Proc. Indian Acad. Sci. (Anim. Sci.), Vol. 93, No. 3, April 1984, pp. 141–158.
© Printed in India.

Aspects of biological evolution based on molecular data sets

SHAMA BARNABAS and JOHN BARNABAS*

Post-graduate School for Biological Studies, Ahmednagar College, Ahmednagar 414001, India

*Jawaharlal Nehru Fellow, National Chemical Laboratory, Pune 411 007, India

Keywords. Biological evolution; phylogenetic trees; eukaryotes.

1. Introduction

Establishment of phylogenetic relatedness of organisms is traditionally done on the basis of evidences from fossil record, comparative anatomy, embryology and other biological disciplines. In recent years, an entirely new body of evidences has come from molecular geneological analysis.

There are a number of approaches to the study of molecular phylogeny. The degree of similarity of a protein in related species can be measured by means of immunological methods. Gel electrophoresis of proteins can also be used to get an estimate of genetic distance among populations of closely related species. The percent nucleotide differences among DNA's in related species can be measured by means of DNA-DNA hybridisation. However, a powerful approach that is making increasingly important contributions to our understanding of biological evolution is based on the phylogenetic schema made on the sequences of proteins and nucleic acids.

In molecular geneological studies, two types of homologous genes are distinguished (Fitch and Margoliash 1970). Orthologous genes are descendants of ancestral genes present in the last common ancestor, and therefore, evolution of orthologous genes reflects the evolution of the species that carry these genes. On the other hand, paralogous genes are descendants of a duplicated gene. The evolution of paralogous genes reflects gene species phylogeny. The underlying assumption in these studies is that genes of basic metabolic importance share the same evolutionary history within an organism, and therefore, the sequences of these genes or those of their products can be used either singly or additively to derive phylogenetic schema and infer from it the evolutionary course. These sequences can also be used to derive a composite tree provided its topology maintains the topology of individual trees.

In our review and discussions, we shall trace the outlines of biological evolution from the phylogenetic trees made on the sequences of proteins and nucleic acids. Wherever necessary, we shall also utilise information derived from morphological, cytological, physiological and biochemical data.

2. Construction of molecular phylogenetic trees

Two types of methods are available for the construction of phylogenetic trees based on sequence data sets. The first type uses matrix numbers which are obtained by a count of

non-matching residues in a set of aligned sequences. The second type is based on the reconstruction of ancestral sequences which generate descendant sequences through minimum number of mutations.

Among the matrix methods, the unweighted pair group (UWPG) method of Sokal and Michener (1958), the average percent standard deviation (APSD) coefficient method of Fitch and Margoliash (1967) and the additive tree method of Moore *et al* (1973b) have been used to derive phylogenetic trees. However, the least squares matrix method of Dayhoff (1976) is particularly useful for distantly related sequences such as those found in the prokaryote kingdom. In this method, a matrix of percentage differences is calculated and corrected for inferred superimposed mutations to yield the 'observed matrix'. Next, for each possible topology, a set of branch lengths which gives a weighted least squares fit of the reconstructed matrix to the 'observed matrix' is determined. The topology which gives the smallest sum of absolute values of the branch lengths is taken as the phylogenetic tree for that set.

The ancestral sequence method is a problem of double minimization of the sum of values of branch lengths on the one hand, and topological configurations on the other. Many variations of the ancestral sequence method including those of Fitch (1971) and Dayhoff (1972) are in current use in many laboratories. In Ahmednagar, we have been deriving trees utilising maximum parsimony (MP) approach outlined by Barnabas *et al* (1972) and developed into a mathematically proven method of Moore *et al* (1973a). Recently, we have developed a simpler variation of the MP method (Barnabas *et al* 1978, 1980) which we currently use. The sequences of proteins and nucleic acids needed for this work are available in the sequence data base compiled by Dayhoff (1972, 1973, 1976, 1978) and Dayhoff and Schwartz (1981).

3. Prokaryote evolution

In recent years, a broad view of prokaryote evolution has been deduced from the phylogenetic trees made on the sequences of proteins and nucleic acids by utilising least squares matrix method (Schwartz and Dayhoff 1978). Recently, Barnabas *et al* (1982) have extended this work to deduce the metabolic innovations in the Precambrian. In these studies, a composite tree made on the sequences of ferredoxins, c-type cytochromes and 5 S ribosomal RNA's was used to infer the evolutionary course. Initially, a phylogenetic tree made on the sequences of 4Fe-4S ferredoxins was affixed with a root by taking into account gene doubling which is present in all ferredoxins. Next, points on the 5 S ribosomal RNA tree were mapped onto the bacterial ferredoxin tree. The 2Fe-2S ferredoxins tree and c-type cytochromes tree were next mapped onto the composite tree maintaining the topology of the composite tree consistent with that of individual trees.

Ferredoxins are iron-sulphur proteins which have an extremely electronegative redox potential as a result of which, they take part in a number of oxidation-reduction reactions. Two major types of ferredoxins can be distinguished depending on the structure of iron-sulphur clusters. The bacterial type contain one or two 4Fe-4S clusters. The sequences of ferredoxins from lostridia, *Peptococcus* and *Megasphaera*, *Chlorobium* (green sulfur bacteria) and *Chromatium* (purple sulfur bacteria) contain two 4Fe-4S clusters, whereas those of *Bacillus* and *Desulfovibrio* contain one 4Fe-4S cluster. These sequences vary from 54 residues in *Peptococcus aerogenes* to 81 residues

in *Bacillus stearothermophilus*. The plant type of ferredoxin, commonly found in chloroplasts, contains one 2Fe-2S cluster. Interestingly, ferredoxins of cyanobacteria (blue-green algae) and *Halobacterium halobium* also contain the plant type of ferredoxins. These sequences vary from 98 residues in *Nostoc muscorum* (blue-green) and 97 in spinach chloroplast to 128 in *Halobacterium halobium*.

The cytochrome c superfamily contains a host of c-type cytochromes which act as redox carriers. In contrast to ferredoxins which act on the hydrogen side of NADP, c-type cytochromes act as electron acceptors of reduced NADP. The sequences of cytochrome c from mitochondria, cytochrome c_2 from *Rhodospirillaceae* (purple nonsulfur bacteria), cytochrome c_{550} from *Paracoccus denitrificans*, cytochrome c_6 from cyanobacteria and chloroplasts, cytochrome c_{551} from *Pseudomonas aeruginosa* and cytochrome c_{555} from *Chlorobium* and *Prosthecochloris* have been utilised for discerning the early biological evolution. They are all homologous proteins but their chain length varies from 82 amino acid residues in cytochrome c_{551} of *Pseudomonas aeruginosa* to 135 residues in *Paracoccus denitrificans*.

Ribosomes which are essential in the biosynthesis of proteins are ribonucleoproteins found in all cell types. They have two subunits of unequal size. In prokaryotes, the subunits have the sedimentation coefficients of 50S and 30S whereas in the eukaryotic cytoplasmic ribosomes, 60S and 40S subunits are present. The 5S ribosomal RNA is associated with the large subunit in both prokaryotes and eukaryotes and may play a role in the non-specific binding of transfer RNA during protein biosynthesis. The 5S ribosomal RNA in different organisms is around 120 nucleotides long.

We have schematically represented the prokaryote evolution and early evolution of eukaryotes in figure 1. This schema combines information from the composite phylogenetic tree made on the sequences of ferredoxins, 5S ribosomal RNA and c-type cytochromes (Schwartz and Dayhoff 1978; Barnabas *et al* 1982) as well as from paleontological evidences and biochemical data. Evidences from other biological disciplines have also been utilised to get insights into the taxonomy of lower eukaryotes.

3.1 *Evolution in the Archaen*

The Precambrian is now divided into three Eons: Hadean, the Archaen and the Proterozoic. The Hadean extended from 4·6 to 3·8 billion years ago; the Archaen from 3·8 to 2·6 billion years ago and the Proterozoic from 2·6 to 0·6 billion years ago. It was the Archaen which saw the appearance of the oldest rocks (Moorbath *et al* 1973) as well as the first life. Various lines of evidence including stromatolites, stable isotope ratios and organic carbon distribution suggest that anoxygenic photosynthesis existed as early as 3·5 billion years ago (Knoll 1982). The primary source of evidence for photosynthesis is the presence of stromatolites in the Pilbara block of Western Australia (Walter *et al* 1980). Similarly, the isotopic evidence in the upper Archaen of the Aldan Shield, Siberia and in the Michipicoten and Woman River banded iron formation of Canada, indicate that sulphate reducers arose between 2·8 to 3 billion years ago (Schidlowski 1979). The transition from the Archaen to the Proterozoic was characterised by large scale growth and stabilisation of continents (Knoll 1982).

Obligate fermentative anaerobes represented by *Megasphera*, *Peptococcus* and lostridia appear near the base of the tree. This result is consistent with what is generally believed—that primitive atmosphere was reducing and that fermentative

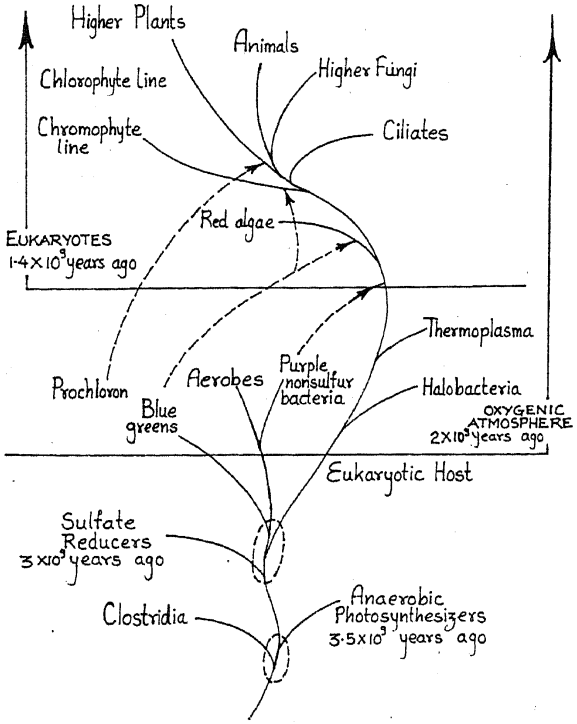

Figure 1. Schematic representation of outlines of evolution of prokaryotes and lower eukaryotes. Prokaryote evolution was inferred from the composite tree based on the sequences of ferredoxins, 5S rRNAs and c-type cytochromes (Barnabas *et al* 1982). In the figure, clostridial line includes *C. pasteurianum* and *C. butyricum* as well as other obligate anaerobes like *Peptococcus aerogenes* and *Megasphaera elsdenii*. Anaerobic photosynthesizers include *Chromatium vinosum* and *Chlorobium limicola*. Sulfate reducer line includes *Desulfovibrio gigas*. *Bacillus stearothermophilus* (not shown) diverges close in time to *Desulfovibrio*. The eukaryotic host branch which is derived from the 5S rRNA sequences from several representative species, also contains *Halobacterium salinarium* and *Thermoplasma*. The 5S rRNA sequence of *Thermoplasma* is determined recently (Luehresen *et al* 1981). The blue-green algae include *Spirulina maxima* and *Synechococcus* ATCC 27144. The aerobes branch include *Pseudomonas fluorescens* and *Azotobacter vinelandii*. Purple nonsulphur bacteria is represented by a few species from *Rhodospirillaceae*. Here, *Rhodopseudomonas globiformis* comes closest to eukaryotic mitochondria. The topological arrangements of protists is shown in the upper part of eukaryotic host branch. The positioning of rhodophytes, ciliates, chromophyte line, and chlorophyte line is done on the basis of morphological, tytological, ultrastructural and biochemical data. The cytochrome c evolutionary tree suggests that animals, plants and fungi must have diverged very close to time (Dayhoff 1978). The serial endosymbiosis is shown in dotted lines. In our schema purple nonsulphur bacterium is the free-living form which gave rise to mitochondria; and the blue-greens are the prokaryote which became plastids in rhodophytes and chromophyte line. However, we have shown 'Prochloron' (Chl a and Chl b containing prokaryote) as the free-living form which gave rise to chloroplasts of chlorophyte line and higher plants (Lewin *et al* 1981).

obligate anaerobes were the first organisms to appear on the earth (Morris 1977). These primitive organisms, like their modern descendants, utilised glycolytic pathways to obtain ATP and reducing equivalents.

Two nitrogen fixing clostridial species *C. Pasteurianum* and *C. butyricum* (figure 1) have the ability to fix dinitrogen. In fact, nitrogen fixing ability is found in a few species of green sulphur bacteria, many species of *Desulfovibrio* and blue green algae as well as in aerobes such as *Rhodospirillaceae* and *Azotobacter*. It is also found in a number of

other aerobes such as *Klebsiella* and *Rhizobium.* Nitrogen fixation by *Klebsiella pneumoniae* is genetically encoded by 17 genes in a cluster of seven or eight operons (Postgate 1982). What is particularly interesting is that *Klebsiella* 'nif' cluster has been transferred genetically to self-transmissible plasmids; and many species which are not known to fix nitrogen have been made to fix it through 'nif' plasmids. These results suggest that 'nif' gene cluster is mobile in nature and therefore it gets distributed in different branches of prokaryotes through gene transfer. Although we cannot rule out such a possibility within closely related groups, it is unlikely that such genetic constructs would find permanent acceptance particularly when the two species are distantly related. Moreover, nitrogen fixation is very expensive in terms of ATP requirements and is extremely sensitive to oxygen. It may be noted that nitrogenase is a iron-sulphur protein like ferredoxins; and thermodynamic data indicate that iron sulphides were plentiful on the primitive earth (Osterberg 1974). In toto, these results suggest that nitrogen fixation is an early acquisition of the prokaryotic stem.

Two anaerobic photosynthesizers, green sulphur bacteria and purple sulphur bacteria, also appear near the base of the tree. In these organisms, photosynthesis is initiated by light reactions in which CO_2 is reduced to carbohydrates at the expense of hydrogen donors such as molecular hydrogen, reduced sulphur compounds or small organic molecules. These organisms contain bacteriochlorophyll (BChl) and carotenoids which harvest light and transfer the excitation energy to a reaction centre containing a BChl-protein complex from which electrons flow through photosystem I either cyclically or non-cyclically (Dutton and Prince 1978).

For assimilation of CO_2, these organisms may use reductive carboxylic acid cycle (reversed Krebs cycle) (Buchanan 1972) and reductive pentose phosphate (Calvin) cycle (Fuller 1978). Recently, with the characterization of ATP citrate lyase (Antranikian *et al* 1982) from green sulphur bacteria, it has become clear that reversed Krebs cycle is operative in this organism. These results collectively suggest that ability to synthesize heme, use photosystem I either cyclically or non-cyclically and reduce CO_2 using either reversed Krebs cycle or Calvin cycle were early innovations of the prokaryote stem.

The green sulphur bacteria uses H_2S as an electron donor and release sulphate into the environment. It is not surprising therefore that *Desulfovibrio* which reduces sulphate to sulphide arose after the divergence of green sulphur bacteria.

3.2 *Evolution in the Proterozoic*

It is now well established that it was around 2 billion years ago that free oxygen began to appear in the atmosphere due to oxygen-releasing photosynthetic activity of cyanobacteria (Cloud 1968). Although the time of the origin of oxygen-releasing photosynthesis is not established, terminal Archaen cratonization must have increased substantially the area available for colonization of benthic cyanobacteria (Knoll 1982). This would indicate that cyanobacteria arose in the early Proterozoic. This Eon also saw the appearance of the first eukaryote around 1·4 billion years ago. This is supported by the finding that in the Beck Spring Dolomite microflora ($1{\cdot}4 \pm 100$ years ago) from Southern California, spheroidal forms exhibiting dark spots are present suggesting the preservation of eukaryotic cell structure (Cloud *et al* 1969). However, the oldest demonstrably nucleated cells are those of the Bitter Springs microflora (0·9 billion years in age) from central Australia (Schopf 1970).

Following the divergence of sulphate reducers, a line referred to as eukaryotic host diverges just prior to the divergence of cyanobacteria and the stem which gave rise to

aerobes (figure 1). The oxygen-releasing photosynthesis of cyanobacteria has photosystems I and II which utilise chlorophyll a and phycobiliproteins as light collecting pigments (Krogmann 1978). Here, photosystem I is an early acquisition of the prokaryotic stem which cyanobacteria has retained; but photosystem II is unique to this group.

The relevant point here is that the ancestor of blue-greens were anaerobes which eventually acquired aerobic respiration when atmosphere became oxygenic due to their own photosynthetic activity. This in turn, would suggest that others which got oxygen respiration acquired it independently. Clearly, oxygen respiration was born not once but many times.

The organisms that diverge from aerobic prokaryotic stem are represented by purple nonsulphur bacteria, *Pseudomonas* and *Azotobacter*. Since these are aerobes, it can be assumed that their common ancestor acquired the ability to respire aerobically. In other words, Krebs cycle and electron transport chain got fully established in the stem line which gave rise to these aerobes. Among these, purple nonsulphur bacteria is a photosynthetic organism which has photosystem I similar to that in the purple sulphur bacteria (Dutton and Prince 1978).

4. Origin of eukaryotes and the proliferation of primitive eukaryotes leading to the present-day eukaryotes

Once aerobic respiration had fully become established, it gave the prokaryotic stem the potential for providing many times more ATP than from fermentation. As a result, much larger and more efficient cell types evolved giving rise to the first eukaryotes.

There have been two theories regarding the origin of eukaryotes. One is that eukaryotes arose as a result of serial endosymbiosis among prokaryotes (Margulis 1981) and the other is that they arose due to compartmentalization of the DNA within the cytoplasm of a single line of prokaryote (Raff and Mahler 1972). In the phylogenetic tree of c-type cytochromes (Schwartz and Dayhoff 1978; Barnabas *et al* 1982), the sequences of cytochrome c_2 of purple nonsulphur bacteria cluster with those of cytochrome c of eukaryotic mitochondria suggesting that ancestors of purple nonsulphur bacteria were probably the free-living organisms that gave rise to mitochondria. Similarly, phylogenies based on 2Fe-2S ferredoxins and cytochrome c_6 sequences identify ancestors of blue-greens as the free-living prokaryotes that gave rise to chloroplasts. The host line for this endosymbiosis is the eukaryotic host line (figure 1). This branch is so called because it is derived from 5S ribosomal RNA from the cytoplasmic ribosomes of eukaryotes. Interestingly, *Halobacterium* and *Thermoplasma* which along with methanogens are classified as Archaebacteria (Fox *et al* 1980) also diverge from this branch. Their position in the eukaryotic host branch is also supported by the observation that parts of the protein synthesizing machinery of *Halobacterium* resemble those of eukaryotes (Bayley and Morton 1978). Similarly, *Thermoplasma* has a flavin-terminated respiratory system resembling the microbodies of the eukaryotic cells (Searcy *et al* 1981). However, on the basis of 16S ribosomal RNA catalogue, their unusual lipids and a number of other unique characteristics, Archaebacteria has been classified as a unique group and not as derived prokaryotes (Woese and Gupta 1981). The arguments of Woese and his collaborators for an ancient archaebacterial and eubacterial divergence are compelling. But, the sequences of 5S ribosomal RNA on our

rooted composite tree shows that *Halobacteria* and *Thermoplasma* diverge from the eukaryotic host line (Dayhoff and Schwartz 1981; Barnabas *et al* 1982). This host line probably separated from the prokaryotic stem around the time of divergence of sulphate reducers (figure 1).

The serial endosymbiotic theory of Margulis (1981) is built around the idea that mitochondria, plastids and cilia arose when smaller prokaryotic cells took up permanent residence inside large prokaryotic cells. Once formed, the eukaryotes began to proliferate and eventually gave rise to the present-day eukaryotes. According to the most widely accepted 5-kingdom classification (Whittaker 1969), the living organisms are classified into five kingdoms. The prokaryotes are represented by *Monera*; and eukaryotes by *Protista, Fungi, Plantae* and *Animalia.*

It seems likely that the first eukaryotes were not unlike the present-day unicellular protists. The best known protists of today include yellow and brown flagellate algae and their allies, the green algae, the red algae, the fungus-like protists, the slime molds, and animal-like protists. The phylogeny of these groups is not fully understood for want of sufficient sequence data. However, a broad view of phylogeny of protists can be obtained by examining the available morphological, ultrastructural, cytological and biochemical data.

The yellow and brown flagellate algae and their allies represent the chromophyte line (Christensen 1964) in which chlorophyll c (Chl c) can be used as a taxonomic marker. The chromophytes contain Chl a and Chl c but Chl b is conspicuously absent. In this line, *Chrysophyceae, Bacillariophyceae, Phaeophyceae* and *Haptophyceae* are phylogenetically related. This is supported by the finding that these groups contain Chl c_1 and Chl c_2 which are not found together elsewhere; and that their carotenoid biosynthetic pathways particularly that leading to fucoxanthin are remarkably similar (Ragan and Chapman 1978). Also, *Xanthophyceae* and *Raphidophyceae* are close to this group; but in them, as well as in *Eustigmatophyceae*, the fucoxanthin pathway is replaced by vaucheriaxanthin pathway (Ragan and Chapman 1978). However, *Raphidophyceae* is considered to be an offshoot of *Xanthophyceae* (Hibberd 1979) while *Eustigmatophyceae* is regarded as a distinct class because of the presence of a unique type of photoreceptor apparatus, absence of golgi bodies from the zoospores and absence of girdle lamella from chloroplasts (Hibberd and Leedale 1972).

The *Cryptophyceae* resemble the chromophyte line in possessing Chl c_2 and the chlorophyte line (Chl b line) in their starch production. The *Cryptophyceae*, like *Rhodophyceae*, contain special accessory pigments, the phycobiliproteins (Glazer *et al* 1976) which are found nowhere else except in cyanobacteria. This implies that both *Cryptophyceae* and *Rhodophyceae* are primitive eukaryotes. However, Ragan and Chapman (1978) indicate that *Cryptophyceae* are probably derived from an ancestral stock near the base of chromophyte and chlorophyte lines.

Dinophyceae is yet another group whose phylogenetic position is difficult to assess. On the one hand, they lack the typical features of eukaryotic chromatin and on the other, they show a characteristic eukaryotic life cycle as well as meiosis (Loeblich 1976). However, the presence of acetylenic xanthophylls and Chl c_2 indicates affinity with chromophytes; and the dinoflagellates could be placed at the base of the chromophyte series (Taylor 1976, 1978; Ragan and Chapman 1978).

Rhodophyceae is regarded as one of the most primitive of the eukaryotes. This is based on the finding that they lack the 9 + 2 structures, typical of eukaryotes. Moreover, their plastids contains Chl a similar to that in cyanobacteria. More significantly,

phycobiliproteins of *Rhodophyceae* show homology with those of cyanobacteria (Glazer *et al* 1976).

Euglenoids are an enigmatic group. They possess Chl a and Chl b like the green algae and have been grouped with them in the chlorophyte series (Christensen 1964) or placed in the division *Euglenophyta* which was kept adjacent to the chlorophyte series (Leedale 1967; Round 1965; Silva 1962). The fact is that the euglenoids may not be closely related to the green algae since they differ in a host of characters. For example, ratios of Chl a to Chl b as well as xanthophylls are different in these groups. Euglenoid chloroplast has an endoplasmic reticulum which is absent in the green algae (Leedale 1967; Taylor 1974). The euglenoids show some resemblance to *Crithidia* in flagellar insertion, microtubular organisation in their periplast and nuclear division (Leedale 1978). Also, the properties of tryptophan biosynthetic enzymes, 28S rRNA and shikimate pathway enzymes are similar between investigated *Euglenophyceae* and protozoa and/or simple *Eumycota* (Ragan and Chapman 1978). It is of special interest to note that two trypanosome flagellates, *Crithidia oncopelti* and *Crithidia fasciculata*, cluster close to *Euglena gracilis* in the cytochrome c evolutionary tree (Dayhoff 1978). Also, *Tetrahymena pyriformis* diverges prior to *Crithidia* in this tree.

The protozoa are not a natural group. They are essentially single-celled eukaryotes and are mostly holozoic or saprozoic. Most of them have a single vesicular nucleus, a few are multinucleate, whereas ciliophora show nuclear dimorphism with a micro and macronucleus (Levine *et al* 1980).

The green algae or the *Chlorophyceae* represent the chlorophyte line of Christensen (1964). They are characterised by the presence of Chl a and Chl b. There is difference of opinion regarding the taxonomic status of green algae (Whittaker and Margulis 1978; Dodge 1974; Ragan and Chapman 1978). This does not concern us here. However, what is interesting is that Stewart and Mattox (1975) have suggested a hypothetical classification based on comparative cytology of many green algae, into two classes representing two different lines of evolution, one of which gave rise to the plants.

One of these classes is the *Chlorophyceae* which includes unicellular, colonial, filamentous and parenchymatous forms. These are characterised by an interzonal spindle which does not persist till the completion of cytokinesis and there is formation of a phycoplast. The key enzyme of glycolate metabolism is glycolate dehydrogenase (Fredrick *et al* 1973). Flagella are inserted anteriorly in the motile cells. Their flagellar basal bodies are associated with four or more relatively narrow, cruciately arranged microtubular roots (Pickett-Heaps 1975). The other class is *Charophyceae* which includes unicellular, filamentous and parenchymatous forms. These are characterised by phragmoplast type of cell division in which interzonal spindle is persistent during cytokinesis and microtubules are laid down at right angles to the plane of cell division. The key enzyme in glycolate metabolism is glycolate oxidase.

Pickett-Heaps and Marchant (1972) have derived the higher plants and also the conjugales from the *Charophyceae* line. Clearly, the *Charophyceae* share with higher plants a phragmoplast system and the possession of glycolate oxidase. Added to this is an observation by Henry and Hall (1977) that charophycean line contains cyanide sensitive, Cu/Zn type of superoxide dismutase similar to that in the cytosol of plants, animals and fungi. This is in contrast to chlorophycean line which contain the cyanide insensitive superoxide dismutase which is prokaryotic in origin.

In figure 1, we have shown three endosymbiotic events that gave rise to plastids in

eukaryotic algae and higher plants. Although the exact number of such events remains unclear, there is little doubt that plastids are polyphyletic in origin (Raven 1970; Schwartz and Dayhoff 1981; Gray and Doolittle 1982). While there are reasons to believe that plastids of rhodophytes and the chromophyte line are cyanobacterial in origin, it is probable that plastids of the chlorophyte line and higher plants arose from 'Prochloron', a non-cyanobacterial oxygen-releasing photosynthetic organism (Lewin 1981). From the cytochrome c_6 phylogeny (Schwartz and Dayhoff 1981) it is evident that there are two cyanobacterial endosymbiotic events. One led to the plastids of the rhodophyte *Porphyra tenera* and the phaeophyte *Alaria esculenta*. The other led to the plastids of *Euglena gracilis* and chrysophyte *Monochrysis lutheri*. We however, include chrysophytes and phaeophytes in the chromophyte line; and hence the two cyanobacterial symbiotic events shown in figure 1 are at the base of rhodophyte and chromophyte line respectively. Also, we favour the view that euglenoid chloroplasts may have arisen due to symbiosis between a chlorophycean alga and a zooflagellate.

There is no agreement on whether the mitochondria arose due to symbiosis or not; and if they are of symbiotic origin, whether they are monophyletic or polypheletic (Raff and Mahler 1972; Dayhoff and Schwartz 1981; Gray and Doolittle 1982). However, in figure 1 we have shown appearance of mitochondria as one endosymbiotic event.

5. Geneological analysis of haemoglobin and evolution of vertebrates

It is conceivable that with the widespread adoption of multicellularity in the evolving eukaryotes, the need for efficient oxygen transport within an organism also arose. This need was eventually met by oxygen-carrying proteins such as haemoglobins and myoglobins. The geneology of these oxygen-carrying proteins can be discerned through their amino acid sequences.

Haemoglobin is a monomer in many species of invertebrates in which it is found (Wittenberg *et al* 1965) as well as in the root nodules of leguminous plants (Kellin and Wang 1945). It is also a single-chained protein in the primitive jawless vertebrate (lamprey). The myoglobin found in the vertebrate muscle tissue is also a monomeric protein. On the other hand, the haemoglobin of vertebrate is a tetramer containing two types of chains. In vertebrates, different kinds of haemoglobins appear during development and growth. In man, the principal kind of haemoglobin in the adult contains two alpha and two beta chains. The minor adult haemoglobin contains two alpha and two delta chains. At the time of birth, a foetal type containing two alpha and two gamma chains is present. Early in foetal life, alpha-like zeta chains and beta-like epsilon chains also appear for a brief period. All these polypeptide chains are paralogous proteins.

The haemoglobin-myoglobin family of proteins contains the same heme group, iron (II) protoporphyrin IX (Perutz 1976), in each polypeptide chain. Therefore, it is the variation in the sequence of the polypeptide chains that distinguishes one globin from another. The x-ray crystallographic data support the sequence homology found in different polypeptide chains of haemoglobins. In fact, three-dimensional structures of haemoglobins of root nodules of leguminous plants (Vainshtein *et al* 1975) an annelid *Glycera dibranchiata* (Padlan and Love 1968), an insect *Chironomus thummi* (Hubber *et al* 1971), a lamprey *Petromyzon marinus* (Hendrickson *et al* 1973), as well as that of

mammalian myoglobins and haemoglobin chains (Perutz *et al* 1960) are remarkably similar.

Ingram (1961) suggested an evolutionary scheme for haemoglobin. According to this scheme, haemoglobin arose by duplication of a single ancestral gene followed by mutation and translocation; and by repetition of the same process, genes for alpha, beta, gamma and delta chains arose in that order. The phylogenetic trees derived by utilising UWPG method and APSD coefficient method by Barnabas *et al* (1971) and Goodman *et al* (1971) as well as those derived by utilising MP method by Goodman *et al* (1974) confirmed Ingram's scheme and further traced the evolutionary history of haemoglobins.

The Precambrian and Cambrian globin ancestors descending from the invertebrate-chordate ancestor to the primordial vertebrate ancestor were single-chained globins. The myoglobin-haemoglobin gene duplication took place in the primitive vertebrates, after the divergence of the lampreys. The alpha-beta gene duplication took place in the common ancestor of teleosts and tetrapods. By the Silurian period, about 400 million years ago, the last common ancestor of teleost and tetrapods, as a result of preceeding alpha-beta gene duplication, had already acquired the ability to form tetrameric haemoglobins. About 250 million years later, in the primitive therian mammalian lineage, after the divergence of monotremes but before the marsupial-placental split, duplication in the beta gene separated the ancestral locus for the primate gamma and the marsupial beta chains from the locus for typical mammalian beta chains. The beta globin gene further underwent a gene duplication in the higher primate stem giving rise to the delta globin gene.

The sequence organisation of globin genes in man has been established. The sequence arrangement of alpha-like genes which are located on chromosome 16 is 5′-ζ2-ζ1-$\psi\alpha$1-α2-α1-3′ and that of beta-like genes which are located on chromosome 11 is 5′-$\psi\beta$2-ε-$^{G}\gamma$-$^{A}\gamma$-$\psi\beta$1-δ-β-3′ (Lauer *et al* 1980; Dayhoff *et al* 1981). The presence of intra-species tandem duplicates at the alpha chain locus is widespread in animal species. Recently, a suggestion has been made that regions coding for alpha chains have been in duplicate since the time of amniote ancestor and that duplicate alpha chain genes have evolved in concert over evolutionary time (Zimmer *et al* 1980).

A phylogenetic tree of representative globin sequences is shown in figure 2. By combining information from this tree as well as from the three-dimensional structures of globins and sequence organization of globin genes, functional innovations of haemoglobins over evolutionary time could be discerned.

Two parameters which give a measure of functional properties of haemoglobins are cooperativity in oxygen-binding and oxygen affinity. The former can be expressed in terms of Hill's coefficient n; whereas the latter can be measured in terms of P_{50} value which is inversely related to it.

The three-dimensional structural analysis of haemoglobins has shown that there are a number of amino acid residues which have defined functional roles (Perutz *et al* 1968). Among them, amino acid residues at $\alpha_1\beta_2$ contact sites and those that are involved in C-terminal salt bridge formation are primarily responsible for the cooperativity in tetrameric haemoglobin. Amino acid residues that are involved in heme contact sites, $\alpha_1\beta_1$ contact, and those that are responsible for the Böhr effect (Kilmartin and Rossi-Bernardi 1969) and 2,3-DPG binding (Arnone 1972) also have important roles in haemoglobin function. Moreover, it has been established that haemoglobin tetramer achieves cooperativity in oxygen binding due to its ability to transit between two

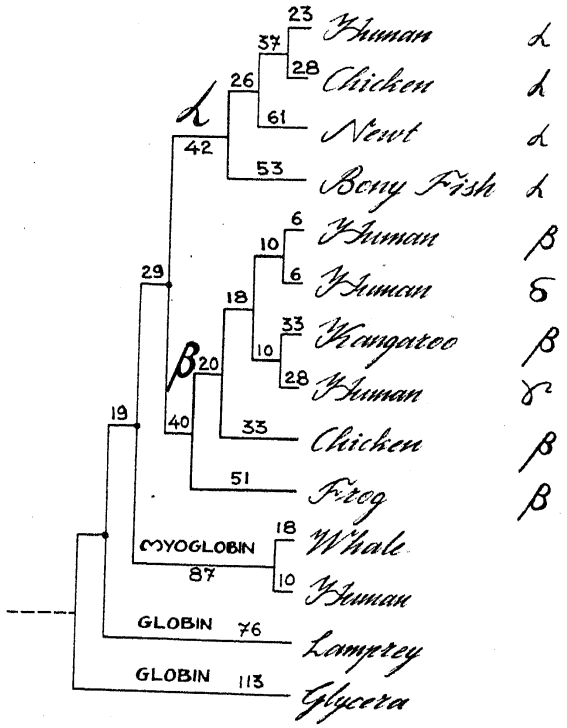

Figure 2. A phylogenetic tree of 14 globin sequences. The tree was drawn by using parsimony method of Barnabas *et al* (1978, 1980). The amino acid sequences are from man (*Homo sapiens*) alpha, chicken (*Gallus gallus*) alpha, newt (*Taricha granulosa*) alpha, bony fish (*Catostomus clarkii*) alpha, human beta, human delta, human gamma, Kangaroo (*Macropus giganteus*) beta, chicken beta, frog (*Rana catesbeiana*) beta, human myoglobin, sperm whale (*Physeter catadon*) myoglobin, lamprey (*Petramyzon marinus*) globin and glycera (*Glycera dibranchiata*) globin. Numbers represent nucleotide replacements between ancestor and descendant sequences, derived from MP method.

conformational states. The low oxygen affinity state, the T state, corresponds to deoxyhaemoglobin whereas the high oxygen affinity (R) state refers to oxyhaemoglobin (Perutz and TenEyck 1971).

In globin genes, there are three coding blocks (exons) separated by two noncoding sequences (introns) (Jeffreys and Flavell 1977; Tilghman *et al* 1978). Eaton (1980) analysed the range of haemoglobin function in terms of the three exons. It was shown that the central exon-encoded fragment has most of the $\alpha_1\beta_2$ contact sites and heme contacts, whereas $\alpha_1\beta_1$ contact sites are located predominently in the right exon-encoded fragment. The other functionally important sites are distributed in the left and the right fragments. Recently Gō (1981) has suggested that there are four domains in globins (F1 to F4). The F1 and F4 correspond to the left and right exons whereas F2 and F3 correspond to the central exon which could be a fusion product of two exons. Interestingly, Jensen *et al* (1981) showed that in soybean leghaemoglobin gene an additional intron is found in the central exon between residues 68 and 69.

In figure 2, the most ancestral branching point is the ancestor of *Glycera* (invertebrate) and lamprey. The haemoglobins of *Glycera* (Vinogradov *et al* 1970) as well as leghaemoglobin (Imamura *et al* 1972) and vertebrate myoglobin (Rossi-Fanelli and Antonini 1958) exhibit a value of n close to unity and a small P_{50} value. However, lamprey haemoglobin exhibits a value of $n = 1{\cdot}2$ at pH 6·8 (Wald and Riggs 1951), an

unusual property for monomeric haemoglobin. The weak cooperativity of lamprey haemoglobin is due to its capability to form dimers when deoxygenated and dissociate when saturated with oxygen (Briehl 1963). Li and Riggs (1970) after comparing the amino acid residues forming the $\alpha_1\beta_2$ contact area in horse haemoglobin with residues at analogous positions in lamprey haemoglobin suggested that a contact similar to the $\alpha_1\beta_2$ contact area of mammalian haemoglobins is present in lamprey haemoglobin. Love *et al* (1971) showed that the overall configuration of the C-helix, CD corner, and FG corner of lamprey haemoglobin is similar to that of mammalian haemoglobins. Also, Hendrickson *et al* (1973) suggested the involvement of the C-helix and FG corner in the formation of lamprey homodimers. These results collectively suggest that rudiments of the $\alpha_1\beta_2$ contact sites were present in the vertebrate globin ancestor. Functional innovations of haemoglobins over evolutionary time have been discussed in detail in a recent paper by Furtado *et al* (1982).

From the available information, we could also deduce the nature of primordial globin gene. Since most of the heme contact sites are in the F2 and F3 regions, it is probable that F2 and F3 domains acted together in the primordial haemoglobin to form the hydrophobic heme pocket. An indirect support for this contention comes from the studies of Craik *et al* (1981) who showed, by excising the central exon-encoded fragment (F2 and F3 together), that it has structural potential for providing a tight and specific binding site for heme and the fit is sharpened by the addition of the N and C-terminal fragments. Of particular interest, is the observation that the complex between the central exon-encoded fragment and heme shows characteristics of cytochromes b_5 and c_{551} (Argos and Rossmann 1979). This suggests an ancient function for the fragment encoded by the central exon. This function is unlikely to be an oxygen-carrying property, since the cytochromes have their origin in the Archean when the atmosphere was anaerobic (Cloud 1968). Since the heme binding site (F2 His) is located in the F3 region, we consider this domain as the most ancestral one. It is conceivable that with the establishment of free oxygen in the atmosphere and the eventual appearance of multicellular eukaryotes, the need for a good oxygen carrying protein became necessary. One could visualise the single-chained primordial globin having the four domains meeting this need. However, it is likely that the central exon was interrupted by an intron in primordial globin gene as in leghaemoglobin gene. Since, introns increase the probability of recombination (Gilbert 1978), the presence of an intron interrupting the central exon would be disadvantageous to organisms where haemoglobins play a crucial role as in higher vertebrates. On the basis of this, the presence of an intron interrupting central exon would be ancient in origin. This intron was eliminated in higher vertebrate globin genes while it was retained in the leghaemoglobin gene. It is interesting to note that the myoglobin gene branch which diverges after the separation of the lamprey branch (figure 2) does not contain an intron interrupting the central exon (Blanchelot *et al* 1983).

As evolutionary time advanced, a gene duplication in the vertebrate globin stem separated the myoglobin branch from the haemoglobin branch. The myoglobin apparently lost the weak cooperativity which its globin ancestor possessed. This is probably due to replacements of uncharged amino acids at some of the $\alpha_1\beta_2$ contact sites by charged amino acids (C3 Glu, C7 Lys, CD2 Asp, FG3 His, and FG4 Lys).

The next major event in the evolutionary history of globins is a gene duplication that separated the alpha and beta globin branches (figure 2). Subsequently, evolving monomeric haemoglobins developed the capability to form tetramers as well as to

transit between two conformational states. The interactions constraining the haemoglobin in the T state are salt bridges formed by the C-terminal amino acid residues, removal of which results in drastic lowering of cooperativity (Kilmartin and Hewitt 1971). The residues involved in salt bridge formation are HC3 α Arg, HC3 β His, NA1 α Val, H9 α Asp, C5 α Lys and FG1 β Asp. From among these, the amino acid residues at HC3 α, H9 α, C5 α and FG1 β evolved only after the alpha-beta divergence.

Among the $\alpha_1\beta_2$ contact sites, some are conserved during evolution, while some others such as C3, C5, C6, C7, CD2, FG4, G3 and HC3 change in the alpha and beta globin branches, each branch conserving a specific amino acid. Replacement of specific amino acid at these sites in either alpha or beta chains, results in increased oxygen affinity, and decreased cooperativity. For example, both HbJ Capetown, FG4 α (Arg $\rightarrow$ Gln) and Hb Malmo FG4 β (His $\rightarrow$ Gln) have high oxygen affinity and low cooperativity with n values of 2·2 and 1·58 respectively (Bellingham 1976). These results suggest that the $\alpha_1\beta_2$ contact sites that evolved after alpha–beta divergence have played a significant role in the development of the two alternate conformations of $\alpha_1\beta_2$ contact area essential for the T $\rightarrow$ R transition; and hence for enhanced cooperativity in oxygen binding. It may be noted that the T state has preference for proton binding over the R state. When this happens T state gains free energy and R $\rightarrow$ T equilibrium is shifted in favour of T state. Clearly, lowering of pH (when more protons are released) causes a shift to the T state (Böhr effect). Thus, in the vicinity of actively metabolising tissues where the pH is lower than in the lungs, oxygen can be unloaded. Conversely, in the lungs where CO_2 is removed from the blood, the pH rises and causes a shift in equilibrium to the R state and as a consequence, haemoglobin can take up oxygen. This behaviour of the haemoglobin tetramer is advantageous to large and mobile animals since by this mechanism efficiency of oxygen transport is greatly increased.

6. Mammalian phylogeny

An overview of evolution of diverging mammals can be obtained by constructing phylogenetic trees based on orthologous globin sequences (Barnabas 1976). The alpha and the beta chain sequences of mammalian globins were combined and were treated as one orthologous set for deriving phylogeny (figure 3).

The living mammals consist of monotremes, marsupials and placentals. Monotremes, though unknown in the fossil record till the Pleistocene epoch, are known to be the most primitive of mammals. On the other hand, marsupials and placentals appear for the first time at the end of Cretaceous (Romer 1966). The final phase of mammalian radiation took place in the Cenozoic era during which time the primitive mammals evolved along various lines giving rise to the successful mammals of modern age. From figure 3, it is evident that the basic pattern of globin gene evolution of mammals resembles closely the traditional taxonomy of mammalian species in which these globin genes are present. In the line leading to mammalian stem, the first branch to separate is that of echidna (*Monotremata*) followed by the marsupial (Kangaroo) branch. In the eutherian (placental) line of descent, *Carnivora* branch (dog) and *Rodentia* branch (mouse) separate in that order prior to the ungulate-primate split. From the fossil evidence, it appears that carnivores arose from a carnivorous palaeoryctoid stock, and that the modern carnivore group developed from their miacoid ancestors by Oligocene time (McKenna 1969). The origin of rodents is obscure

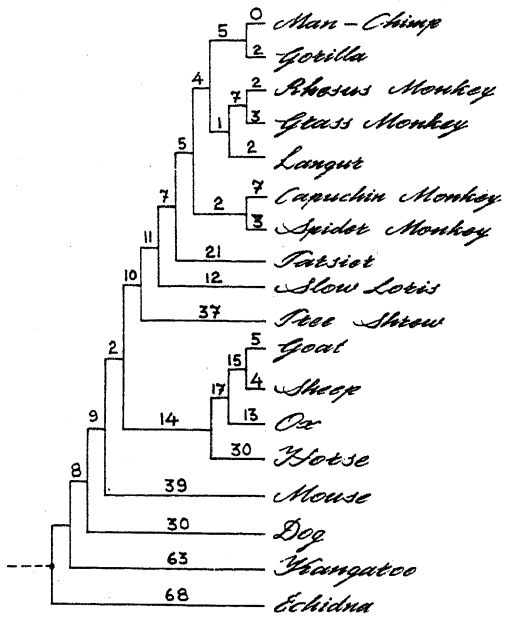

Figure 3. A phylogenetic tree of combined alpha-beta chain sequences of haemoglobins. The amino acid sequences are from man, chimpanzee (*Pan troglodytes*), gorilla (*Gorilla gorilla*), rhesus monkey (*Macaca mulatta*), grass monkey (*Ceropithecus aethiops*), langur (*Presbytis entellus*), spider monkey (*Ateles geoffroyi*), capuchin monkey (*Cebus apella*), tarsier (*Tarsius*), slow loris (*Nycticebus coucang*), tree shrew (*Tupaia glis*), goat (*Capra sp*), sheep (*Ovis aries*), ox (*Bos taurus*), horse (*Equus caballus*), mouse (*Mus musculus*), dog (*Canis familiaris*), kangaroo and echidna (*Tachyglossus aculeatus*).

in the fossil record. However, when they appear, they possessed many features shared with early primates and primate-like insectivores (McKenna 1969; Romer 1966).

In the ungulate line of descent, perrisodactyl branch (horse) separate prior to branches of artiodactyls. In the artiodactyl line represented by the ruminants (ox, sheep and goat), ox branch separates prior to branches of caprines (sheep and goat). The most primitive of the ungulates, the condylarths, appear in the sediments of early Paleocene age (Colbert 1970). The ancestors of artiodactyls were also condylarths. Two major radiations of the artiodactyls appear to have taken place in succession: one throughout Eocene and the other at the close of Eocene and beginning of Oligocene (Van Vallen 1971).

The order primates includes not only lemurs, monkeys and apes but man himself. It seems clear that primates arose from insectivore ancestors. Ancestral primates had become defined by the beginning of Cenozoic times. An initial radiation of primitive primates was in the Paleocene and Eocene and these are represented by lemurs, lorises and tarsiers. Of special interest are the primate-like forms which appear in the fossil record in the middle Paleocene (Romer 1966). After the divergence of these prosimians, there was another radiation in the upper Eocene and following geologic periods; and these are represented by monkeys, apes and man (Colbert 1970). In figure 3, the most ancestral branch in the primate line of descent is that of the tree shrew. The living tree shrew is often regarded as a primate although there is a controversy about it. The next branch to separate from the primate stem is the prosimian branch represented by slow loris. This is followed by the tarsier branch which is placed between lorises on the one hand, and the anthropoids on the other. This arrangement is in agreement with Colbert

(1970) who places tarsier in a separate suborder *Tarsoidea.* However, the position of tarsier in primate classification is still controversial. The anthropoid branch further separates into platyrrhines (capuchin monkey and spider monkey) and catarrhines. The latter branch further separates into Old World monkeys (langur, grass monkey and rhesus monkey) and hominoids (gorilla, chimpanzee and man). In the Old World monkey line, langur which is a colobine monkey separates from the cercopithecines (rhesus monkey and grass monkey). Clearly, the primate phylogeny derived from haemoglobin sequences closely parallels the traditional taxonomy of primates except that man and chimpanzee appear as a single line. In fact, this single line is due to the fact that chimpanzee carries the same adult haemoglobin as man. This indicates that either the common ancestor of man and chimpanzee lived closer in time than what is generally accepted or the alpha and beta chains of haemoglobins have not changed in man and chimpanzee since the time of their common ancestor. However, from paleontological evidences it is clear that the fossil apes known as dryopithecines which appeared in the middle of Miocene times were ancestral to apes and man (Romer 1966). Also, a form known as *Ramapithecus* which appeared in the Siwalik hills of the Punjab, 12–14 million years ago, may be the first step towards man (Day 1973).

Acknowledgements

One of the authors (JB) thanks the Jawaharlal Nehru Memorial Fund, Teen Murti House, New Delhi, for the award of Jawaharlal Nehru Fellowship. His thanks are also due to Dr L K Doraiswamy, Director, NCL, Pune, for permission to work at NCL. This work is supported by grants from the DST, New Delhi.

References

Antranikian G, Herzberg C and Gottschalk G 1982 Characterisation of ATP citrate lyase from *Chlorobium limicola; J. Bacteriol.* **152** 1284–1287

Argos P and Rossman M G 1979 Structural comparisons of heme binding proteins; *Biochemistry* **18** 4951–4960

Arnone A 1972 X-ray diffraction study of binding of 2,3-diphosphoglycerate to human haemoglobin; *Nature (London)* **237** 146–149

Barnabas S 1976 *Variation and comparative evolution of haemoglobins in some species of three mammalian orders: Primates, rodents and ungulates during their descent*, Ph.D. thesis, Poona University, Pune

Barnabas J, Goodman M and Moore G W 1971 Evolution of haemoglobin in primates and other therian mammals; *Comp. Biochem. Physiol.* **B39** 455–482

Barnabas J, Goodman M and Moore G W 1972 Descent of mammalian alpha globin chain sequences investigated by the maximum parsimony method; *J. Mol. Biol.* **69** 249–278

Barnabas J, Mathew P A, Ratnaparkhi M V and Barnabas S 1978 Evolutionary genetics of mammalian haemoglobins and myoglobins during Cenozoic history; *Indian J. Biochem. Biophys.* **15** 388–400

Barnabas J, Ratnaparkhi M V and Mathew P A 1980 Molecular phylogeny of mammalian globin chain sequences deduced by using a parsimony probe, in *Biomolecular structure, conformation, function and evolution* (ed) R Srinivasan (Oxford and New York: Pergamon Press) Vol 2 635–642

Barnabas J, Schwartz R M and Dayhoff M O 1982 Evolution of major metabolic innovations in the Precambrian; *Origins of Life* **12** 81–91

Bayley S T and Morton R A 1978 Recent developments in the molecular biology of extremely halophilic bacteria; *CRC Crit. Rev. Microbiol.* **6** 151–205

Bellingham A J 1976 Haemoglobins with altered oxygen affinity; *Br. Med. Bull.* **32** 234–238

Blanchelot A, Wilson V, Wood D and Jeffreys A J 1983 The seal myoglobin gene: an unusually long globin gene; *Nature (London)* **301** 731–734

Briehl R W 1963 The relation between the oxygen equilibrium and aggregation of subunits in lamprey haemoglobins; *J. Biol. Chem.* **238** 2361–2366

Buchanan B 1972 in *The Enzymes* (ed) P D Boyer (New York: Academic Press) Vol 6 193–216

Christensen T 1964 The gross classification of algae in *Algae and man* (ed) D F Jackson (New York: Plenum Press) 59–64

Cloud P 1968 Atmospheric and hydrospheric evolution on the primitive earth; *Science* **160** 729–736

Cloud P E Jr., Licari G R, Wright L A and Troxel B W 1969 Proterozoic eukaryotes from eastern California; *Proc. Natl. Acad. Sci. U.S.A.* **62** 623–630

Colbert E H 1970 *Evolution of the vertebrates*, (New Delhi: Wiley Eastern Private Ltd.)

Craik C S, Buchman S R and Beychok S 1981 O_2 binding properties of the product of the central exon of β-globin gene; *Nature (London)* **291** 87–90

Day M D 1973 The fossil history of man, article 13 in *Readings in genetics and evolution* (ed) J J Head (Oxford: University Press)

Dayhoff M C 1972, 1973, 1976, 1978, *Atlas of protein sequence and structure*, National Biomedical Research Foundation Georgetown University Medical Centre, Washington, DC

Dayhoff M O and Schwartz R M 1981 Evidence on the origin of eukaryotic mitochondria from protein and nucleic acid sequences; *Ann. N.Y. Acad. Sci.* **361** 92–104

Dayhoff M O, Schwartz R M, Chen H R, Barkar W C and Orcutt B C 1981 *Nucleic acid sequence data base*, (Georgetown University Medical Center, Washington, DC National Biomedical Research Foundation) Vol 1

Dodge J D 1974 Fine structure and phylogeny in the algae; *Sci. Prog. Oxf.* **61** 257–274

Dutton P L and Prince R C 1978 in *The photosynthetic bacteria*, (eds) R K Clayton and W R Sistrom (New York: Plenum Press) 525–576

Eaton W A 1980 The relationship between coding sequences and function in haemoglobins; *Nature (London)* **284** 183–185

Fitch W M 1971 Towards defining the course of evolution minimum change for a specific tree topology; *System. Zool.* **20** 406–416

Fitch W M and Margoliash E 1967 Construction of phylogenetic trees; *Science* 155 279–284

Fitch W M and Margoliash E 1970 The usefulness of amino acid and nucleotide sequences in evolutionary studies; *Evol. Biol.* **4** 67–109

Fox G E, Stackebrandt E, Hespell R B, Gibson J, Maniloff J, Dyer T A Wolfe R S, Balch W E, Tanner R S, Magrum L J, Blackmore R, Zablen L B, Gupta R, Bonen L, Lewis B J, Stahl D A, Luehrsen K R, Chen K N and Woese C R 1980 The phylogeny of prokaryotes; *Science* **209** 457–463

Frederick S E, Gruber P J and Tolbert N E 1973 The occurrence of glycolate dehydrogenase and glycolate oxidase in green plants: an evolutionary survey; *Plant Physiol.* **52** 318–323

Fuller R C 1978 in *The photosynthetic bacteria* (eds) R K Cleyton and W R Sistrom (New York: Plenum Press) 691–705

Furtado M, Mathew P A and Barnabas J 1982 Evolutionary analysis of functional properties in haemoglobins; *Proc. Indian Natn. Sci. Acad.* **B47** 937–947

Gilbert W 1978 Why genes in pieces? *Nature (London)* **271** 501–502

Glazer A N, Appell G S, Hixon C S, Bryant D A and Brown D M 1976 Biliproteins of cyanobacteria and rhodophyta; homologous family of photosynthetic accessory pigments; *Proc. Natl. Acad. Sci. U.S.A.* **73** 428–431

Gō M 1981 Correlation of DNA exonic regions with protein structural units in haemoglobin; *Nature (London)* **291** 90–92

Goodman M, Barnabas J, Matsuda G and Moore G W 1971 Molecular evolution in the descent of man; *Nature (London)* **233** 604–613

Goodman M, Moore G W, Barnabas J and Matsuda G 1974 The phylogeny of human globin genes investigated by the maximum parsimony method; *J. Mol. Evol.* **3** 1–48

Gray M W and Ford Doolittle W 1982 Has the endosymbiont hypothesis been proven? *Microbiol. Rev.* **46** 1–42

Hall D O, Rao K K and Commack R 1975 The iron-sulphur proteins: structure, function and evolution of a ubiquitous group of proteins; *Sci. Prog. (Oxford)* **62** 285–317

Hendrickson W A, Love W E and Karle J 1973 Crystal structure analysis of sea lamprey haemoglobin at 2A resolution; *J. Mol. Biol.* **74** 331–361

Hendrickson W A 1973 Structural effects accompanying ligand change in crystalline lamprey haemoglobin; *Biochem. Biophys. Acta.* **310** 32–38

Henry L E A and Hall D O 1977 Superoxide dismutase in green algae: An evolutionary survey; *Special issue of Plant Cell Physiol.* 377–382

Hibberd D J 1979 The structure and phylogenetic significance of the flagellar transition region in the chlorophyll c containing algae; *Biosystems* **11** 243–261

Hibberd D J and Leedale G F 1972 Observations on the cytology and ultrastructure of the new algal class; *Eustigmatophyceae; Ann. Bot.* **36** 49–71

Huber R, Epp O, Steigmann W and Formanek H 1971 The atomic structure of erythrocruorine in the light of the chemical sequence and its comparison with myoglobin; *Eur. J. Biochem.* **19** 42–50

Imamura T, Riggs A and Gibson Q H 1972 Equilibria and kinetics of ligand binding by leghaemoglobin from soybean root nodules; *J. Biol. Chem.* **247** 521–526

Ingram V M 1961 Gene evolution and the haemoglobins; *Nature (London)* **198** 704–708

Jeffreys A J and Flavell R A 1977 The rabbit beta globin gene contains a large insert in the coding sequence; *Cell* **12** 1097–1108

Jensen E O, Paludan K, Hyldig-Nielson J J, Jorgensen P and Marker K A 1981 The structure of a chromosomal leghaemoglobin gene from soybean; *Nature (London)* **291** 677–679

Keilin D and Wang Y L 1945 Haemoglobin in the root nodules of leguminous plants; *Nature (London)* **155** 227–229

Kilmartin J V and Rossi-Bernardi L 1969 Inhibition of CO_2 combination and reduction of the Bohr effect in haemoglobin chemically modified at its alpha amino groups; *Nature (London)* **222** 1233–1246

Kilmartin J V and Hewitt J A 1971 The effect of removal of C-terminal residues on cooperative interactions in haemoglobin; *Cold. Spr. Harb. Symp. Quant. Biol.* **36** 311–314

Knoll A H 1982 Tectonics, productivity and ecosystems on the early earth, The North American Paleontological Convention, *Proc.* **2** 307–311

Krogmann D W 1978 in *The biology of blue green algae*, (eds) N G Carr and B A Whitton (Berkeley: Univ. California Press) 80–98

Lauer J, Shen C K J and Maniatis T 1980 The chromosomal arrangement of human alpha-like genes, sequence homology and alpha-globin gene deletions; *Cell* **20** 119–130

Leedale G F 1967 *Euglenoid flagellates* (Englewood Cliffs, New Jersey: Prentice Hall)

Leedale G F 1978 Phylogenetic criteria in euglenoid flagellates; *Biosystems* **10** 183–187

Levine N D, Corliss J O, Cox F E G, Deroux G, Grain J, Honigberg B M, Leedale G F, Loeblich III A R, Lom J, Lynn D, Merinfeld E G, Page F C, Poljansky G, Sprague V, Vavra J and Wallace F G 1980 *A newly revised classification of the Protozoa J. Protozool.* **27** 37–58

Lewin R A 1981 Prochloron and the theory of symbiogenesis; *Ann. N.Y. Acad. Sci.* **361** 325–328

Li S L and Riggs A 1970 The amino acid sequence of haemoglobin V from the lamprey *Petromyzon marinus; J. Biol. Chem.* **245** 6149–6169

Loeblich III A R 1976 Dinoflagellate evolution: speculation and evidence; *J. Protozool.* **23** 13–28

Love W E, Klock P A, Lattman E E, Padlan E A and Ward K B Jr., 1971 The structure of lamprey and bloodworm haemoglobins in relation to their evolution and function; *Cold. Spr. Harb. Symp. Quant. Biol.* **36** 349–357

Luehresen K R, Fox G E, Kilpatrick M W, Walker R T, Domdey R, Krupp G and Gross H J 1981 The nucleotide sequence of the 5s *r*RNA from archaebacterium *Thermoplasma acidphilum; Nucl. Acids Res.* **9** 965–970

Margulis L 1981 *Symbiosis in cell evolution* (San Francisco; W H Freeman and Co.)

Mckenna M C 1969 The origin and early differentiation of therian mammals; *Ann. N. Y. Acad. Sci.* **167** 217–240

Moorbath S, O'Nions R K and Pankhurst R J 1973 Early Archaen age for the Isua iron formation, West Greenland; *Nature (London)* **245** 138–139

Moore G W, Barnabas J and Goodman M 1973a A method for constructing maximum parsimony ancestral amino acid sequences on a given network; *J. Theor. Biol.* **38** 459–485

Moore G W, Goodman M and Barnabas J 1973b An iterative approach from the standpoint of additive hypothesis to the dendrogram problem posed by molecular data sets; *J. Theor. Biol.* **38** 432–457

Morris J G 1977 The physiology of obligate anaerobiosis; *Adv. Microb. Physiol.* **12** 169–246

Osterberg R 1974 Origins of metal ions in Biology; *Nature (London)* **249** 382–383

Padlan E A and Love W E 1968 Structure of the haemoglobin of the marine annelid worm *Glycera dibranchiata* at 5·5Å resolution; *Nature (London)* **220** 376–378

Perutz M F 1976 Structure and mechanism of haemoglobin; *Br. Med. Bull.* **32** 195–208

Perutz M F, Muirhead A, Cox J M and Goaman L C G 1968 Three-dimensional Fourier synthesis of horse

oxyhaemoglobin at 2·8Å resolution: The atomic model; *Nature (London)* **219** 131–139
Perutz M F, Rossman M G, Cullis A F, Muirhead H, Will G and North A C T 1960 Structure of haemoglobin; *Nature (London)* **185** 416–422
Perutz M F and TenEyck L F 1971 Stereochemistry of cooperative effects in haemoglobin; *Cold. Spr. Harb. Symp. Quant. Biol.* **36** 295–310
Pickett-Heaps J D 1975 Aspects of spindle evolution; *Ann. N.Y. Acad. Sci.* **253** 352–361
Picket-Heaps J D and Marchant H J 1972 The phylogeny of the green algae: a new proposal; *Cytobios* **6** 255–264
Postgate J R 1982 Biological nitrogen fixation; fundamentals: *Philos. Trans. R. Soc. Lond.* **B296** 375–385
Raff R A and Mahler H R 1972 The non-symbiotic origin of mitochondria; *Science* **177** 575–582
Ragan M A and Chapman D J 1978 *A biochemical phylogeny of the protists* (New York: Academic Press)
Raven P H 1970 A multiple origin of plastids and mitochondria; *Science* **169** 641–646
Romer A S 1966 *Vertebrate paleontology* (Chicago: University of Chicago Press)
Round F E 1965 *The biology of the algae* (London: Edward Arnold)
Rossi-Fanelli A and Antonini E 1958 Studies on the oxygen and carbon monoxide equilibria of human myoglobin; *Arch. Biochem. Biophys.* 77 478–498
Schidlowski M 1979 Antiquity and evolutionary status of bacterial sulphate reduction: sulphur isotope evidence; *Origins of Life* **9** 299–311
Schopf J 1970 Precambrian micro-organisms and evolutionary events prior to the origin of vascular plants; *Biol. Rev.* **45** 319–352
Schwartz R M and Dayhoff M O 1978 Origin of prokaryotes, eukaryotes, mitochondria and chloroplasts; *Science* **199** 395–403
Schwartz R M and Dayhoff M O 1981 Chloroplast origins: inferences from protein and nucleic acid sequences; *Ann. N.Y. Acad. Sci.* **361** 260–272
Searcy D C, Stein D B and Searcy K B 1981 A mycoplasma-like archaebacterium possibly related to the nucleus and cytoplasm of eukaryote cells; *Ann. N.Y. Acad. Sci.* **361** 312–324
Silva P C 1962 Classification of algae, in *Physiology and biochemistry of algae* (ed.) R A Lewin (New York: Academic Press)
Sokal R R and Michener C D 1958 A statistical method for evaluating systematic relationships; *Kansas Univ. Sci. Bull.* **38** 1049–1438
Stewart K D and Mattox K R 1975 Comparative cytology, evolution and classification of the green algae with some consideration of the origin of other organisms with chlorophylls a and b; *Bot. Rev.* **41** 104–135
Taylor F J R 1974 Implications and extensions of the serial endosymbiosis theory of the origin of eukaryotes; *Taxon* **23** 229–258
Taylor F J R 1976 Flagellate phylogeny: a study in conflicts: *J. Protozool.* **23** 28–40
Taylor F J R 1978 Problems in the development of an explicit hypothetical phylogeny of the lower eukaryotes; *Biosystems* **10** 67–89
Tilghman S M, Tiemeir D C, Seidman J G, Peterlin B M, Sullivan M, Maizel J V and Leder P 1978 Intervening sequence of DNA identified in structural portion of a mouse beta globin gene; *Proc. Nàtl. Acad. Sci. U.S.A.* **75** 725–729
Vainshtein B K, Harutyunyan E H, Kuranova I P, Borisov V V, Sosfenov N I, Pavlovsky A G, Grebenko A I and Konareva N V 1975 Structure of leghaemoglobin from lupin root nodules at 5A resolution; Nature *(London)* **254** 163–164
Van Vallen L 1971 Toward the origin of artiodactyls; *Evolution* **25** 523–529
Vinogradov S, Machlik C A and Chao L L 1970 The intracellular haemoglobin of a Polychaeta: Some properties of the haemoglobins of *Glycera dibranchiata; J. Biol. Chem.* **245** 6533–6538
Wald G and Riggs A 1951 The haemoglobin of sea lamprey *Petromyson marinus; J. Gen. Physiol.* **35** 45–53
Walter M R, Buick R and Dunlop J S R 1980 Stromatolites 3400–3500 myr old from the North Pole area, Western Australia; *Nature (London)* **284** 443–445
Whittaker R H 1969 New concepts of kingdoms of organisms; *Science* **163** 150–160
Whittaker R H and Margulis L 1978 Protist classification and the kingdoms of organisms; *Biosystems* **10** 3–18
Wittenberg B A, Briehl R W and Wittenberg J B 1965 Haemoglobins of invertebrate tissues; *Biochem. J.* **96** 363–371
Woese C R and Gupta R 1981 Are archaebacteria merely derived 'Prokaryotes?' *Nature (London)* **289** 95–96
Zimmer E A, Martin S L, Beverley S M, Kan Y W and Wilson A C 1980 Rapid duplication and loss of gene coding for the alpha chain of haemoglobins; *Proc. Natl. Acad. Sci. U.S.A.* 77 2158–2162