
6. W. M. Fitch, *ibid.* **8**, 13 (1976); M. Goodman, G. W. Moore, J. Barnabas, G. Matsuda, *ibid.* **3**, 1 (1974).
7. M. O. Dayhoff, L. T. Hunt, P. J. McLaughlin, D. D. Jones, in *Atlas of Protein Sequence and Structure*, M. O. Dayhoff, Ed. (National Biomedical Research Foundation, Washington, D.C., 1972), vol. 5, p. 17.
8. J. W. Schopf, *Annu. Rev. Earth Planet. Sci.* **3**, 213 (1975).
9. P. Cloud, *Paleobiology* **2**, 351 (1976).
10. J. W. Schopf, *Origins Life* **7**, 19 (1976).
11. T. R. Mason and V. Von Brunn, *Nature (London)* **266**, 47 (1977).
12. P. E. Cloud, Jr., *Science* **160**, 729 (1968).
13. J. W. Schopf and D. Z. Oehler, *ibid.* **193**, 47 (1976).
14. M. O. Dayhoff, W. C. Barker, P. J. McLaughlin, *Origins Life* **5**, 311 (1974).
15. M. O. Dayhoff, C. M. Park, P. J. McLaughlin, in *Atlas of Protein Sequence and Structure*, M. O. Dayhoff, Ed. (National Biomedical Research Foundation, Washington, D.C., 1972), vol. 5, p. 7; W. M. Fitch and J. S. Farris, *J. Mol. Evol.* **3**, 263 (1974).
16. M. O. Dayhoff, Ed., *Atlas of Protein Sequence and Structure* (National Biomedical Research Foundation, Washington, D.C., 1976), vol. 5, suppl. 2, p. 311.
17. M. O. Dayhoff, in *ibid.*, p. 1.
18. S. B. Needleman and C. D. Wunsch, *J. Mol. Biol.* **48**, 443 (1970).
19. R. E. Dickerson, R. Timkovich, R. J. Almassy, *ibid.* **100**, 473 (1975).
20. W. C. Barker, R. M. Schwartz, M. O. Dayhoff, in *Atlas of Protein Sequence and Structure*, M. O. Dayhoff, Ed. (National Biomedical Research Foundation, Washington, D.C., 1972), vol. 5, p. 51. Respiratory protein sequences and our alignments of them are collected here as are tabulations of computer alignment scores for the ferredoxins.
21. T. Hase, K. Wada, H. Matsubara, *J. Biochem. (Tokyo)* **79**, 329 (1976); M. Tanaka, M. Haniu, K. T. Yasunobu, K. K. Rao, D. O. Hall, *Biochemistry* **13**, 5284 (1974); _____; *ibid.* **14**, 5535 (1975); K. Wada, T. Hase, H. Tokunaga, H. Matsubara, *FEBS Lett.* **55**, 102 (1975); M. Tanaka, M. Haniu, K. T. Yasunobu, M. C. W. Evans, K. K. Rao, *Biochemistry* **14**, 1938 (1975); T. Hase, N. Ohmiya, H. Matsubara, R. N. Mullinger, K. K. Rao, D. O. Hall, *Biochem. J.* **159**, 55 (1976); T. Hase, K. Wada, M. Ohmiya, H. Matsubara, *J. Biochem. (Tokyo)* **80**, 993 (1976).
22. R. V. Eck and M. O. Dayhoff, *Science* **152**, 363 (1966).
23. R. E. Buchanan and N. E. Gibbons, Eds., *Bergey's Manual of Determinative Bacteriology* (Williams & Wilkins, Baltimore, Md., ed. 8, 1974); R. Y. Stanier, M. Doudoroff, E. A. Adelberg, *The Microbial World* (Prentice-Hall, Englewood Cliffs, N.J., ed. 3, 1970). Nutritional and ecological descriptions of the bacterial genera used in the text were taken from these references.
24. E. Broda, *The Evolution of the Bioenergetic Process* (Pergamon, Oxford, 1975).
25. T. V. Desikachary, in *The Biology of Blue-Green Algae*, N. G. Carr and B. A. Whitton, Eds. (Univ. of California Press, Berkeley, 1973), p. 473.
26. R. A. Raff and H. R. Mahler, *Science* **177**, 575 (1972); T. Uzzell and C. Spolsky, *Am. Sci.* **62**, 334 (1974).
27. L. Margulis, *Origin of Eukaryotic Cells* (Yale Univ. Press, New Haven, Conn., 1970); _____, in *Handbook of Genetics*, R. C. King, Ed. (Plenum, New York, in press), vol. 1.
28. R. Monier, in *Ribosomes*, M. Normura, A. Tissieres, P. Lengyel, Eds. (Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., 1974), p. 141.
29. R. M. Schwartz and M. O. Dayhoff, in (*16*), p. 293. Sequences of 5*S* ribosomal RNA's, a model of their secondary structure, and our alignment of them are collected here, as are tabulations of computer alignment scores.
30. C. R. Woese, C. D. Pribula, G. E. Fox, L. B. Zablen, *J. Mol. Evol.* **5**, 35 (1975); C. D. Pribula, G. E. Fox, C. R. Woese, *FEBS Lett.* **64**, 350 (1976); H. A. Raue, T. J. Stoff, R. J. Panta, *Eur. J. Biochem.* **59**, 35 (1975); J. Benhamou and B. R. Jordan, *FEBS Lett.* **62**, 146 (1976); P. I. Payne and T. A. Dyer, *Eur. J. Biochem.* **71**, 33 (1976).
31. K. Nishikawa and S. Takemura, *J. Biochem. (Tokyo)* **76**, 935 (1974).
32. M. O. Dayhoff and W. C. Barker, in (*16*), p. 25. Sequences of c-type cytochromes and our alignment of them are collected here, as are tabulations of computer alignment scores; R. P. Ambler, T. E. Meyer, M. D. Kamen, *Proc. Natl. Acad. Sci. U.S.A.* **73**, 472 (1976); A. Aitken, *Nature (London)* **263**, 793 (1976).
33. P. H. Raven, *Science* **169**, 641 (1970).
34. L. Margulis, J. C. G. Walker, M. Rambler, *Nature (London)* **264**, 620 (1976).
35. M. Brunori, C. Greenwood, M. T. Wilson, *Biochem. J.* **137**, 113 (1974).
36. B. Haslett, C. R. Bailey, J. A. M. Ramshaw, M. D. Scawen, D. Boulter, *Biochem. Soc. Trans.* **2**, 1329 (1974); J. A. M. Ramshaw, M. D. Scawen, D. Boulter, *Biochem. J.* **141**, 835 (1974); M. D. Scawen and D. Boulter, *ibid.* **143**, 237 (1974); M. D. Scawen, J. A. M. Ramshaw, D. Boulter, *ibid.* **147**, 343 (1975); A. Aitken, *ibid.* **149**, 675 (1975).
37. R. A. Lewin, *Nature (London)* **261**, 697 (1976).
38. W. C. Barker and M. O. Dayhoff, in (*16*), p. 67. Flavodoxin sequences are collected here; M. Tanaka, M. Haniu, K. T. Yasunobu, D. C. Yoch, *Biochem. Biophys. Res. Commun.* **66**, 639 (1975).
39. H. Vogel, M. Bruschi, J. Le Gall, *J. Mol. Evol.* **9**, 111 (1977).
40. M. O. Dayhoff, R. V. Eck, C. M. Park, in *Atlas of Protein Sequence and Structure*, M. O. Dayhoff, Ed. (National Biomedical Research Foundation, Washington, D.C., 1972), vol. 5, p. 89.

41. Supported by NASA contract NASW 3019 and NIH grants GM-08710 and RR-05681. We thank W. C. Barker and B. C. Orcutt for their helpful discussions and criticism and M. J. Gantt for technical assistance.