Classical Plant Taxonomic Ambiguities Extend to the Molecular Level

Michael Syvanen,[1] Hyman Hartman,[2] and Peter F. Stevens[3]

[1] Department of Medical Microbiology and Immunology, The Medical School, University of California at Davis,
Davis, California 95616, USA
[2] Computer Science Department, University of California at Berkeley, Berkeley, California 94720, USA
[3] The Herbaria, Harvard University, Cambridge, Massachusetts 02138, USA

Summary. The molecular evolution of cytochrome c from angiosperms is compared to that from vertebrates. On the basis of a cladistic analysis from 26 plant species, compared to that from 27 vertebrate species, we find that although the vertebrate sequences yield reasonably well-defined minimal trees that are congruent with the biological tree, the plant sequences yield multiple minimal trees that are not only highly incongruent with each other, but none of which is congruent with any reasonably biological tree. That is, the plant sequence set is much more homoplastic than that of the animal. However, as judged by the relative rate test, the extent of divergence, and degree of functional constraint, cytochrome c evolution in plants does not appear to differ from that of vertebrates.

Key words: Cytochrome c – Homoplasy – Plant systematics – Parsimony trees

Introduction

Botanists have long had difficulty in classifying flowering plants into unambiguous higher taxonomic groups (Stevens 1984, 1986; Dahlgren and Bremer 1986). The problem in plant taxonomy is the widespread existence of convergences, or more generally homoplasies, of gross morphological characters (Vavilov 1922; Went 1971: Dahlgren and Bremer 1986). Whether plants are qualitatively different from animals in this regard is unclear because these two kingdoms do not share homologous organs; thus, the criteria that define higher taxonomic groups within the two kingdoms defy comparison. We have

Offprint requests to: M. Syvanen

therefore examined this question by determining the molecular phylogenies of the strictly comparable cytochromes c from plants and vertebrates. There have been earlier indications that molecular systematics in plants is confronting difficulties (Boulter et al. 1979: Martin et al. 1985) that apparently result from highly homoplastic data sets (Peacock and Boulter 1975: Peacock 1981), though whether or not plants and animals differ quantitatively in this regard has not been explicitly addressed.

Methods

Sequences of the plant cytochrome c were determined by Boulter (1974); we obtained these sequences from computer files maintained by the National Biomedical Research Foundation (N.B.R.F.). Sequences for the vertebrate cytochrome c were also obtained from the N.B.R.F. except for those of the tuna, loris, rattlesnake. lizard, and lamprey, that were obtained from files maintained by Morris Goodman. Minimal replacement length trees were determined using a computer program entitled PAUP provided by David Swofford of the Illinois Natural History Survey. This program infers phylogenies using maximum parsimony where the overall number of character changes in the data set are minimized. The raw cytochrome c sequences were edited in the following ways: ambiguous assignments (such as B, Z, and X in the one letter amino acid code) were replaced with "?." Plant cytochromes c vary in length at their C termini these length differences were coded and given the same value as a single amino acid. Alignment of the sequences is the same as given by Dayhoff (1978).

Results

Greater Homoplasy in Plant Phylogenetic Trees

We first compare the amount of homoplasy found when either plant or animal cytochrome c sequence
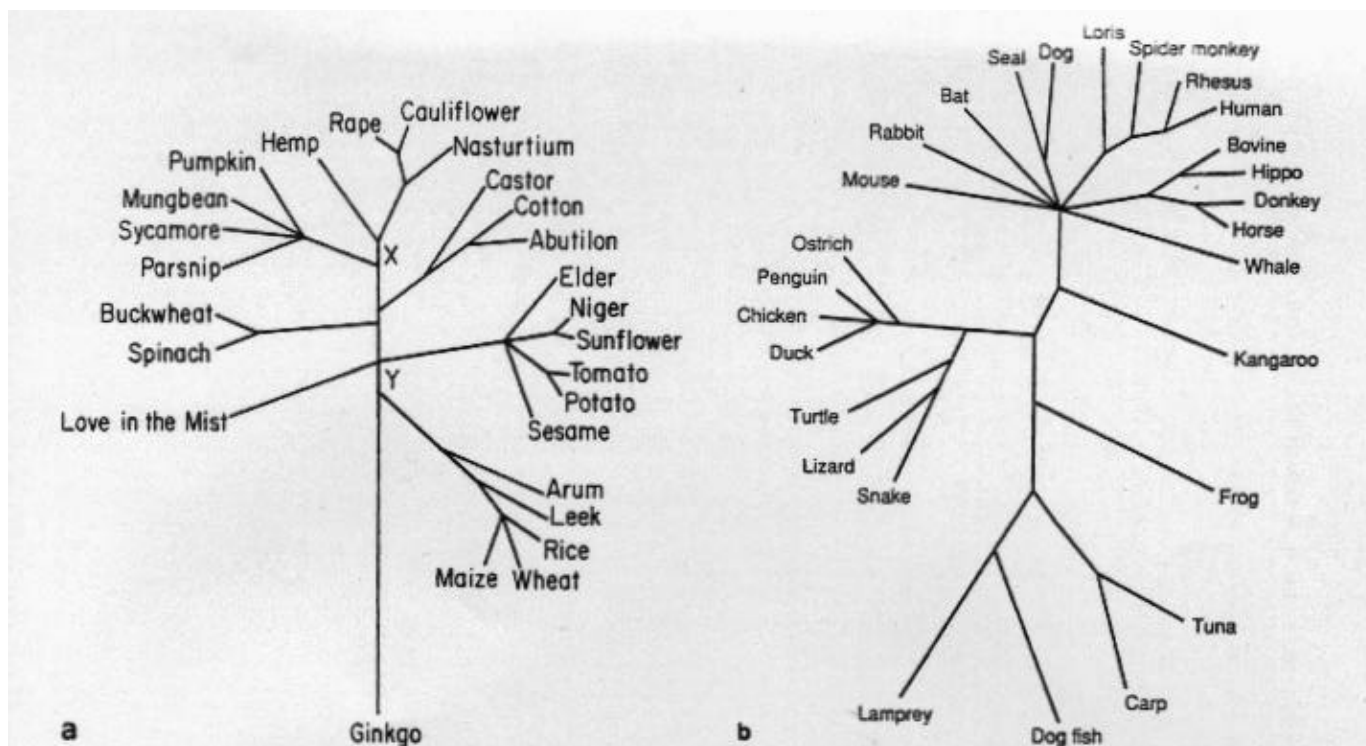
Fig. 1. Unrooted phylogenetic trees of (a) plants and (b) animals. These trees were constructed starting from an initial best-estimate of relatedness based upon all available characters. a The relative position of the internal nodes between points x and y have the greatest uncertainty. b The relative arrangement of the various vertebrate classes is not disputed, although we have arbitrarily assumed that the various mammalian orders radiated from a point.

differences are used to determine the lengths of their respective phylogenetic trees. Table 1 lists the different plant and vertebrate species considered in this study and Fig. 1 presents unrooted trees for those species based upon traditional taxonomic characters. (Note: this is *not* a minimal cytochrome c tree; those are given below.) For the plants (Fig. 1a), the tree represents little more than a guess. and there is a fair amount of uncertainty in the linkage arrangements. There are a number of uncontroversial points: (1) angiosperms (taxa 1-25) are monophyletic, (2) monocotyledons (taxa 21-25) are monophyletic, and (3) the arrangement of the dicotyledons (taxa 1-20) into their respective families is not questioned. Therefore, the major uncertainty in Fig. la is in the relationships of the dicotyledon families with one another; this area of ambiguity lies especially between points x and y along the main trunk. The vertebrate tree (Fig. 1b) is not problematic, except where we show all the mammalian orders radiating from a point; because the cytochrome c sequences (see below) do not resolve these nodes in any case, we need not concern ourselves with them. The length of the trees in Fig. 1 was calculated using the Swofford program (see Methods); ancestral sequences were determined at each node and related to the extant sequences at the end of each branch by assuming the minimal number of amino acid substitutions (see Table 2). The length of the tree,

therefore, is given in units of amino acid substitutions. An ancestral and descendant sequence may be equivalent, which means that the calculated distance between some of the nodes may be zero or otherwise different from the relative lengths of the various branches shown in Fig. 1. In addition, with this kind of unrooted tree analysis, the single ancestral sequence for all of the taxa remains unidentified.

The consistency index provides an indication of the amount of homoplasy. This index is defined as the ratio of the minimum number of amino acid changes needed to account for the differences seen in the protein sequences when these changes are not imposed upon any tree topology, divided by the length of the given tree. There is a difference in the consistency index between plants and animals. For example, with plants, the 84 character differences from the entire data set require 161 amino acid changes when imposed upon the topology of the tree in Fig. la (consistency index = 0.50), whereas with animals, the 85 amino acid differences require only 125 corresponding changes (consistency index = 0.68). The nature of the homoplastic changes seen in these two cases is displayed by a histogram in Fig. 2. The difference between plants and animals is striking. The same change occurring three or more times in the plant tree is not uncommon, but only one such change is seen in the animal tree. There

Table 1. Distance matrix for (a) plant and (b) vertebrate cytochromes c

TAXA		1	2	3	4	5	6	7	8	9	10	11	12
a) Plant cytochromes c													
1	RAPE	0.000											
2	NASTURT	0.080	0.000										
3	PUMPKIN	0.027	0.097	0.000									
4	PUMPKIN	0.0 18	0.080	0.044	0.000								
5	MUNGBEA	0.044	0.097	0.044	0.044	0.000							
6	HEMP	0.072	0.108	0.081	0.072	0.072	0.000						
7	SESAME	0.081	0.117	0.063	0.099	0.072	0.108	0.000					
8	CASTOR	0.072	0.108	0.081	0.090	0.090	0.082	0.045	0.000				
9	COTTON	0.072	0.090	0.081	0.072	0.072	0.073	0.055	0.036	0.000			
10	ABUTILO	0.090	0.108	0.081	0.090	0.090	0.082	0.055	0.045	0.027	0.000		
11	TOMATO	0.080	0.124	0.088	0.097	0.097	0.108	0.072	0.054	0.072	0.063	0.000	
12	POTATO	0.07 1	0.115	0.080	0.088	0.080	0.099	0.072	0.063	0.063	0.081	0.044	0.000
13	ELDER	0.062	0.088	0.071	0.062	0.062	0.081	0.099	0.081	0.063	0.072	0.080	0.080
14	BOXELDE	0.097	0.115	0.106	0.097	0.097	0.117	0.099	0.081	0.063	0.072	0.080	0.080
15	NIGELLA	0.128	0.147	0.147	0.128	0.128	0.110	0.156	0.148	0.148	0.139	0.147	0.174
16	NIGER	0.099	0.081	0.117	0.099	0.117	0.092	0.128	0.101	0.092	0.092	0.117	0.135
17	SUNFLOW	0.124	0.097	0.124	0.124	0.106	0.126	0.099	0.099	0.099	0.099	0.106	0.124
18	PARSNIP	0.108	0.108	0.126	0.126	0.135	0.128	0.128	0.101	0.119	0.119	0.126	0.135
19	BUCKWHE	0.107	0.134	0.125	0.107	0.107	0.127	0.100	0.108	0.108	0.108	0.107	0.116
20	SPINACH	0.142	0.159	0.150	0.142	0.142	0.171	0.162	0.162	0.144	0.162	0.177	0.168
21	RICE	0.088	0.133	0.071	0.106	0.080	0.126	0.063	0.099	0.099	0.099	0.088	0.080
22	WHEAT	0.088	0.133	0.115	0.106	0.133	0.144	0.135	0.117	0.117	0.135	0.124	0.115
23	LEEK	0.107	0.107	0.116	0.107	0.107	0.099	0.117	0.090	0.090	0.090	0.116	0.125
24	ARUM	0.072	0.090	0.072	0.072	0.063	0.083	0.073	0.073	0.073	0.073	0.063	0.081
25	MAIZE	0.054	0.099	0.072	0.072	0.090	0.101	0.073	0.064	0.083	0.083	0.054	0.072
26	GINKGO	0.155	0.182	0.173	0.173	0.173	0.167	0.167	0.139	0.157	0.148	0.136	0.155

b) Vertebrate cytochrome c

		1	2	3	4	5	6	7	8	9	10	11	12
1	DOG	0.000											
2	SEAL	0.010	0.000										
3	BAT	0.029	0.038	0.000									
4	HUMAN	0.106	0.115	0.106	0.000								
5	RHESUS	0.096	0.106	0.096	0.0 10	C 000							
6	SPMONKE	0.135	0.144	0.135	0.058	0.067	0.000						
7	LORIS	0.07	0.087	0.077	0.067	0.058	0.077	0.000					
8	HORSE	0.067	0.077	0.077	0.125	0.115	0.135	0.096	0.000				
9	DONKEY	0.058	0.067	0.067	0.115	0.106	0.144	0.087	0.010	0.000			
10	BOVINE	0.058	0.067	0.077	0.125	0.115	0.154	0.096	0.029	0.019	0.000		
11	HIPPO	0.038	0.038	0.058	0.106	0.096	0.135	0.077	0.067	0.058	0.058	0.000	
12	RABBIT	0.048	0.058	0.048	0.087	0.077	0.106	0.048	0.067	0.058	0.067	0.048	0.000
13	MOUSE	0.038	0.048	0.048	0.087	0.077	0.106	0.048	0.067	0.058	0.058	0.038	0.019
14	WHALE	0.029	0.038	0.048	0.096	0.087	0.115	0.058	0.058	0.048	0.048	0.029	0.019
15	KANGARO	0.067	0.077	0.077	0.096	0.106	0.106	0.087	0.077	0.087	0.087	0.067	0.058
16	CHICKEN	0.096	0.096	0.096	0.125	0.115	0.144	0.106	0.115	0.106	0.115	0.096	0.077
17	PENGUIN	0.096	0.096	0.096	0.125	0.115	0.144	0.106	0.125	0.115	0.125	0.096	0.077
18	OSTRICH	0.096	0.096	0.096	0.125	0.115	0.144	0.106	0.125	0.115	0.125	0.096	0.077
19	DUCK	0.087	0.087	0.077	0.115	0.106	0.135	0.096	0.115	0.106	0.115	0.077	0.067
20	SNAKE	0.23 1	0.231	0.23 1	0.173	0.183	0.221	0.212	0.250	0.240	0.240	0.231	0.202
21	LIZARD	0.125	0.125	0.125	0.135	0.125	0.163	0.135	0.135	0.125	0.144	0.125	0.106
22	TURTLE	0.087	0.087	0.087	0.144	0.135	0.154	0.115	0.115	0.106	0.115	0.087	0.087
23	FROG	0.115	0.115	0.125	0.173	0.163	0.192	0.135	0.144	0.135	0.135	0.115	0.106
24	CARP	0.117	0.117	0.136	0.184	0.184	0.214	0.165	0.146	0.136	0.136	0.117	0.136
25	TUNA	0.173	0.173	0.183	0.212	0.212	0.240	0.202	0.192	0.183	0.173	0.173	0.173
26	DOGFISH	0.163	0.173	0.173	0.231	0.221	0.250	0.192	0.173	0.163	0.163	0.173	0.163
27	LAMPREY	0.135	0.144	0.154	0.192	0.192	0.221	0.192	0.163	0.154	0.144	0.154	0.163

Unique amino acid sequences for 26 plant and 27 vertebrate cytochromes c have been determined these are maintained by N.B.R.F. The snake, loris, and lizard sequences are from the data base maintained by Morris Goodman, Wayne State University. The plant

13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
0.000														
0.053	0.000													
0.147	0.165	0.000												
0.099	0.126	0.111	0.000											
0.097	0.106	0.147	0.090	0.000										
0.126	0.126	0.150	0.109	0.117	0.000									
0.116	0.125	0.157	0.118	0.125	0.136	0.000								
0.150	0.159	0.202	0.180	0.159	0.126	0.098	0.000							
0.097	0.115	0.193	0.144	0.115	0.162	0.125	0.168	0.000						
0.115	0.115	0.174	0.153	0.133	0.126	0.143	0.133	0.106	0.000					
0.107	0.134	0.156	0.127	0.125	0.145	0.117	0.143	0.134	0.152	0.000				
0.054	0.081	0.140	0.083	0.063	0.119	0.091	0.153	0.072	0.126	0.091	0.000			
0.081	0.099	0.140	0.092	0.090	0.110	0.082	0.153	0.063	0.090	0.091	0.045	0.000		
0.155	0.145	0.178	0.145	0.182	0.164	0.119	0.191	0.127	0.145	0.174	0.157	0.130	0.000	0.000

13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
0.000														
0.019	0.000													
0.048	0.058	0.000												
0.077	0.087	0.115	0.000											
0.077	0.087	0.096	0.019	0.000										
0.077	0.087	0.106	0.019	0.019	0.000									
0.067	0.077	0.106	0.038	0.038	0.038	0.000								
0.212	0.212	0.231	0.221	0.231	0.231	0.212	0.000							
0.106	0.115	0.144	0.087	0.096	0.096	0.077	0.202	0.000						
0.087	0.077	0.106	0.077	0.077	0.067	0.077	0.240	0.115	0.000					
0.087	0.106	0.125	0.106	0.115	0.106	0.115	0.260	0.144	0.096	0.000				
0.126	0.117	0.136	0.155	0.155	0.146	0.155	0.282	0.194	0.136	0.136	0.000			
0.163	0.163	0.183	0.173	0.183	0.173	0.183	0.269	0.212	0.173	0.154	0.107	0.000		
0.163	0.154	0.192	0.183	0.192	0.183	0.183	0.269	0.202	0.183	0.192	0.175	0.192	0.000	
0.154	0.144	0.163	0.173	0.183	0.173	0.183	0.298	0.192	0.183	0.202	0.146	0.183	0.163	0.000

cytochrome c is 112 amino acids long with 34 variable positions and the vertebrate sequence is 102 amino acids long with 34 variable positions. The distances given in the table are the fraction of sites that differ over the entire gene in any pair-wise comparison
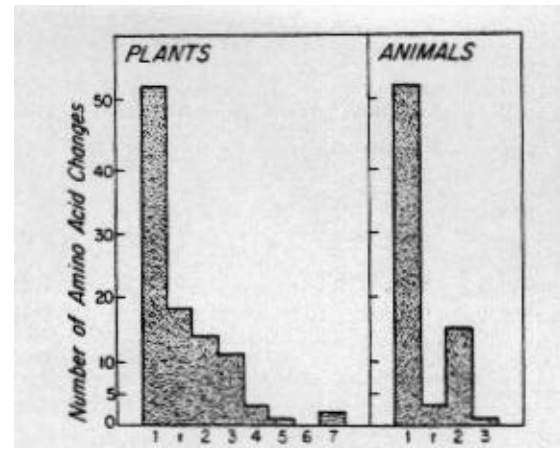
Table 2. Lengths of the phylogenetic and minimal trees based on cytochrome c

	Length	Consistency index
Plants		
Phylogenetic tree	161	0.50
Minimal tree Fig. 3a	138	0.59
Minimal tree Fig. 3b	137	0.59
Minimal tree Fig. 3c	139	0.58
Animals		
Phylogenetic tree	129	0.66
Minimal tree Fig. 4a	121	0.71
Minimal tree Fig. 4b	120	0.70
Minimal tree Fig. 4c	120	0.71

Types of Amino Acid Changes

Fig. 2. Homoplastic changes seen in the plant and animal phy. logenetic trees. Swofford's program PAUP using the different cytochromes c calculated the sequence length of the phylogenetic trees shown in Fig. 1. This program determines the lengths between nodes. so the profiles in Fig. la and b provide the topology, whereas branch lengths are determined from the protein sequences. "Type of amino acid changes" refers to the number of times that the same amino acid change appears in the tree. Therefore, homoplastic changes are given by "r' and $n>2$. No distinction is made between parallel and convergent changes. whereas "r'' indicates reversals. On two occasions the same reversal occurs twice: these are each counted twice. Designation of "ancestral" is arbitrary in identifying reversals.

are also many more reversals in the plant tree than are seen in the animal tree. (However, because the tree is unrooted and character change polarity is undetermined, "true" ancestral sequences are unknown; this makes identifying reversals problematical.)

Multiple (and Highly Incongruent) Minimal Plant Cytochrome c Trees

Presence of the homoplasies in the plant tree may be due quite possibly to serious error in the tree shown in Fig. la, and perhaps a better (i.e., more parsimonious) tree for the cytochrome c sequences could be found. To do this we used another feature of the program PAUP that is designed to find shorter trees—one branch is swapped through the tree and the length of each resulting tree is measured; this process is continued until a shorter tree can no longer be found. In the first analysis, the tree in Fig. la was used as a starting point after which the significantly shorter tree shown in Fig. 3a was obtained. This is not necessarily the shortest possible tree, only the shortest one the program could find beginning with that particular topology. Therefore, we repeated the analysis, but provided other starting topologies; in Fig. 3b, the starting topology is taken from Boulter (1974), whereas in Fig. 3c the starting topology is obtained by arbitrarily scrambling the taxa in a bifurcated tree. As can be seen in Fig. 3, each of these yields different trees, trees that are extensively incongruent with each other, but, nevertheless, are all about the same length (137—139 in Table 2). In addition, none of the minimal trees clearly resembles the tree from Fig. la. There are major discrepancies with all of the minimal trees that cause us to question their accuracy; for example, in none of the minimal trees do the five monocotyledons form a single cluster: if they are on a single

branch, some dicotyledons are included or they may be placed on different branches. Most of the families do cluster, though some of the trees break up the Asteraceae. Abutilon and cotton (Fig. 3b and c) cluster in a way that suggests they form a paraphyletic group, although relationships between the Euphorbiaceae (castor) and Malvaceae (abutilon and cotton) have been suggested. Besides the minimal trees shown in Fig. 3, we determined approximately four other minimal trees that were found beginning with either random initial topologies or with no initial topology; again, minimal trees highly incongruent with each other and with those in Fig. 3 were seen.

Is this difficulty in using the amino acid differences provided by the plant cytochrome c sequences a problem with cytochrome c itself or is it unique to plants? To test this, we performed a similar minimal tree analysis using the animal cytochrome c sequences. Figure 4a gives the minimal tree calculated from the initial topology presented in Fig. lb. The tree in Fig. 4a reproduces information that has been published on cytochrome c evolution in animals (Langley and Fitch 1974; Dayhoff 1978). These sequences reconstruct realistic phylogenetic relationships, even though we can see that the rate of primate cytochrome c evolution is faster than other mammalian orders and that snake evolution is also unusually fast. These aberrant rates do not interfere with their correct assignment, however. In order to
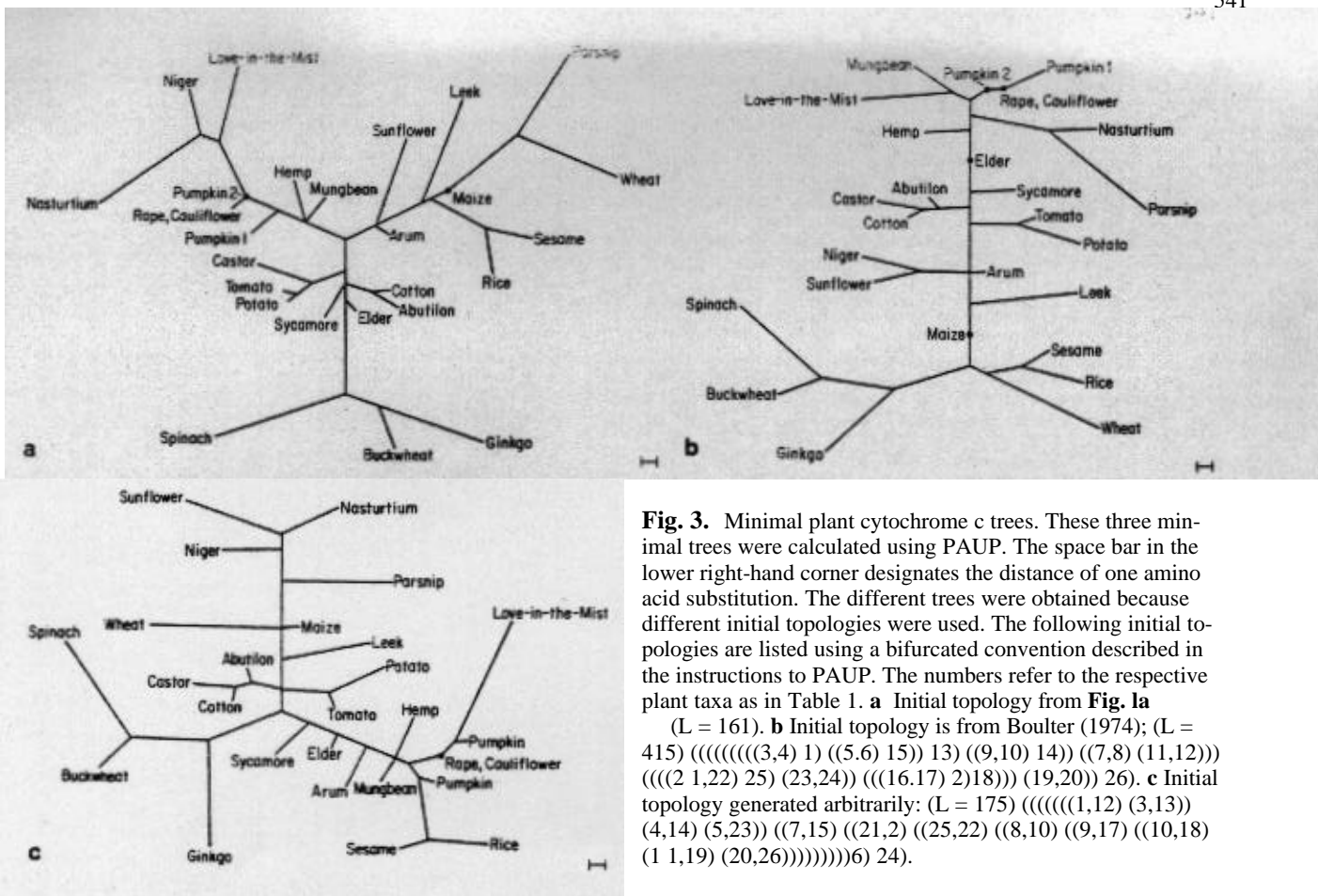
Fig. 3. Minimal plant cytochrome c trees. These three minimal trees were calculated using PAUP. The space bar in the lower right-hand corner designates the distance of one amino acid substitution. The different trees were obtained because different initial topologies were used. The following initial topologies are listed using a bifurcated convention described in the instructions to PAUP. The numbers refer to the respective plant taxa as in Table 1. **a** Initial topology from **Fig. la**
(L = 161). **b** Initial topology is from Boulter (1974); (L = 415) (((((((((3,4) 1) ((5.6) 15)) 13) ((9,10) 14)) ((7,8) (11,12))) ((((2 1,22) 25) (23,24)) (((16.17) 2)18))) (19,20)) 26). **c** Initial topology generated arbitrarily: (L = 175) (((((((1,12) (3,13)) (4,14) (5,23)) ((7,15) ((21,2) ((25,22) ((8,10) ((9,17) ((10,18) (1 1,19) (20,26))))))))6) 24).

see if other equally parsimonious but highly incongruent vertebrate trees could be found, we repeated the minimal tree search using either six different and arbitrary initial topologies or no initial topology; two of the resulting trees are shown in Fig. 4b and c. A few inconsistencies are observed in these minimal trees. The most striking is caused by unexpected similarity between the snake and the primates (Fig. 4c), and a separation of the branch containing birds, reptiles, and amphibians from that of the fish (Fig. 4b and c). However, even with these inconsistencies, broad relationships are preserved.

The situation with the animal trees contrasts sharply with that of plants. This is apparent in a number of ways. The difference in the length between the initial animal tree (129) and the minimal animal trees (120 or 121) is quite small. That is, a biologically reasonable phylogenetic tree is nearly as short as is the minimal replacement tree. This is in contrast to the plants, where the minimal trees are much shorter than the initial phylogenetic tree. In addition, as can be seen in Table 2, the plant <u>minimal</u> trees have more homoplasy (consistency index = 0.59) than do the animal minimal trees (consistency index = 0.70-0.71). Finally, the animal trees show greater congruency with one another than do the plant trees. An estimation of this can be made using a congruency measure proposed by Waterman

and Smith (1978). In this procedure the number of steps required to shift the topology of one tree into a second is computed (Table 3). Using this approach we find that the four vertebrate trees (from Figs. 1b, 4a, b, and c) are. on average, 7.3 steps removed from each other (in the six pair-wise comparisons the values range from 4 to 12), whereas the four plant trees (from Figs. la, 3a, b, and c) are, on average, 31.3 steps removed from each other (where in the six pair-wise comparisons the values range from 27 to 37).

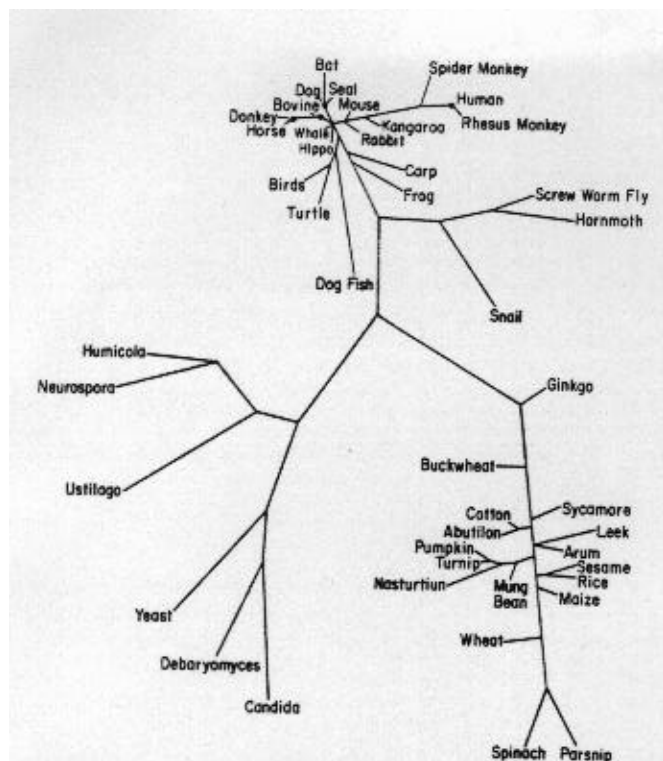*Plant and Vertebrate Cytochromes c Show Comparable Evolution*

We could imagine that the types of mutations that are tolerable to plants differ from those that are tolerable to animals and this causes the nature of cytochrome c evolution to be different in these two cases. Or possibly, evolution of plant cytochrome c is so extensive that the phylogenetic tracks have been lost. If so, we might expect to see less functional constraint or more variation among the plant sequences. However, from inspection of the sequences, both plants and vertebrates have 34 variable sites (noted in legend to Table 1) and molecular distance between sequences is quite comparable (see distance matrix Table 1). This latter point is shown

Fig. 4. Minimal vertebrate cytochrome c trees. The space bar in the lower right-hand corner designates one amino acid substitution. **a** Initial topology from Fig. 1b. **b** No initial topology assigned. In this case PAUP computes its own (see documentation to PAUP by David Swofford). **c** Initial topology assigned with random number table; (((((((((7,4) (12,20)) ((8,13) (18,2))) (((23,10) (16,25)) ((15,21) (24,11)))) (17,27)) (3,19) (9,5)) (26,22)) (14,6))1).

by the average molecular distance between plant sequences as compared to vertebrate sequences (Table 4a). Furthermore, not only is the magnitude of sequence differences nearly the same for both plants and vertebrates, but we can show that their rates of evolution are roughly comparable. This is illustrated by the relative rate test (Sarich and Wilson 1967) that measures both the extent and the variation in rate of sequence evolution (Table 4b). In this test the molecular distance between an outgroup taxon (for example, the ginkgo) and each member of a monophyletic cluster (for example, the 25 angiosperms) is determined. Note that the molecular distance is the number of differences between two species under comparison, not actual lengths through

a tree. If the rates of evolution for the different members of the monophyletic cluster are approximately the same, then the distance between each of these members and the outgroup taxa will be the same. and therefore, the variation in rate for the different lineages can be revealed by the standard deviation of those distances. As can be seen (Table 4b), both the mean and the variation in distances for the plants are basically the same as that seen with vertebrates. Thus, both the extent of divergence and variations in rate are about the same in plants as in animals.

Strictly speaking, we cannot say that the rates are exactly the same for plants and animals. The lineage giving rise to conifers and ginkgo probably separated from that giving the angiosperms in the Upper Car-

Fig. 5. Minimal eukaryotic cytochrome c tree. The plant and vertebrate sequences shown in Table I as well as the indicated invertebrate and fungal sequences were used in the tree construction. Only those regions of the sequence homologous throughout the entire set were used to construct the tree.

Table 3. Tree topology comparisons

Plants				Animals			
Trees compared	la	3a	3b	Trees compared	1b	4a	4b
3a	35			4a	4		
3b	26	37		4b	7	3	
3c	34	27	29	4c	12	10	8

The topologies from the unrooted trees in Figs. 1a and 3 (plants) and from 1b and 4 (animals) are compared against one another, respectively. This congruency test (Waterman and Smith 1978) is performed by converting the topology of one of the two trees under comparison into the topology of the second. This is done by moving branches that may have one or more associated taxa from one point in the tree to a second point by following trunks. The magnitude in the table refers to the number of nodes that are crossed as a branch is moved through a tree. Trifurcated and higher nodes are weighted the same as bifurcated nodes, otherwise the distance between nodes does not influence the magnitude. In order to find the shortest path between the two topologies, we constructed intermediate trees that were not justified either on the basis of any phylogenetic or parsimony considerations. The operation of moving branches was performed manually, and a path with the shortest possible distance was sought

boniferous about 260 x 10⁶ years ago (Rothwell 1982). [Although divergence of the extant angiosperms is generally considered to have occurred about 145 X 10⁶ years ago in the Cretaceous, some

Table 4. Average molecular distances and relative rate lest for plant and vertebrate cytochromes c

	(a) Average molecular distance			(b) Relative rate test
	Mean distance	Standard deviation	n	Mean distance
Plants	11.5	3.8	25	17.4
Vertebrates	11.2	4.8	26	17.9

[a] This is the average distance between all pairs in the set Character difference distances (d_{ij}) are determined by aligning two cytochrome c sequences from different taxa and counting the number of amino acids that differ; these are derived from the fractional values shown in Table 1:

$$\text{Mean distance} = (2 \sum_{i \neq j} d_{ij}) / n (n - 1)$$

[b] For plants, distances are measured from ginkgo (defined as the outgroup taxa) to each of the other angiosperms. For animals, the distance is from the lamprey to each of the other vertebrate taxa. The mean gives the average distance for each of these sets of measurements

of the early angiosperms may be considerably older (Doyle and Donoghue 1986).] The jawless fish/other vertebrate division probably occurred in the Upper Devonian about 380 x 106 years ago; thus, plant cytochrome c evolution appears to be about 25% faster than animal cytochrome c evolution.

The Minimal Eukaryotic Tree

In addition to showing molecular clock behavior, plant cytochrome c evolution is consistent with overall eukaryotic evolution. If we construct a single minimal tree using both animal and plant cytochrome c sequences as well as those from some invertebrates and fungi. the tree shown in Fig. 5 is obtained. This is a minimal tree as determined by PAUP. A few salient points: the relationship of the various kingdoms to one another is reasonable, the relationships among the metazoan phyla are accurate, and again we see that the various vertebrate classes are conventionally related. But the relationships among the plants themselves are again highly incongruent with any conventional expectation.

Discussion

Homoplasy appears to rule plant evolution; this is seen for both morphological as well as molecular character sets. The fact that homoplasy is so extensive among the angiosperms has made it impossible to identify higher taxonomic groups; this distinguishes evolution in plants from that in animals. In a number of respects the plant cytochrome c appears

to have evolved in a well-behaved fashion, not differently from other eukaryotic cytochrome c. The plant sequences diagnose their own kingdom (Fig. 5) and their rate of change appears relatively uniform (Table 4b). We have compared the plant sequences to the vertebrate sequences in order to control for homoplasies caused by random genetic drift and covariation (Fitch 1967; Peacock and Boulter *1975;* Peacock 1981) in the protein. We may conclude that evolution in plants is qualitatively different from that in animals. Gross morphological characters may be attributed to natural selection, whereas protein sequence characters are explained by stochastic events according to the neutral theory. The fact that homoplasy in plants is seen with both sets of characters strongly suggests that the homoplasy is not caused by natural selection acting at the organismal level but rather is the result of underlying genetic mechanisms. How the molecular genetics of plants and animals differ in this regard remains to be determined.

References

Boulter D (1974) The evolution of plant proteins with special reference to higher plant cytochrome c. Curr Adv Plant Sci 4(Commentaries in Plant Sci No 8):1-16

Boulter D, Peacock D, Guise A, Gleaves JT, Estabrook G (1979)

Relationship between the partial amino acid sequences of plastocyanin from members of ten families of flowering plants. Phytochemistry 18:603—608

Dahlgren R, Bremer K (1986) Major clades of angiosperms. Cladistics 1:349—368

Dayhoff MO (1978) Atlas of protein sequence and structure—supplement 3. National Biomedical Research Foundation, Washington DC

Doyle JA, Donoghue MJ (1986) Seed plant phylogeny and the origin of angiosperms: an experimental cladistic approach. Bot Rev *52:34 - 431*

Fitch WM (1967) A method for estimating the number of invariant amino acid coding positions in a gene; use of cytochrome-c as a model case. Biochem Genet 1:65—71

Langley CH, Fitch WM (1974) An examination of the constancy of the rate of evolution. J Mol Evol 3:161—177

Martin PG, Boulter D, Penny D (1985) Angiosperm phylogeny studied using sequences of five macromolecules. Taxon 34:393—400

Peacock D (1981) Data handling for phylogenetic trees. In: Gutfreund H (ed) Biochemical evolution. Cambridge University Press, pp 88 - 115

Peacock K, Boulter D (1975) Use of amino acid sequence data in phylogeny and evaluation of methods using computer simulation. J Mol Biol 95:513—527

Rothwell (l982) New interpretations of the earliest conifers. Rev Palaeobot Palynol 37:7—28

Sarich VM, Wilson AC (1967) Rates of albumin evolution in primates. Proc Natl Acad Sci USA 58:142—148

Stevens PF (1984) Metaphors and typology in the development of botanical systematics 1690—1960 or the art of putting new wine in old bottles. Taxon 33:169—2 11

Stevens PF (1986) Evolutionary classification in botany, 1960— 1985. J Arnold Arbor 67:313—339

Vavilov NI (1922) The law of homologous series in variation. J Genet 12:47—89

Waterman M, Smith T (1978) On the similarity of dendrograms. J Theor Biol 73:789—800

Went FW (1971) Parallel evolution. Taxon 20:197—226