

## Eukaryote Evolution: A View Based on Cytochrome *c* Sequence Data

P. J. McLaughlin and M. O. Dayhoff

National Biomedical Research Foundation,  
Georgetown University Medical Center, Washington, D.C.

Received March 19, 1973

**Summary.** We have compared the amino acid sequences of cytochrome *c*'s from 45 species of organisms representing all five kingdoms, including one species each for the Protista and Monera. We have made a phylogeny for these data by reconstructing probable ancestral sequences which generate the present descendants through a minimum number of mutations. Several trials with different data sets produced the same minimal configuration. Assuming the occurrence of no major shifts in mutation acceptance rate, we find an early differentiation between prokaryote and eukaryote stocks. Afterward the eukaryote stem gave rise first to the protozoan flagellate branch and later to the multicellular green plant branch; after this the fungi and multicellular animal stems diverged from each other. A probable ancestral sequence was estimated for each kingdom of multicellular organisms. The basic eukaryote ancestor was probably a non-photosynthetic, heterotrophic flagellate. The photosynthetic apparatus could have been a later symbiotic acquisition in the plant ancestry. The dicotyledons had differentiated into two stocks before the emergence of a monocotyledon line as did the Ascomycetes before the emergence of the Basidiomycetes. The mollusc and chordate lines may have had a common acoelomate ancestor at the divergence of the arthropod stock. The numbers of mutations on all of the branches of the phylogenetic tree were calculated as well as the numbers of mutations and repeated mutations at each amino acid position.

**Key words:** Eukaryote Phylogeny — Cytochrome *c* — Evolutionary Tree — Ancestral Sequences — Mutations.

### Introduction

Our knowledge of evolution is greatly increased by several new sequences of cytochrome *c*, especially the first one from a unicellular organism (Pettigrew, 1972). This protistan is *Crithidia oncopelti*, a trypanosomatid flagellate parasitic in insects. *Crithidia* has a functional mitochondrial cytochrome system containing a cytochrome *c* comparable with those of multicellular organisms. The crithidial protein contains trimethyllysine, as do proteins from green plants and fungi; it reacts with mammalian cytochrome oxidase, although at a reduced rate (Hill, Chan, and Smith, 1971); and the sequence is homologous with other cytochrome *c* sequences. Cytochrome *c* sequences are now known for representatives of the five kingdoms of organisms—the Fungi, the multicellular green plants (Plantae), the multicellular animals

(Animalia), the unicellular organisms (Protista), and the bacteria or prokaryotes (Monera) (Whittaker, 1969). We are attempting to trace the outlines of evolution which differentiated these major groups.

## Technique

For the purposes of working out the evolutionary history of the species and genes with which we are concerned, each amino acid in a sequence is treated as an independent characteristic of the species or population in which the protein is found. In determining these histories we have applied the principle that we will not propose more events than the minimum number necessary to connect the available data and to correlate these data with major pieces of information from sources other than the sequences themselves. A second main premise in our technique is that an ancestor was most likely to have, as any particular amino acid characteristic, that amino acid which is present in the majority of the closest known relatives in the lines of descendants or antecedents. On this basis we can reconstruct probable ancestral sequences from the known sequences of living organisms and determine the numbers of mutations which probably occurred in each line of descent.

We consider an evolutionary history as a tree which represents the lines of descent connecting the sequences from living organisms. A computer program handles a tree as a topological configuration in which each junction of lines is a node representing an ancestor. By comparing the known sequences residue by residue, the program reconstructs an ancestral sequence at each node and determines the location and the minimum possible number of mutations within a given configuration. Branches can be moved from one location to another in the configuration and, for each configuration considered, the number of mutations is totaled. From all of the configurations investigated we select the one having the minimum total number of mutations as the most probable tree. To present a configuration as an evolutionary tree, the number of mutations on each branch is converted to Accepted Point Mutations (PAMs): the percent difference increased by an estimate of the superimposed mutations which probably occurred but are unobserved.

The details of the above technique are given in McLaughlin, Hunt, and Dayhoff (1972) and Dayhoff, Park, and McLaughlin (1972). These sources include a recent improvement over earlier programs (Dayhoff and Park, 1969) which conducts a detailed evaluation for possible amino acid assignments in groups of blanks in connected ancestral sequences.

For the study reported here, we used the evolutionary tree of cytochrome *c* shown in Dayhoff, Park, and McLaughlin (1972) as the initial configuration within each of the three kingdoms of multicellular organisms. The new sequences were added starting with those most closely related to previously known sequences. Groups of sequences in distant parts of the tree were sometimes represented by common ancestral sequences. After all of the new sequences were placed, we reexamined the placement of the first sequences added. Finally, the best configuration for all five kingdoms was resolved and the branch lengths and mutations of the complete tree were calculated. Over a thousand configurations were examined in all of the present series of trials.

## Data

In this study, the sequences are aligned with each other in a manner similar to the cytochrome alignments on pages D-9 and D-367 in Dayhoff, Hunt, McLaughlin, and Barker (1972). In Fig. 1 are aligned sequences representing each kingdom, including some sequences which have been reported recently. This alignment differs from that on p. D-9 (Dayhoff *et al.*, 1972) in the following respects: (1) Newer sequences have

Table 1. Organisms and classification

Species	Common names	Kingdom	Class	Order; Family
<i>Rhodospirillum rubrum</i>		Monera	Schizomycetes	Pseudomonadales
<i>Criithidia oncopelti</i>		Protista	Mastigophora	Protomonadida
<i>Ginkgo biloba</i>	ginkgo	Plantae	Gymnospermae	Ginkgoaceae
<i>Fagopyrum esculentum</i>	buckwheat	Plantae	Dicotyledonae	Polygonaceae
<i>Triticum aestivum</i>	wheat	Plantae	Monocotyledonae	Poaceae
<i>Lycopersicon esculentum</i>	tomato	Plantae	Dicotyledonae	Solanaceae
<i>Helianthus annuus</i>	sunflower	Plantae	Dicotyledonae	Asteraceae
<i>Phaseolus aureus</i>	mung bean	Plantae	Dicotyledonae	Leguminosae
<i>Cucurbita maxima</i>	pumpkin	Plantae	Dicotyledonae	Cucurbitaceae
<i>Brassica napus</i>	rape	Plantae	Dicotyledonae	Brassicaceae
<i>Brassica oleracea</i>	cauliflower	Plantae	Dicotyledonae	Brassicaceae
<i>Sesamum indicum</i>	sesame	Plantae	Dicotyledonae	Brassicaceae
<i>Ricinus communis</i>	castor	Plantae	Dicotyledonae	Pedaliaceae
<i>Gossypium barbadense</i>	cotton	Plantae	Dicotyledonae	Euphorbiaceae
<i>Abutilon theophrasti</i>	abutilon	Plantae	Dicotyledonae	Malvaceae
<i>Saccharomyces oviformis</i>	baker's yeast	Fungi	Ascomycetes	Endomycetales
<i>Candida krusei</i>		Fungi	Deuteromycetes	Moniliales
<i>Debaryomyces hloekeri</i>		Fungi	Ascomycetes	Endomycetales
<i>Ustilago sphaerogena</i>	rust fungus	Fungi	Basidiomycetes	Ustilaginales
<i>Neurospora crassa</i>		Fungi	Ascomycetes	Sphaeriales
<i>Humicola lanuginosa</i>		Fungi	Deuteromycetes	Hyphomycetales
<i>Haematobia irritans</i>	screw-worm fly	Animalia	Insecta	Diptera
<i>Drosophila melanogaster</i>	fruit fly	Animalia	Insecta	Diptera
<i>Manduca sexta</i>	tobacco horn worm moth	Animalia	Insecta	Diptera
<i>Samia cynthia</i>	silkworm moth	Animalia	Insecta	Lepidoptera
<i>Helix aspersa</i>	garden snail	Animalia	Gastropoda	Lepidoptera
<i>Entosphenon tridentatus</i>	pacific lamprey	Animalia	Agnatha	Stylommatophora
<i>Squalus sucklii</i>	puget sound dogfish	Animalia	Chondrichthyes	Petromysontiformes
<i>Cyprinus carpio</i>	carp	Animalia	Osteichthyes	Squaliformes
<i>Katsuwonus vagrans</i>	bonito	Animalia	Osteichthyes	Cypriniformes
<i>Thunnus thynnus</i>	tuna	Animalia	Osteichthyes	Perciformes
<i>Rana catesbiana</i>	bullfrog	Animalia	Amphibia	Perciformes
<i>Chelydra serpentina</i>	snapping turtle	Animalia	Reptilia	Anura
				Chelonia



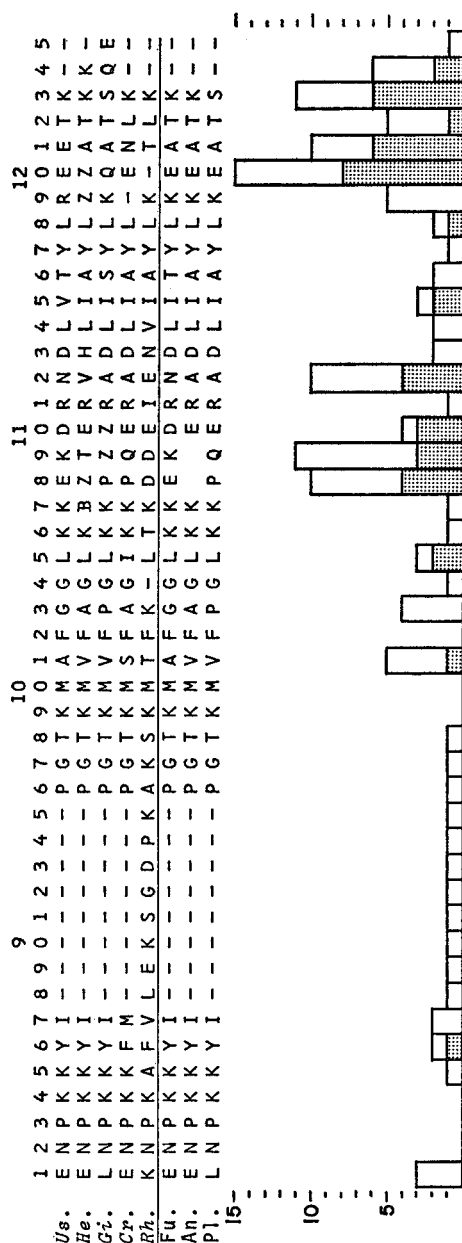


Fig. 1. Cytochrome alignment. One sequence from each kingdom is shown. Below these are probable ancestral sequences for the three multicellular kingdoms. The histogram shows below each position the number of total mutations (whole bars) and the number of repeated mutations (stippled portions) at that position in 45 sequences. A, Ala; B, Asx; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; Y, Tyr; Z, Glx; -, Gap; (blank), undecided

added one more position at each end of the alignment. (2) The regions of gaps at positions 61–63 and 88–95 in the eukaryote sequences have been shifted to the left by one and two positions respectively. These changes were based on improved data for the probability of mutation acceptance (Dayhoff, Eck, and Park, 1972). (3) The gap at position 124 in the *Rhodospirillum* sequence has been shifted from the position immediately to the right of the gap at position 120. This change was due to new sequence information, mainly that of *Crithidia*. The newer sequences are aligned like their near relatives.

Most of the sequence data used here are listed in Dayhoff *et al.* (1972). In addition are the recently reported sequences of *Crithidia* (Pettigrew, 1972), *Gingko* (Ramshaw, Richardson, and Boulter, 1971), buckwheat and cauliflower (Thompson, Richardson, and Boulter, 1971a), tomato (Scogin, Richardson, and Boulter, 1972), pumpkin (Thompson, Richardson, and Boulter, 1971b), rape (Richardson, Ramshaw, and Boulter, 1971), cotton and abutylon (Thompson, Notton, Richardson, and Boulter, 1971), correction of baker's yeast (Lederer, Simon, and Verdiere, 1972), correction of *Candida* (Lederer, 1972), *Humicola* (Morgan, Hensley, and Riehm, 1972), rust (Bitar, Vinogradov, Nolan, Weiss, and Margoliash, 1972), snail (Brown, Richardson, Boulter, Ramshaw, and Jefferies, 1972) and elephant seal (Augusteyn, McDowall, Webb, and Zerner, 1972). The organisms referred to in this work are listed by scientific and common names in Table 1, with the omission of the Mammalia and Aves. The kingdom classifications are those proposed by Whittaker (1969).

## Results and Discussion

### 1. The Detailed Minimum Evolutionary Tree

The results of several trials in which different taxonomic areas were studied in detail have been combined in an overall evolutionary tree of the cytochrome *c* data (Fig. 2). The angle at the bottom of the tree is the earliest point in time within the area of this tree. Theoretically this point should be the divergence of two moneran populations, one of which developed into the surviving bacterial group which includes *Rhodospirillum*, and the other of which eventually evolved into an ancestor of the eukaryotes. The angle is not a node at which an ancestral sequence has been estimated, as are all the other junctions of lines in the tree, but instead is a bend within the line connecting *Rhodospirillum* with the rest of the tree. Branch lengths were averaged down through the tree to locate this bend as the midpoint. To do this we must assume that there has been no major modification of the overall rate of mutation acceptance in cytochromes. No major modifications have been demonstrated in such trees. Furthermore, a technique for evaluating the relationships of distant proteins from separate genes has supported our assumption. Comparisons of some of the above sequences with bacterial cytochromes  $c_{551}$  and  $c_{553}$  do not show any changes in evolutionary rates (Barker and Dayhoff, 1973).

This bend at the base of the tree places a time orientation on the configuration, but no matter what the time orientation, the evolutionary distances and the pattern of branch connections remain the same. The kingdoms Fungi and Animalia are directly connected to each other and

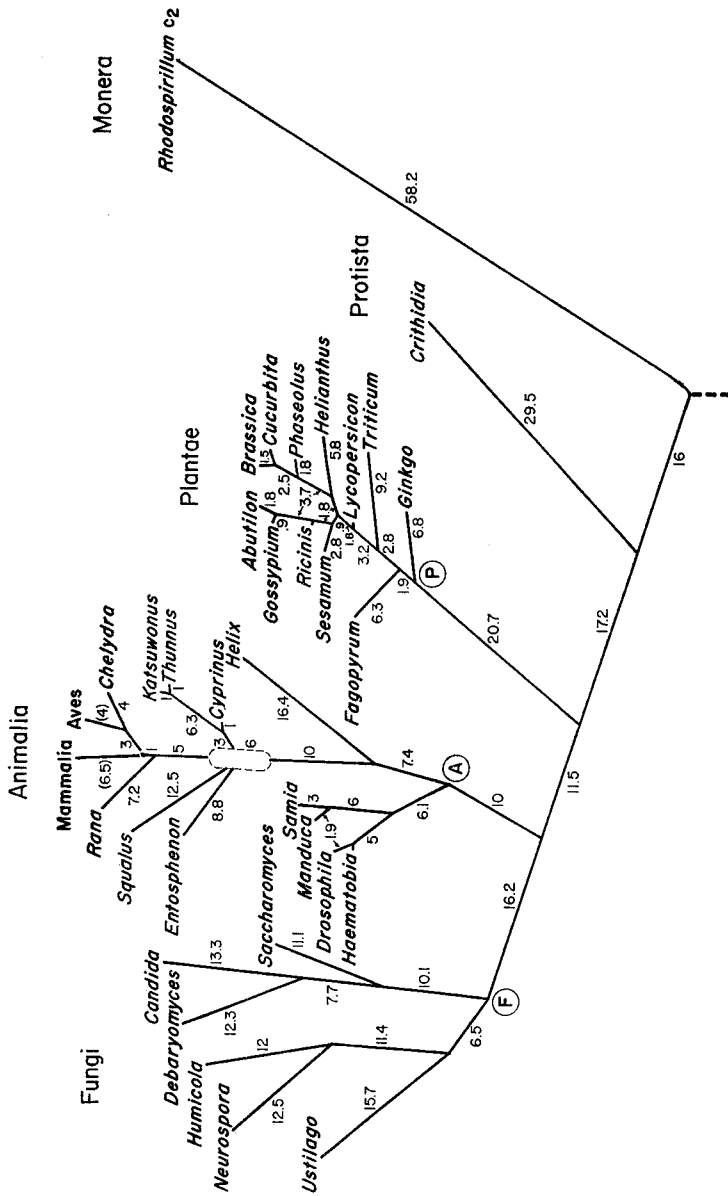


Fig. 2. The detailed cytochrome *c* evolutionary tree. The order of branching for the five kingdoms is the same as configuration 1 in Fig. 4. The progression of time is toward the top of the tree. The lengths of the branches are drawn in proportion to the numbers beside the branches, which are PAMs or Accepted Point Mutations estimated to have occurred on these branches

indirectly connected to the other kingdoms. The three multicellular kingdoms are more closely related to each other than any of them is to the non-multicellular organisms. The moneran and protistan branches are connected directly to each other and indirectly to the other kingdoms, but the greatest amount of evolutionary differentiation occurs on the line to the moneran representative.

Several sequences with only a few mutations between them have been determined for Mammalia and Aves. The two classes are shown on this tree as single lines which are averages of the distances to the different species. The details within these two classes can be found in Dayhoff, Park, and McLaughlin (1972). The dashed oval on the vertebrate stem is an area of uncertainty concerning the order of divergence of branches leading to members of the classes Agnatha, Chondrichthyes, and Osteichthyes. It is not possible to resolve this uncertainty now.

There was an uncertainty about the branching order of *Saccharomyces*, *Candida*, and *Debaryomyces* in an earlier cytochrome tree (Dayhoff, Park, and McLaughlin, 1972). The addition of new sequence information has made the branching order shown here a minimum score and all others slightly higher. This minimum configuration, however, does not accord with the taxonomic organization of these fungi—there are both Ascomycetes and Deuteromycetes on both major stems of the fungus tree. This is probably due to the artificial nature of the class Deuteromycetes, most members of which are undoubtedly related to various members of the Ascomycetes. The evidence does show a divergence of two lines in the Ascomycetes before the differentiation of a member of the Basidiomycetes from one line.

The Fungi and Plantae areas of the tree appear to have different patterns of branching. In the Fungi there is a relatively short main stem and long branches to the individual sequences, whereas in the Plantae there is a long main stem with short branches to the plant sequences and short distances between these branchings. All of the studied species of Plantae are in the phylum Tracheophyta and have diverged from each other rather recently in evolutionary time. None of the groups which differentiated early in plant evolution (mosses, ferns, etc.), and thus would be expected to connect to earlier parts of the long plant stem, are represented in the cytochrome information. Conversely, the fungal groups shown here must have differentiated early in the evolution of this kingdom.

Recently Boulter, Ramshaw, Thompson, Richardson, and Brown (1972) worked out a phylogenetic tree for *Ginkgo* and 14 dicotyledonous species using the ancestral sequence technique described in Dayhoff and Park (1969). We have conducted trials with the same group of sequences using our recently improved ancestral sequence technique and have arrived at the same minimum configuration as Boulter's group in the placement of branches. Our work differs from theirs in the lengths of some of the branches. The differences may be due to our modification of the technique or to interpretation of the spinach, niger and elder sequences, which were presented by Boulter *et al.* (1972) only as differences from other sequences and have not been published elsewhere. These three sequences are not shown on the final tree because of this uncertainty. We have also conducted trials in which: (1) we added to the above sequences one representing a probable common



ancestor of the Fungi and Animalia; (2) the *Triticum* sequence was added to the previous 16 sequences; (3) the ancestral sequence and *Triticum* were present but spinach, elder and niger were removed; (4) the sequences used in (3) were used without the ancestral sequence but with several representative sequences from the rest of the tree. As a result of these trials we developed the same minimum configuration as Boulter's group but did not get minimum scores, as they did, if spinach and *Fagopyrum* were placed in other locations. However, we did find in some trials that a minimum score was also obtained if *Ginkgo* were placed between *Fagopyrum* and *Triticum*. This second configuration is contradictory to fossil and morphological evidence and may be due to a combination of the very diverse evolutionary distances in this region and the rather high rate of repeated mutations in the Plantae (see below). Lawrence (1951) reviewed several schemes of tracheophyte phylogeny and noted that "Opinions have differed as to whether the monocotyledonous plants are more advanced or more primitive than the dicotyledonous plants". Some authors (Hutchinson, 1964; Takhtajan, 1969) consider the dicotyledonous plants to have diverged into two or more stocks before the differentiation of a monocotyledonous ancestor from one of these stocks. Our results on the location of the *Triticum* branch, the only monocotyledon sequence, agree with this general idea of plant phylogeny.

A minimum number of mutations was found with the branch to the single molluscan sequence, *Helix aspera*, located on the stem leading to the chordates after the divergence of the insect stock (Fig. 2). This same result was found by Brown, Richardson, Boulter, Ramshaw, and Jefferies (1972) in forming a tree including a somewhat different group of sequences. Our trials revealed that two other locations of the molluscan branch—on the insect line or on the common animal stem before the divergence of insect and chordate lines—have only slightly higher scores than the minimum, and both have the same score; thus this placement is not very certain.

The several theories of invertebrate phylogeny based on morphology and ontogeny (summarized in Hyman, 1951; Kerkut, 1960) propose either that the molluscs arose from the same line as the arthropods or that the molluscs diverged from the common arthropod-chordate stem. These theories are generally based on a major division into Protostomia (represented here only by insects and a mollusc) and the Deuterostomia (represented here by the chordates), which are characterized by two different systems of embryological development. However, there are many exceptions to the characteristics for each group. In discussing the origin of the molluscs, Vagvolgyi (1967) proposed that the common ancestor of the mollusc and annelid (and presumably arthropod) lines was a flatworm-like, non-segmented, acoelomate animal with spiral cleavage and a trochophore larva. He argues that segmentation was not a primary but a later development in molluscs and that coelom development is not homologous in

different lines of the animal kingdom. Perhaps we might consider that a burst of experimentation in the early ancestors of molluscs gave rise not only to the molluscan style of coelom formation but also to another type—the enterocoel of the chordate line. It may be that the cleavage methods of the Deuterostomia were a new development on this line after an early divergence of the Protostomia into an annelid-arthropod line and a mollusc line. It is unlikely that the questions of invertebrate phylogeny can be decided on anatomical or embryological evidence alone, but it is possible that future knowledge of protein, and nucleic acid, sequences from a greater variety of invertebrate phyla may supply the answers.

The computer program calculates the number of total mutations and the number of repeated mutations at each position in an alignment of sequences. These numbers are shown in histogram form in Fig. 1. The number of repeated mutations includes back mutations and parallel mutations. It is derived by subtracting the number of different amino acids minus one from the number of total mutations calculated throughout the tree at that position. These numbers were calculated from a configuration linking 45 sequences.

The total number of mutations counted at all positions is 477 and the number of repeated mutations at all positions is 168. Thus 35 % of the mutations are repeated mutations. A configuration with 15 distantly related sequences from all parts of the tree had 22 % repeated mutations in a total of 343 mutations. As might be expected, these figures indicate that many repeated as well as unique mutations are not revealed by comparison of small numbers of sequences, and that as more sequences are included in a configuration increasingly greater numbers of repeated mutations are revealed among the various branches.

There are 17 positions in the alignment at which no mutations have been found, while there are 12 positions with 10 or more. Some regions of the alignment have several adjacent positions which are relatively highly mutable, and other regions have adjacent positions with few changes indicating that the structures in sections of the chain can have similar levels of constraint. However, constraints often operate on single positions; for example, the invariant position 71 between positions having 14 and 15 mutations, invariant position 43 between positions having 7 and 10 mutations, and invariant position 10 between positions having 7 and 8 mutations. A full explanation of the molecular nature of the constraints awaits the elucidation of all of the interactions in which cytochrome *c* is involved.

## *II. The Evolution of Kingdoms*

After finding the most probable locations of branches within each of the kingdoms of multicellular organisms, these three kingdoms were combined

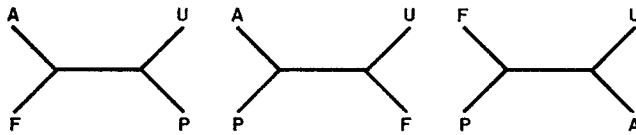


Fig. 3. Three possible configurations for a tree of four kingdoms. No orientation in time is shown. A, Animalia; F, Fungi; U, Unicellular or Protista; P, Plantae

in one configuration with the single sequence from the protistan kingdom (from *Crithidia oncopelti*). For four main branches, representing the four eukaryote kingdoms, there are three possible configurations; that is, there are three ways in which the four eukaryote kingdoms could have diverged from each other (Fig. 3). Trials of these configurations with the ancestral sequence technique showed that the first configuration listed has the minimum total number of mutations and is thus the most probable course of events. The second configuration is 0.9 % larger and the third is 1.2 % larger.

The next step was the addition of the sequence of cytochrome *c*<sub>2</sub> from *Rhodospirillum*, which represents the moneran kingdom. For five main branches representing the five kingdoms, there are 15 possible configurations, which are shown in Fig. 4. Of the 15 configurations, only one is the actual course of evolutionary history. The problem is to differentiate this one actual history from the 14 non-real configurations.

Three series of trials were conducted to answer this question and in each series the total number of mutations was determined for each of the 15 configurations. The first series used 27 sequences, which included actual sequences from all of the orders for which data had been published by mid-1972, with the substitution of two sequences of probable ancestors to represent the classes Mammalia and Aves. The second series differed from the first in the addition of recently determined sequences from a mollusc and two fungi and corrections of previous fungal sequences. The tetrapod sequences were replaced by a single sequence representing a common tetrapod ancestor. In the third series, five sequences were used. Two sequences were those from the moneran and protistan species. The other three were a basic ancestral sequence for each of the three multicellular kingdoms (Fig. 1).

In each of these three series, the configuration with the minimum total number of mutations was the same one, the first configuration in Fig. 4. In all of the series, all of the other configurations had higher mutation totals. The next higher configurations were those labeled 2, 3, and 4 in all of the series, and in the third series the mutation totals for these configurations were 3.3 % higher than that of the minimum configuration. The configura-

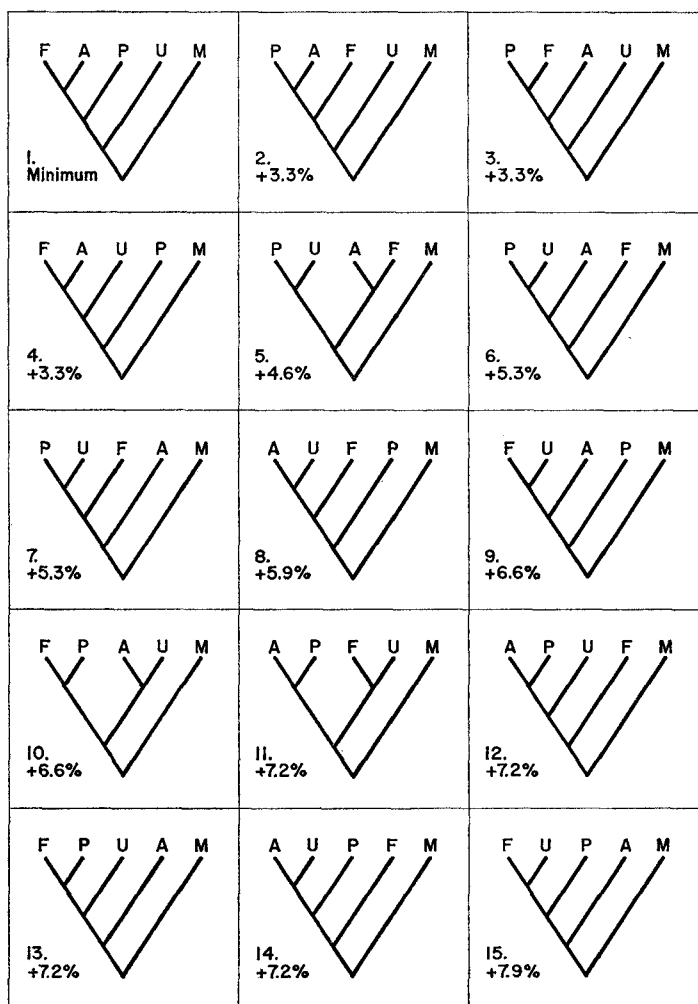


Fig. 4. All of the fifteen possible configurations for a tree of five kingdoms. Configuration 1 has the minimum number of mutations in trials using ancestral sequences to represent each kingdom of multicellular organisms. The percent increase in number of mutations above the minimum is shown for each of the other configurations. The kingdoms are F, Fungi; A, Animalia; P, Plantae; U, Protista (unicellular); M, Monera. By bending the moneran line, the configurations are drawn as if time progresses upward, thus emphasizing the difference between prokaryotes and eukaryotes which is biologically evident. The number of possible configurations and the relative locations of branches are not affected by the placement of the bend

tions of eukaryote kingdoms in Fig. 3 are comparable with configurations 1, 2, and 3 in Fig. 4 in that order, and the results of the trials with four kingdoms are basically the same as the results of the series with five kingdoms.

Throughout the history of biology, many varied proposals have been made concerning the evolution of major groups of organisms and about their classification into kingdoms and phyla. The classification which we have followed is that of five kingdoms proposed by Whittaker (1969). His classification is based on the levels of organization (prokaryotic vs. eukaryotic; primarily unicellular vs. multicellular) and on the main modes of nutrition (photosynthetic, absorptive and ingestive). He separated the Fungi and Plantae kingdoms on the basis that "The nutritive mode and way of life of the fungi differ from those of the plants." Furthermore, he states that the two groups have separate derivations from the protists. The cytochrome sequence evidence also indicates separate derivations of the two kingdoms (Fig. 2).

Traditional opinions have placed the flagellates in the position of a basic ancestor of the eukaryotes (Hyman, 1940). The occurrence of flagella with the basic  $(9 + 2)$  pattern of fibrils in almost all major groups of eukaryotes is strong evidence for this placement (Allsopp, 1969). Such views also propose that the flagellate ancestor was photosynthetic, an "Uralga" which developed from photosynthetic bacteria (Klein and Cronquist, 1967). From this stage the fungi and protozoan flagellates were supposed to have evolved by loss of the photosynthetic apparatus and, in the case of protozoa, by "... the development of the preexisting but rudimentary pinocytosis and gullet mechanisms into efficient foodcapturing systems" (Klein and Cronquist, 1967). The order in which the kingdoms diverge from each other on the cytochrome evolutionary tree differs from these traditional schemes and is difficult to reconcile even with the main points. Assuming the orientation in time placed on the cytochrome tree, it seems unlikely that a complex photosynthetic process was developed in eukaryotes only to be lost more than once, in the protozoan flagellate line and in the fungus-animal stem (where it eventually was replaced in various branches by absorption and ingestion, both of which are found in remote simpler organisms).

However, the symbiotic theory of eukaryote evolution (Margulis, 1970) proposes that a symbiotic association of monerans produced an ancestral eukaryote that was flagellated, nonphotosynthetic and had mitochondria. Such an ancestor would fit easily on the beginning of the eukaryote stem in our tree (Fig. 5). The mitochondria are supposed to arise from aerobic bacteria acquired by an anaerobic host. Although cytochrome *c* functions in the mitochondria, several types of evidence have shown that in eukaryotes the structural gene for this protein is located not in the mitochondria but in the nucleus (Sherman and Stewart, 1971). Thus the cytochrome tree traces the evolution of the anaerobic host genome in the symbiotic formation and differentiation of eukaryotes. In the symbiotic theory, photosynthesis evolved by the association of heterotroph ancestors with blue-green algae,

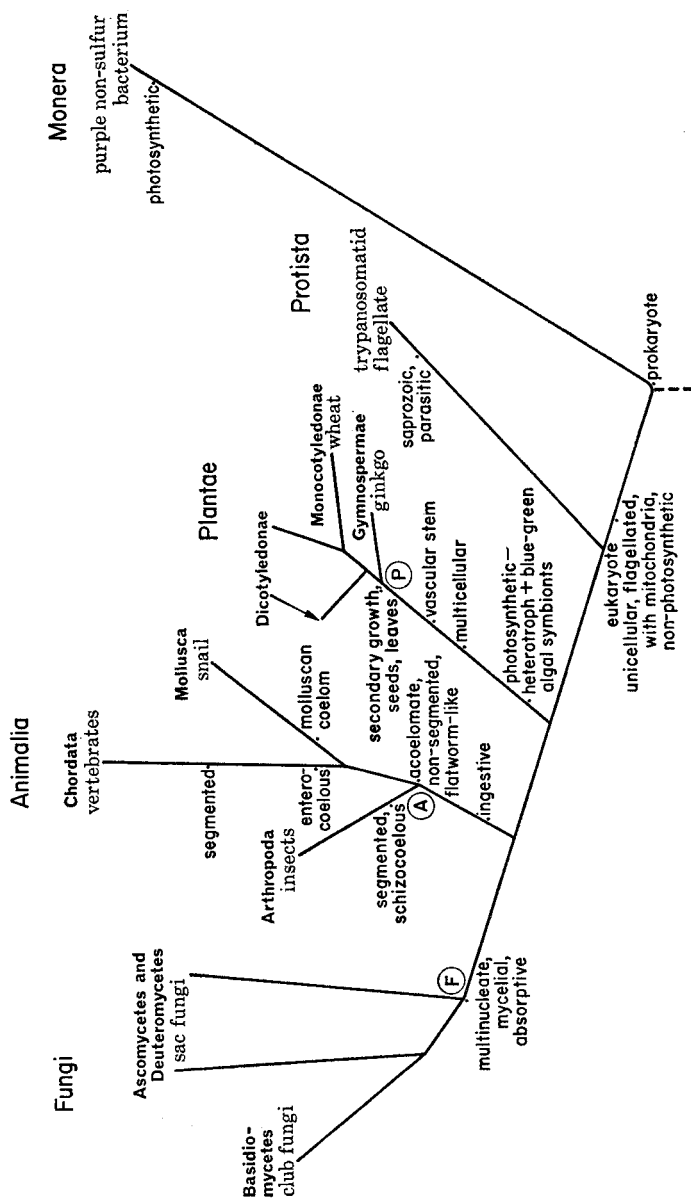


Fig. 5. Tree outlining the evolution of kingdoms and of phyla or classes which can be derived from cytochrome sequences. Probable ancestral stages are described and apply to the regions of the tree indicated by solid circles

which became plastids within the heterotrophs, and this occurrence at the beginning of the plant stem would be consistent with our location of the Plantae in the midst of ancestors and related branches which are non-photosynthetic.

"In evolution ingestive nutrition was a development secondary to the absorptive nutrition of most monerans and many eukaryotic unicells. Both protozoans with food vacuoles and metazoans with digestive tracts have probably evolved from absorptive flagellates, and in this evolution internalized the process of food absorption and added to it the process of ingestion" (Whittaker, 1969). If the ancestral forms throughout most of the differentiation of the eukaryote kingdoms had an absorptive mode of nutrition, then the location of the Fungi in our tree would be simply a continuance of this basic mode, while the Plantae branch specialized in photosynthetic nutrition and the Animalia branch specialized in ingestive nutrition. Both Margulis (1970) and Whittaker (1969) indicate a derivation of protozoan flagellates close to that of the Animalia. However, the only protozoan on our tree is widely separated from the Animalia branch and it is quite possible that protozoa with ingestive nutrition developed this means independently of the metazoa. The line between absorptive and mainly ingestive nutrition seems easy to cross. The cytostome of protozoan flagellates cannot be considered truly homologous with the multicellular archenteron of the metazoa, whatever the theoretical derivation of the archenteron. The single protistan representative on our tree has absorptive nutrition, but members of this family live by parasitism and other families in the order Protomonadida have ingestive nutrition. It may be that the protozoan flagellates as a group developed ingestive feeding and subsequently parasitic degeneration brought a return to absorption as the main mode of nutrition in the trypanosomatids. Along with other aspects, the electron transport systems of trypanosomatids (Hill and Anderson, 1970) are atypical among flagellates; it is possible that adaptation to parasitic life has affected the cytochrome *c* in some way that is not shown by the present information.

Both the kingdoms Monera and Protista are quite varied assemblages of organisms. Because the present data include only one species within each of these kingdoms, it is not possible at present to determine the course of evolution of other phyla within these kingdoms. The branches here labeled Protista and Monera can represent only the subgroups for which data are known. The ciliates and suctorians are morphologically greatly different from the protozoan flagellates (Kudo, 1966; Hyman, 1940) and may well have a separate derivation from some eukaryote stock. The various algal groups probably differentiated very early in eukaryote evolution (Margulis, 1970; Whittaker, 1969; Klein and Cronquist, 1967) and we cannot tell whether their branches arose on the vascular plant stem, on the branch to the Protomonadida, or on an ancestral eukaryote stem. The single

line to the moneran side of the evolutionary tree should be expanded with future protein data into multiple branchings to the blue-green algae and the various types of bacteria.

Among those proteins for which a rate of mutation acceptance can be calculated, cytochrome *c* is one of the slowest (McLaughlin and Dayhoff, 1972). This slow evolutionary speed is evident in sequences from mammals, where the very few differences between different orders do not clearly reveal the evolutionary history. On the other hand, if sequences are compared over the great evolutionary distances between separate kingdoms, roughly half of the amino acids are the same. The evolutionary rate of this protein is well suited to discerning the differentiation of kingdoms, phyla and classes of organisms. The taxonomic levels which can be distinguished by biochemical characteristics also depend on the degree of elevation given to morphological differentiations in different groups of organisms—family rank in the Mammalia may not be equivalent to family rank in the Dicotyledonae for overall evolutionary distance.

Because cytochrome *c* evolves slowly, it is possible to estimate the amino acid characteristics of this protein in early ancestral organisms which have long been extinct. We can do this only in kingdoms from which several sequences have been determined. We show three probable ancestral sequences in Fig. 1. These sequences were estimated in a trial which included 27 sequences representing all of the kingdoms and phyla. The blanks are positions at which the computer program found nearly equal choices of more than one amino acid.

The ancestral sequences in Fig. 1 are located at the circled letters in the branching structure of the trees in Figs. 2 and 5. The common ancestor for Fungi is at the circled F, which is the earliest divergence point in the fungal kingdom that we can discern now. This stage in fungus evolution probably had mycelia and was multinucleate or multicellular. Unicellularity in yeasts may have been a secondary development rather than a primary condition (Whittaker, 1969). The ancestor called Plantae, located at the circled P, is actually a common ancestor at the divergence of the single gymnosperm branch from the stock leading to the angiosperms. We cannot estimate any earlier ancestors in the Plantae due to the lack of sequence evidence from less specialized groups of plants. The Animalia sequence, at the circled A, is the common ancestor of the insect and mollusc-chordate branches. This ancestor may well have been a non-segmented, acoelomate, multicellular animal somewhat like the present free-living flatworms.

*Acknowledgements.* We are grateful to Dr. Lois T. Hunt and M. J. Gantt for their help in preparing the manuscript. This work was supported by NASA Contract NASW 2288 from the National Aeronautics and Space Administration and Grants GM-08710 and RR-05681 from the National Institutes of Health.



## References

- Allsopp, A.: *New Phytol.* **68**, 591 (1969).
- Augusteyn, R. C., McDowall, M. A., Webb, E. C., Zerner, B.: *Biochim. biophys. Acta (Amst.)* **257**, 264 (1972).
- Barker, W. C., Dayhoff, M. O.: *Biophys. Soc. Abstracts* **13**, 205a (1973), and unpublished results.
- Bitar, K. G., Vinogradov, S. N., Nolan, C., Weiss, L. J., Margoliash, E.: *Biochem. J.* **129**, 561 (1972).
- Boulter, D., Ramshaw, J. A. M., Thompson, E. W., Richardson, M., Brown, R. H.: *Proc. roy. Soc. B.* **181**, 441 (1972).
- Brown, R. H., Richardson, M., Boulter, D., Ramshaw, J. A. M., Jefferies, R. P. S.: *Biochem. J.* **128**, 971 (1972).
- Dayhoff, M. O., Eck, R. V., Park, C. M.: A model of evolutionary change in proteins. In: *Atlas of protein sequence and structure*. M. O. Dayhoff, editor. p. 89. Washington, D. C.: National Biomedical Research Foundation 1972.
- Dayhoff, M. O., Hunt, L. T., McLaughlin, P. J., Barker, W. C.: Data section. In: *Atlas of protein sequence and structure*. M. O. Dayhoff, editor. p. D-1. Washington, D. C.: National Biomedical Research Foundation 1972.
- Dayhoff, M. O., Park, C. M.: Cytochrome *c*: building a phylogenetic tree. In: *Atlas of protein sequence and structure*. M. O. Dayhoff, editor. p. 7. Silver Spring, Maryland: National Biomedical Research Foundation 1969.
- Dayhoff, M. O., Park, C. M., McLaughlin, P. J.: Building a phylogenetic tree cytochrome *c*. In: *Atlas of protein sequence and structure*. M. O. Dayhoff, editor. p. 7. Washington, D. C.: National Biomedical Research Foundation 1972.
- Hill, G. C., Anderson, W. A.: *Exp. Parasit.* **28**, 356 (1970).
- Hill, G. C., Chan, S. K., Smith, L.: *Biochim. biophys. Acta (Amst.)* **253**, 78 (1971).
- Hutchinson, J.: *The genera of flowering plants (angiospermae)*, vol. 1. Dicotyledones. London: Oxford Univ. Press 1964.
- Hyman, L. H.: *The invertebrates: protozoa through ctenophora*, vol. 1. New York: McGraw-Hill 1940.
- Hyman, L. H.: *The invertebrates: platyhelminthes and rhynchocoela. The acoelomate bilateria*, vol. 2. New York: McGraw-Hill 1951.
- Kerkut, G. A.: *Implications of evolution*. Oxford: Pergamon 1960.
- Klein, R. M., Cronquist, A.: *Quart. Rev. Biol.* **42**, 105 (1967).
- Kudo, R. R.: *Protozoology*, 5th ed. Springfield, Illinois: Charles C. Thomas 1966.
- Lawrence, G. H.: *Taxonomy of vascular plants*. New York: Macmillan 1951.
- Lederer, F.: *Europ. J. Biochem.* **31**, 144 (1972).
- Lederer, F., Simon, A.-M., Verdier, J.: *Biochem. biophys. Res. Commun.* **47**, 55 (1972).
- Margulis, L.: *Origin of eukaryotic cells*. New Haven: Yale Univ. Press 1970.
- McLaughlin, P. J., Dayhoff, M. O.: Evolution of species and proteins: a time scale. In: *Atlas of protein sequence and structure*. M. O. Dayhoff, editor. p. 47. Washington, D. C.: National Biomedical Research Foundation 1972.
- McLaughlin, P. J., Hunt, L. T., Dayhoff, M. O.: *J. Human Evol.* **1**, 565 (1972).
- Morgan, W. T., Hensley, C. P., Jr., Riehm, J. P.: *J. biol. Chem.* **247**, 6555 (1972).
- Pettigrew, G. W.: *FEBS Letters* **22**, 64 (1972).
- Ramshaw, J. A. M., Richardson, M., Boulter, D.: *Europ. J. Biochem.* **23**, 475 (1971).
- Richardson, M., Ramshaw, J. A. M., Boulter, D.: *Biochim. biophys. Acta (Amst.)* **251**, 331 (1971).
- Scogin, R., Richardson, M., Boulter, D.: *Arch. Biochem. Biophys.* **150**, 489 (1972).

- Sherman, F., Stewart, J. W.: *Ann. Rev. Genet.* **5**, 257 (1971).
- Takhtajan, A.: *Flowering plants—origin and dispersal*. Edinburgh: Oliver and Boyd 1969.
- Thompson, E. W., Notton, B. A., Richardson, M., Boulter, D.: *Biochem. J.* **124**, 787 (1971).
- Thompson, E. W., Richardson, M., Boulter, D.: *Biochem. J.* **124**, 783 (1971a).
- Thompson, E. W., Richardson, M., Boulter, D.: *Biochem. J.* **124**, 779 (1971b).
- Vagvolgyi, J.: *Syst. Zool.* **16**, 153 (1967).
- Whittaker, R. H.: *Science* **163**, 150 (1969).

Dr. P. J. McLaughlin  
Dr. M. O. Dayhoff  
National Biomedical Research Foundation  
Georgetown University Medical Center  
3900 Reservoir Road, N.W.  
Washington, D.C. 20007  
U.S.A