CHAPTER 30

Temporal Patterns of Plant and Metazoan Evolution Suggest Extensive Polyphyly

Michael Syvanen

The current work has determined the divergence times between major eukaryotic clades based on an analysis of 18S ribosomal RNA. A trifurcation rate test is employed which renders it unnecessary to assume that the molecular clocks in the different lineages under comparison are the same. This test suggests divergence times between some of the major clades that are consistently earlier than would be suggested by the fossil record. For example, the trifurcation rate test suggests a molecular divergence time for monocots and dicots at about 175–205 Myr ago, while the fossil record shows that the angiosperm radiation occurred 110 Myr ago. Similar discrepancies are seen between molecular and paleontological estimates of divergence times when the lineages being compared include angiosperms, gymnosperms, bryophytes and some of the metazoan phyla. This suggests that major clades are polyphyletic in that the same modern characters evolved in different lineages (i.e. extensive parallelism has occurred). This discrepancy between molecular time estimates and paleontological estimates is not as extensive with the animals in that most of the major phyla diverged at a time consistent with the Cambrian radiation. There are two exceptions – the Cnidaria and Porifera diverged from the lineage leading to other metazoan phyla about 400 Myr before the Cambrian radiation. A single simple explanation for these widespread parallelisms, ambiguous higher taxonomic categories and polyphyly is that asexual transfer of genetic information between remotely related plants and animals was a major factor in shaping their macroevolutionary patterns.

INTRODUCTION

When Haeckel (1866) took up Darwin's challenge and began a program to unravel metazoan phylogeny, the subsequent effort forced the question whether shared characters were due to descent from common ancestry or whether they arose independently in parallel lineages. Thus, in 1870, Lankester coined the phrase "homoplasy" and formalized the question: Were characters homologous (modern definition) or homoplastic? This problem has bedeviled biologists since. As vexing as the problem was to zoologists (they were, after all, able to find consensus by the 1920s on what constituted the major metazoan phyla), it has proven nearly insoluble in botany. This is because the morphological characters upon which phylogenetic trees are constructed create radically different branching patterns depending upon the weighting of the characters. Botanists have, at different times, come to agree on taxonomic categories, but these periods of agreement have been fleeting (Stevens, 1984). Paleontologists have the potential to record the phylogenetic changes through time and hence, identifying clades as they arise. This requires, however, identifying an ancestral form that shares the

characters of the resulting clade of descendants. Identifying such ancestral forms has proven problematic. Thus, it has been frequently suggested that major plant clades are polyphyletic, i.e. that they arose from multiple ancestors, none of whom share the clade's defining characters. The difficulty in determining whether clades are monophyletic or polyphyletic is especially problematic when attempting to reconstruct the evolutionary history of the modern flowering plants – the angiosperms – but is also seen in other higher taxonomic groups such as the gymnosperms, ferns and mosses.

The introduction of molecular evolution studies was greeted with great expectations; it was assumed that the taxonomic ambiguities of plant classification would be resolved. These hopes, however, were quickly dashed (Peacock and Boulter, 1975; Boulter and Gilroy, 1992). Phylogenetic tree constructions based on molecular sequences have confronted the same problem that confronts morphological character arrays – namely homoplasy. We documented this in an earlier study of cytochrome *c* sequences and showed, furthermore, that the problem of homoplasy of the plant sequences was significantly different from homoplasy among animal sequences (Syvanen et al., 1989). Molecular sequence information has a potential advantage over morphological character arrays due to the possibility of inferring temporal relationships through the operation of what is commonly called the molecular clock. That is, if we know the rate at which two lineages have diverged from a common ancestor and the extent of divergence, then we can calculate the time of divergence. Molecular clock estimations can therefore be tested against the fossil record. Molecular clock estimations have traditionally assumed that the rate of evolution in two lineages under comparison is constant (Sarich and Wilson, 1973).

Recently, the 18S ribosomal RNA sequences from a large number of different land plants have been determined and deposited in public gene banks; this development was greeted with great expectations, and again it appears there are some difficulties in finding a consensus interpretation. For example, the first 90 plant rRNA sequences that were deposited in GenBank are referenced in the annotations as: "Darwin's abominable mystery revisited: Ribosomal RNA insights into flowering plant evolution, Unpublished (1991) by Hamby et al." However, no paper ever appeared with only some summaries of these data appearing (Doyle et al., 1994). Published analyses of a larger dataset seems to contain more encouraging results in that both the angiosperm and gymnosperm descended from their own unique ancestor (Chaw et al., 1995, 1997; Soltis and Soltis, 1997; Soltis et al., 1999), though it is clear that topologies based on 18S rRNA are frequently inaccurate. More recently, a number of studies on a variety of proteins have appeared that have been hailed as a major breakthrough in plant classification. Unfortunately the new topologies conflict with 18S rRNA trees and have also resulted in the demise of the monocot–dicot division which used to be one of the more solid higher taxonomic groupings of the angiosperm (Donoghue and Alverson, 2000; Donoghue and Doyle, 2000; Mathews and Donoghue, 1999; Soltis et al., 1999).

So far the 18S rRNA sequences have been used to define phylogenetic topologies, a concerted effort to infer times of divergence using molecular clock assumptions has not been attempted. The major reason for this is that the rate of 18S rRNA evolution appears to be highly variable among different lineages, thereby invalidating the assumption of a constant molecular clock (Romano and Palumbi, 1996; Sogin et al., 1996; Sorhannus, 1996). In the current chapter I describe a procedure for estimating divergence times when there are unequal rates of evolution in the lineages being compared; this is straightforward application of an approach suggested by Fitch and Margoliash (1967) and has found application on numerous occasions (Chapter 28). The different approach I am taking is to average corrected distances for numerous members of individual clades and to use these averages in constructing trees. In addition, taxa that introduce inconsistencies are removed from the analysis. This test will be used to estimate the time of major radiations of the land plants (where the metazoan and fungi are used as outgroups), and the time of metazoan radiation (where plants and fungi are used as outgroups).

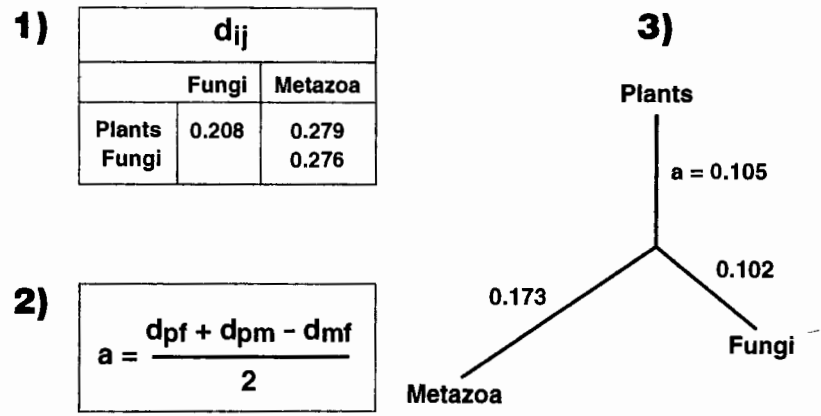

FIGURE 30.1 The trifurcation test. Estimation of evolutionary rates when outgroup taxa are of unequal rates. (1) The distance matrix gives an estimate of the number of nucleotide substitutions that have occurred since the separation of two groups under comparison. Because of the large variation in d_{ij}s among individual taxa, the numbers are averages from many different comparisons. (2) The subscripts p, f, and m refer to plant, fungi and metazoa respectively. Shown is the example for calculating the number of substitutions that occurred in the plant lineage back to the trifurcation. (3) Time to trifurcation in 1.05×10^9 years for fungi and metazoa.

COMPUTATIONS

Computer methods

All computer calculations were performed using GCG software (Genetics Computer Group, Madison, WI) in a VMS or UNIX operating environment.

Sequence alignment

In the current analysis, over 60 sequences were chosen with representatives from angiosperms, gymnosperms, bryophytes, fungi and metazoa from a much larger group. The 18S sequences were recovered from GenBank and aligned using the GCG program Pileup. The alignments were then edited manually to correct alignment errors and to remove uninformative sequences. The 5′ and 3′ ends of each were trimmed such that the 5′ sequence began with the highly conserved TTAAGCCATG and the 3′ sequence ended with the relatively conserved GGGCGGTCG 3′. The highly variable regions, as defined from plant, fungi and metazoan comparisons, were deleted from the sequence. This step had the effect of removing nearly 200 characters that were phylogenetically informative among, for example, vascular plants but not between the different kingdoms. After adjusting the alignment, 1650 nucleotide positions were compared. In a second alignment, with a larger group of metazoans, more extensive editing was performed and only 1350 nucleotide positions were compared.

The trifurcation test

Figure 30.1 illustrates application of the trifurcation test of Fitch and Marguliash (1967) to fungi, f, plants, p, and metazoans, m. If the length of each line in units of the number of nucleotide substitutions since divergence from the common ancestor is given as a, b, c in the figure, then the number of substitutions that separate each taxa, i.e. the number of nucleotide differences divided by the number compared (d_{ij}) is

$$d_{pm} = a + b$$
$$d_{pf} = a + c$$
$$d_{mf} = b + c$$

and a, b and c are easily solved since d_{ij} values are given in the distance matrix calculated from aligned sequences, and is shown in Figure 30.1. These simple relations were first suggested by Fitch and Margoliash (1967) and have been incorporated into programs designed to calculate trees and iterations of this simple calculation are

TABLE 30.1 Taxa used in the present study

GeneBank designation	Domain/phyla	Species
Plants		
AASRG18S	Dicot	*Asarum canadense*
ACURG18S	Monocot	*Acorus calamus*
ADPRG18S	Magnoliophyta	*Antidaphne viscoidea*
AKEERG	Dicot	*Akebia quinata* Houtt
ARU42494	Dicot	*Acer rubrum*
CUORGE	Dicot	*Caulophyllum thalictroides*
D29773	Monocot	*Trachycarpus wagnerianus*
D29776	Dicot	*Magnolia acuminata*
D85299	Gnetophyta	*Welwitschia mirabilis*
D85303	Filicophyta	*Asplenium nidus*
DAU38314	Dicot	*Dillenia alata*
DCU42532	Dicot	*Drosera capensis*
EGLRG18S	Magnoliophyta	*Englerina woodfordioides*
EOARG18S	Magnoliophyta	*Exocarpos bidwillii*
GACRGE	Dicot	*Glaucidium palmatum*
GAU43012	Gymnosperm	*Gnetum africanum*
GDORGE	Monocot	*Gladiolus buckerveldii*
GILRG18S	Magnoliophyta	*Ginalloa*
GSU42541	Dicot	*Geranium* sp.
GUU42417	Gymnosperm	*Gnetum urens*
ISPRGE	Monocot	*Isophysis tasmanica*
LH18SRRNA	Bryophyta	*Lophocolea heterophylla*
LUORGEA	Dicot	*Lepuropetalon spathulatum*
NSU42787	Dicot	*Nepenthes* sp.
OEU42791	Monocot	*Oncidium excavatum*
PIN18SRR	Gymnosperm	*Pinus luchuensis*
RH18SRRNA	Bryophyta	*Riella helicophylla*
RP18SRRNA	Bryophyta	*Rhacocarpus purpurascens*
SBIRGE	Dicot	*Sabia swinhoei*
SNDNA18RR	Filicophyta	*Salvinia natans*
SPIRG18S	Dicot	*Spinacia oleracea*
SRMRG18S	Monocot	*Sparganium eurycarpum*
TUERG18S	Magnoliophyta	*Tupeia antarctica*
ZMU42796	Monocot	*Zea maize*
Choanoflagella		
CFGRGDL	Choanoflagellida	Rosette agent
Fungi		
BBU59062	Basidiomycota	*Bondarzewia berkeleyi*
BRU42477	Ascomycota	*Botryosphaeria ribis*
NEORR18S	Chytridiomycota	*Neocallimastix* sp.
RDU42660	Ascomycota	*Reddellomyces donkii*
UHRRNA18S	Ascomycota	*Urnula hiemalis*
Metazoans		
apu43190	Porifera	*Axinella polypoides*
bgu65223	Mollusca	*Biomphalaria glabrata*
bgu65223	Mollusca	*bloodfluke planorb*
bl18srrna	Echinodermata	*Brissopsis lyrifera*
cc18srrna	Echinodermata	*Centrostephanus coronatus*

TABLE 30.1 Contd

GeneBank designation	Domain/phyla	Species
ccu42452	Porifera	*Clathrina cerebrum*
cpz86107	Cnidaria	*Coryne pusilla*
eb18srrna	Echinodermata	*Echinodiscus bisperforatus*
et18srrna	Echinodermata	*Eucidaris tribuloides*
ewu29492	Arthropoda	*Eusimonia wunderlichi*
funrg18s	Vertebrata	*Fundulus heteroclitus*
humrge	Vertebrata	*Homo sapiens*
ll18rr	Mollusca	*Littorina littorea*
mh18srna	Tardigrada	*Macrobiotus hufelandi*
oeu88709	Mollusca	*Ostrea edulis*
pau42453	Cnidaria	*Parazoanthus axinellae*
peu29494	Nemertea	*Prostoma eilhardi*
pm18srrnx	Echinodermata	Sand urchin
ps18srrn3	Vertebrata	*Polyodon spathula*
ratrge4a	Vertebrata	*Rattus norvegicus*
rp18srr	Vestimentifera	*Ridgeia piscesae*
scu29493	Arthropoda	*Scolopendra cingulata*
sebrg18s	Vertebrata	*Sebastolobus altivelis*
sv18rr	Mollusca	*Scutopus ventrolineatus*
th18srrna	Echinodermata	*Temnopleurus hardwickii*
xelrge14	Vertebrata	*Xenopus laevis*

used in algorithms that determine distance trees of unequal rates. If the time, T, of the trifurcation is known, for some clades, then the clock in each lineage is simply a/T, b/T and c/T, respectively. Hence, it is not necessary to assume that the rate of evolution in each lineage is constant to calibrate the clock.

Clearly, before these rates can be used to determine divergence times of taxa within lineage a, for example, we must assume that the rate of evolution in the lines leading to the taxa under comparison are the same and equal to a. This assumption can be tested using the relative rate test (Sarrich and Wilson, 1973) and deviations can be accommodated (see below).

DETERMINATION OF DIVERGENCE TIMES FOR THE LAND PLANTS

The plant, animal and fungi kingdoms diverged from a common ancestor about 1×1000 Myr ago (Yokoyama and Harry, 1993; Doolittle et al., 1996). Extensive analysis of numerous protein sequences that show clear clock-like behavior fix this time within an error of about 10%. It appears that the metazoa and fungi are about 5% more closely related to each other than either is to plants (Baldauf and Palmer, 1993; Nikoh et al., 1994; Sidow and Thomas, 1994; Drouin et al., 1995; Kumar and Rzhetsky, 1996; Sogin et al., 1996; VandePeer et al., 1997). Hence, if metazoa and fungi diverged 1.0×10^9 Myr ago, then plants diverged from the trifurcation point 1.05×10^9 Myr ago. This will be one of the calibration points I will use in calculating the divergence times. The trifurcation test will first be used to determine divergences in the major plant clades where we will use fungi and two metazoan phyla as our outgroup. In another section, we will apply this test to metazoan phyla where plants and fungi will be used as outgroups.

Selection of taxa and multisequence alignment

There are over 300 nearly complete 18S rRNA sequences of different land plants deposited in the public databases. About 30 were chosen for

the purposes of this study. They were primarily collected by Soltis et al. (1997) (Table 30.1). The three mosses and the fern were grouped together and are collectively referred to as "bryophytes," though they are more accurately called lower archaegoniates.

A group of fungi and metazoan taxa were chosen from a much larger group found in the GenBank. These were analyzed by picking representative samples which were characterized by the fact that, after alignment, the simple d_{ij} among both the fungi and metazoa were greater than 0.10. The amount of divergence among the metazoans is highly variable.

These sequences were aligned and corrected distances were calculated (Table 30.2). The average distance between all members of each assemblage was computed and Table 30.2 presents these averages. On the basis of 18S rRNA sequences alone, the seven groups shown in Table 30.2 define clearly delineated clades. These clades are seen using either parsimony or distance methods. For purposes of presentation I am showing the results where Cnidaria and Porifera are the metazoan outgroup, though the results using a larger number of protostomes and deuterostome did not affect the final calculations of plant divergence times (not shown). However, representatives of these latter taxa were present during sequence alignment and their presence influenced the final alignment.

Sources of error

There are three sources of error that contribute to any estimated divergence times using the data in Table 30.2. The first is the variance in raw d_{ij} values among, for example, different pairs of monocots and dicots. In the present analysis, this variation contributes the least error to the final estimate, since we are using averages from a relatively large number of taxa resulting in a small standard error.

A second source of variation is in the choice of the outgroup taxa. This is illustrated in Figure 30.2. In this example, averages from the same set of plant and fungi sequences are compared against individual metazoan taxa and the average distance to trifurcation for the plants is computed. As can be seen, there are significant differences depending on the choice of the metazoan. Similar variations in plant distance to trifurcation is seen if the metazoan taxa are held constant and individual fungi are used. As with the error due to variation in d_{ij}, this error is minimized by averaging over relatively large numbers of taxa so the ultimate standard error is small.

The third source of error is due to the estimated substitution rate, i.e. have we calibrated the clock properly? This error contributes the greatest uncertainty to our final estimated divergence times. Recall that we are using a 1.05 Byr time to trifurcation to calibrate a clock

TABLE 30.2 Average distances between major plant clades. Just the cnidarian and porifera taxa were used as the metazoan out group but representative of the other phyla were included in determining the final sequence alignment. The numbers outside the parentheses are d_{ij} values. The numbers inside the parentheses are the comparable patristic distances from the tree in Figure 30.3. A statistical correction for multiple substitutions was applied. A gamma distribution of multiple hits was assumed (Jin and Nei, 1990), though using a Poisson distribution (Jukes and Cantor, 1969) gave comparable results

	1	2	3	4	5	6	7
1. Monocots	–	4.32 (4.3)	9.62 (9.6)	10.09 (9.9)	21.50 (21.6)	22.92 (22.6)	19.97 (19.6)
2. Dicots		–	9.44 (9.5)	10.07 (9.8)	21.52 (21.5)	22.46 (22.5)	19.71 (19.5)
3. Gymnosperms			–	9.97 (10.4)	22.52 (22.2)	23.05 (23.2)	19.76 (20.2)
4. Bryophytes				–	21.52 (20.1)	20.20 (21.0)	17.40 (18.0)
5. Fungi					–	19.70 (19.7)	16.71 (16.7)
6. Cnidaria						–	14.80 (14.8)
7. Porifera							–

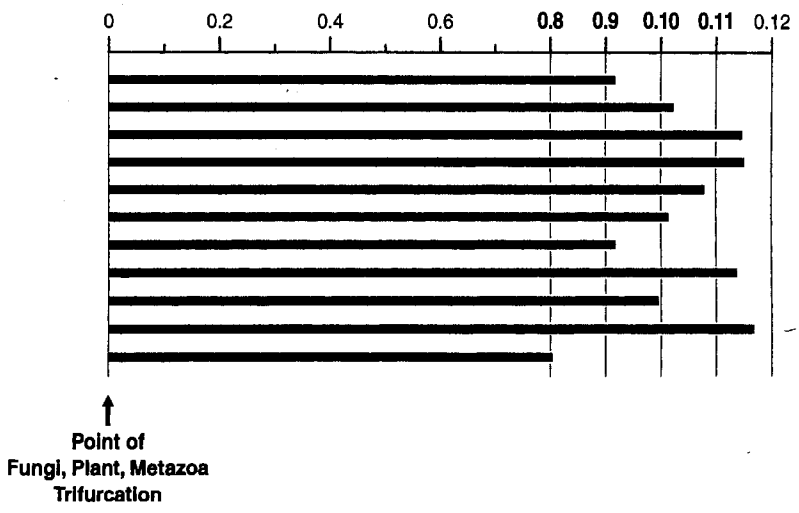

FIGURE 30.2 Variation in the rate due to choice of one outgroup in the trifurcation test.

that will be used to estimate 200–500 Myr divergence times. As mentioned above, in order for the trifurcation test to work, it is best that, for example, the different plant lineages being compared have comparable rates of evolution. To a first approximation, this is the case, as can be seen by applying the relative rate test to the four major plant assemblages; monocots, dicots, gymnosperms and bryophytes in Table 30.2. For example, the distance from the fungi to these four groups are 21.5, 21.5, 22.5 and 21.5 respectively, while the distance from the Porifera is 20.0, 19.7, 19.8 and 17.1 respectively. However, on closer examination, there are significant differences between the rates of evolution among the four plant assemblages since their divergence. This is seen most clearly in the complete distance tree shown in Figure 30.3. (This tree is constructed by a reiteration of the trifurcation test and a weighted least squares distance average for individual branches.) As can be seen, the bryophytes evolve the slowest and the gymnosperms the fastest. This then raises the question: which of the rates is representative of the land plants as a whole? This problem was dealt with by calculating divergence times by two means – in the first, the slowest evolving lineage is considered to be representative of plants as a whole, and in the second, the fastest evolving lineage is. In Figure 30.4, the branching times are given as the average of these two values and the error bar gives the range between slowest and fastest.

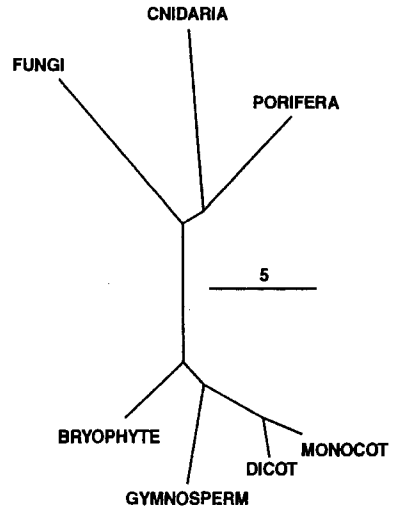

FIGURE 30.3 Distance tree based on data in Table 30.2. The space bar gives a distance of five substitutions per hundred nucleotides.

The tree in Figure 30.5 can be used to estimate the overall robustness of the approach I have employed in this analysis. The distances through the finally computed tree (i.e. the patristic distances) are shown in parentheses next to the d_{ij} values in Table 30.2. As can be seen, there is quite good agreement between these two values. If we

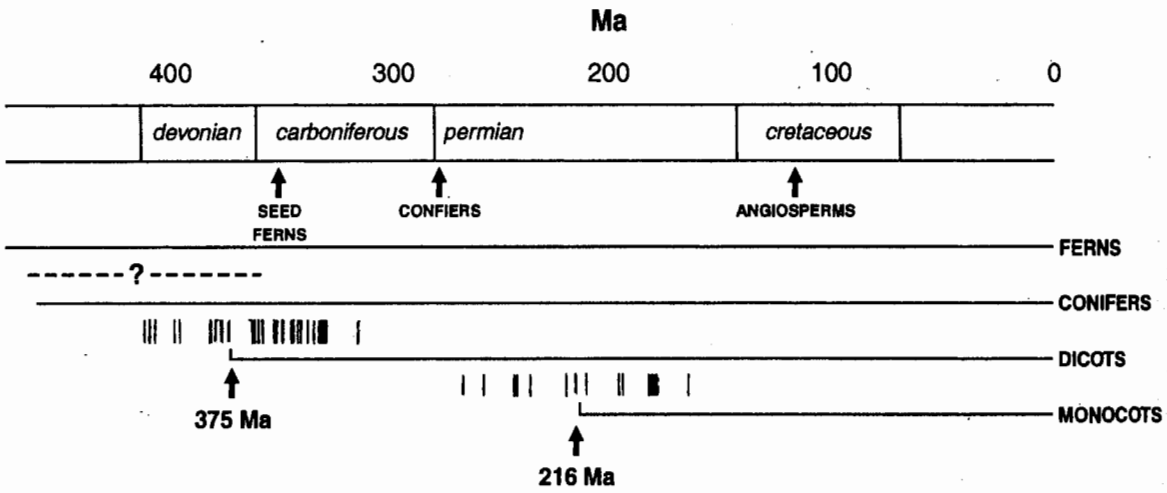

FIGURE 30.4 Times of divergence of major plant clades. The times calculated in this study are compared to the geologic periods. The first appearance of the different groups of plants are shown by the arrows. The scatter of points between the gymnosperm/angiosperm split and the monocot/dicot diversion reflect the magnitude of variation when individual taxa are used to compute patristic distances.

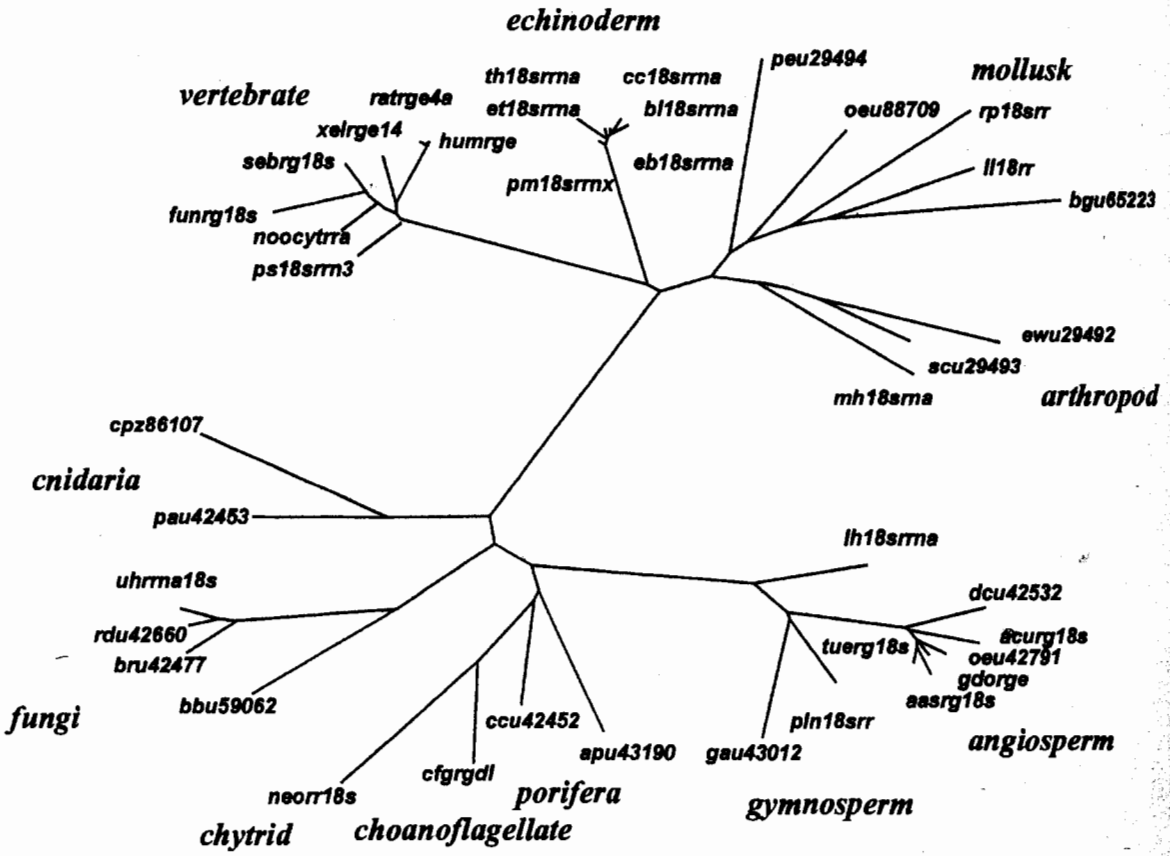

FIGURE 30.5 Untrooted tree showing larger groups of metazoans. This is a Jin/Nei nearest neighbor tree provided in the GCG package and is basically a distance tree that does not assume constant. Arrow denotes trifurcation point to the prokaryotes.

exclude lower archaegoniates, the agreement is 98% or better. In real datasets, patristic distances and d_{ij} values rarely show such close agreement; the reason for the agreement in the present case is because the values in Table 30.2 represent the average from large numbers. Further, all characters that had experienced too many changes, i.e. were approaching or at saturation, were deleted from the alignment.

DETERMINATION OF DIVERGENCE TIMES OF METAZOAN EVOLUTION

Porifera and Cnidaria in the above discussion were included as an outgroup. In this section I will apply the trifurcation rate test to some metazoan phyla. In this case, plants and fungi will serve as the outgroups. Six metazoan phyla are analyzed and include representatives from vertebrate, echinoderms, arthropods and mollusks as well as Porifera and Cnidaria.

The number of metazoan 18S sequences deposited in GenBank is about 4500. A number of criteria were used to pick the 29 used in this study. First, I picked out about 100 full-length 18S sequences somewhat arbitrarily to include equal numbers of protosomes and deuterosomes. An effort was made to align them automatically – many were too highly diverged to align and these were eliminated. Finally, the database was broken up into smaller groups and distance trees were computed. There is huge variation in the rate of evolution in the 18S rRNA sequences. The rates varied from those that were similar to plant and fungi rates to others that were two to three times faster. It was found that the trifurcation rate test completely failed when those fast evolving sequences were compared with the slower evolving sequences.

The failure manifested itself in tree inconsistencies; namely, we could see obvious examples of "long branch attractions" and, more importantly, cases where patristic distances through the minimal tree scattered widely from the d_{ij} values. (These examples are not shown here.) To avoid this problem, all sequences that gave rise to long branches were excluded from the analysis. After identifying long branches, sequences were recompiled, aligned and edited, and new trees constructed. Figure 30.5 shows that tree. For the purpose of an outgroup comparison, five land plants and five fungi were included. This tree has a number of features that are expected. Arthropods and mollusks define a subclade to the exclusion of vertebratae and echinoderms – this is consistent with known relationships between protosomes and deuterosomes. It is also clear that the major metazoan clade displays a greatly accelerated rate of evolution compared with the rest of the eukaryotes, even though the fastest evolving metazoan taxa were removed from the analysis, as can be seen from branch lengths from the root (noted by the arrow). These patterns are consistent with earlier studies. There are two unexpected details implied by this tree. One is that Cnidaria (e.g. jellyfish) and Porifera (the sponges) diverged from the other metazoans at a time very close to the fungal, plant and metazoan radiation. The second is that Porifera appears in a clade that includes choanoflagellates (such as chlamydomonas and volvox) and chytrids. (The association with plants that is seen is too deep to be of much significance.) The association of Porifera with choanoflagellates and chytrids is more significant. This result suggests that if the sponges are considered metazoan, then the metazoans are at least biphyletic. A clade that includes what was previously thought of as (1) a metazoan, (2) a single-celled plant, and (3) a fungus, is not really totally unexpected. Namely, it has been noted, even prior to molecular evidence, that these three taxa seem to have similar choanocyte flagella (Mohri et al., 1995).

The metazoan phyla shown in Figure 30.6 were submitted to the same trifurcation rate test as was done for the plants above (Table 30.3) and results are summarized in Figure 30.6. As can be seen, when distances of multiple taxa with the various clades are averaged, we can obtain a distance tree whose patristic distances reasonably reconstruct d_{ij} values. As is clear from the tree in Figure 30.7, the Cnideria and Porifera diverged from the other metazoans about 1 Byr ago. In addition, the major diversification of metazoan phyla shows a diversification of 430–500 Myr ago, times not at great odds with the metazoan radiation of 530 Myr ago. This latter fact lends credence to the major finding of this analysis – namely that Cnidaria and

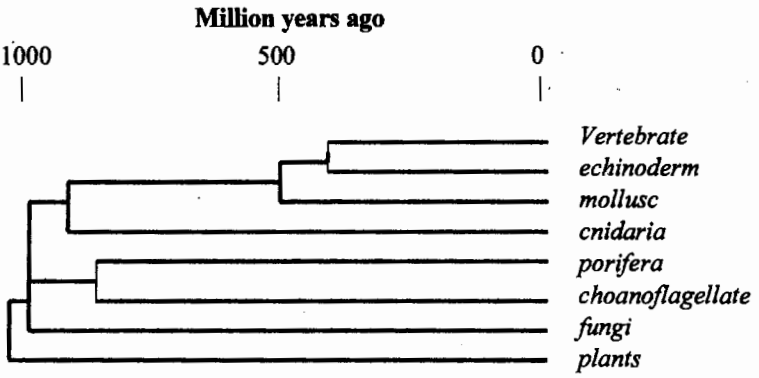

FIGURE 30.6 Time of divergence of major metazoan groups.

TABLE 30.3 Average distances between major metazoan clades. The taxa used in the averages are shown in the tree in Figure 30.5. See Table 30.2 legend for details. Patristic distances are for a minimal distance tree not shown (see Figure 30.6 for topology)

	1	2	3	4	5	6	7
1. Plants		12.31 (12.3)	10.50 (11.1)	12.62 (12.1)	18.28 (18.3)	18.27 (17.6)	17.44 (18.1)
2. Fungi			9.15(9.2)	10.62 (10.26)	16.30 (16.3)	16.53 (15.6)	15.21 (16.1)
3. Porifera				8.43(9.0)	15.86 (15.2)	16.53 (14.6)	15.35 (15.1)
4. Cnidaria					15.76 (16.2)	15.36 (15.7)	14.81 (16.1)
5. Vertebrata						11.01 (10.89	10.06 (10.1)
6. Mollusca							10.31 (10.2)
7. Echinoderm							

Porifera had diverged from the other metazoans 0.85–1.0 Byr ago.

FURTHER DISCUSSION

Evolution of plants

The rates of land plant 18S ribosomal RNA sequence evolution estimated in the current work, coupled with the divergence data, indicate that the higher taxonomic groups such as monocots and dicots, gymnosperms and angiosperms have roots that are much deeper in time than is shown by the fossil record (Figure 30.4). Let us first consider the monocots and dicots within the angiosperm clade. To be sure, taken by itself, the 18S rRNA data support a monophyletic ancestry for the angiosperms because they show that all angiosperms evolved from a common ancestor that excludes other extant plant clades. However, the angiosperms are defined by multiple morphological characters which are apparent in the fossil record, not only by their 18S rRNA. This record shows that modern flowering plants appeared only 110 Myr ago, though some, but not all, of their defining characters have been found as far back as 140 Myr ago (see Chapter 29). However, the data in this paper, and in earlier molecular clock calculations, place, for example, the monocot–dicot divergence time estimates back greater than 170 Myr ago (Wolfe et al., 1989; Savard et al., 1994). These early divergence times have been seen for cytochrome *c* evolution as well (Syvanen, 1994). Among the dicots, there are also multiple families whose divergence also goes back as far; in the current work, I have seen that all extant Rosidae are descended from a common ancestor but they diverged from the other dicots at this early date (not shown). This leads to the inescapable conclusion that multiple plant lineages

existed in Jurassic (and probably the late Permian) that left no fossils recognizable as angiosperms, but evolved to become what today are the flowering plants. Krassilov (1977, 1998; see Chapter 29) describes this as the "angiospermization" of multiple lineages.

The divergence time of 340 Myr ago for the angiosperm and gymnosperm means that the angiosperms had already diverged from the line leading to modern gymnosperm before the appearance of fossil gymnosperms, early after the appearance of the first progymnosperms. This means the tracheophytes are also likely to be polyphyletic. And finally, the results in this paper show that the line leading to tracheophytes diverged from that leading to the modern bryophytes in the early Devonian, a time prior to the appearance of any vascular plants in the fossil record. The conclusion that major clades of vascular plants may be polyphyletic has also been drawn from an analysis of Devonian and Carboniferous fossils (reviewed in Bell, 1992).

One common explanation for this timing problem is that the plant fossil record is incomplete; this is known as the *hidden-forests* hypothesis by critics of paleobotany. This is not an attractive idea; it demands gaps in the fossil record that extend over tens of millions of years and involves multiple lineages. A much more parsimonious explanation is that major taxonomic groups are polyphyletic.

Polyphyly means that major characters that are shared by different angiosperm lineages must have arisen by extensive parallelisms. (E. Mayr has attempted to define this situation as "parallelophyly," and V. Krassilov has coined "pachyphyly.") To use Lankester's terminology, these characters are homoplastic, not homologous. Classically, parallel developments are explained by the general morphogenetic potential found in the various plant lineages and are inherited linearly and differentially expressed by environmental factors. This, in fact, was the hypothesis that Went explicitly addresses in his 1972 review; he argues against the possibility of independent evolution because patterns of homoplastic characters that he described occur throughout the same geographical areas. In different geographical regions, different suites of homoplastic characters are encountered, that is, different lineages on different continents display a lack of homoplasy even though they contain the same morphogenetic potential. However, when the lineages share a habitat, parallel development occurs. Vavilov (1922) also noticed the tendency of plants to vary in parallel and called the phenomenon "homologous variation" which may be an accurate explanation for variation that is determined by transfer of homologous genes.

Today, given what we know of molecular genetic mechanisms and molecular evolution, I would propose that the simplest explanation of polyphyly at this level is that horizontal gene transfer has played a major role in the evolution of higher plants. I am not claiming any examples of horizontally transferred genes among the 18S rRNAs studied here. Rather, I am claiming that many of the major morphological traits, as revealed in the fossil record, that are shared and therefore characterize higher taxonomic groups, are controlled by genes that have horizontally transferred during the emergence of that group (Syvanen, 1994). This hypothesis will be given a direct test once those genes controlling major developmental patterns that distinguish angiosperm families and genera are identified and phylogenetic analysis of them is possible.

Evolution of the metazoa

The current finding that Porifera, Cnidaria and the other metazoans diverged from each other near the time of the plant, fungi and metazoan trifurcation is strongly supported. If this is the case, then either metazoans are polyphyletic (at least triphyletic) or metazoans do not include Porifera and Cnidaria. Indeed for more than a century, the sponges were not even considered metazoa, but given their own kingdom of parazoa. West and Powers (1993) uncovered molecular evidence that sponges had an unexpected close affinity to some protozoans but they provided a different interpretation to the one given here. In either case, this situation points out a major case of parallel evolution. Sponges appear in the fossil record at the same moment as the other metazoans – during the Cambrian radiation. During this time, we may presume that the

ancestor to the modern sponges learned how to utilize $CaCO_3$ or, for the glass sponges, silicates for skeleton construction. This ability appeared simultaneously in sponges, in other metazoans and in numerous protozoan lineages. At this stage, the simplest explanation is that the genes controlling the construction of these hard inorganic skeletons spread like an infection across multiple lineages. The abrupt appearance of hard fossils at the beginning of the Cambrian has been understood for over 100 years as representing an example of massive parallelism and has been cited as supporting a major role for horizontal gene transfer in metazoan evolution (Reanney, 1976; Erwin and Valentine, 1984; Syvanen, 1985; Jeppsson, 1986).

A number of molecular studies, based on protein molecular clocks, have placed the major metazoan radiation (such as divergence of chordates from other in vertebratae or deutrosomes from protostomes) to times back in the Precambrian (Wray et al., 1996; Bromham et al., 1998, 2000; Gu, 1998; Cutler, 2000). The current study does not support these estimates: indeed the protostome/deuterostome divergence estimated here is consistent with the time of the Cambrian radiation. This difference between the rRNA and protein time estimates will have to be resolved in the future.

REFERENCES

Baldauf, S.L. and Palmer, J.D. (1993) Animals and fungi are each other's closest relatives: congruent evidence from multiple proteins. *Proc. Natl Acad. Sci. USA* **90**: 11558–11562.

Bell, P.R. (1992) *Green Plants; Their Origin and Diversity*, Dioscorides Press, Portland, OR.

Boulter, D. and Gilroy, J.S. (1992) Partial sequences of 18s ribosomal RNA of two genera from each of six flowering plant families. *Phytochemistry* **31**: 1243–1246.

Bromham, L., Rambaut, A., Fortey, R. et al. (1998) Testing the Cambrian explosion hypothesis by using a molecular dating technique. *Proc. Natl Acad. Sci. USA* **95**: 12386–12389.

Bromham, L.D. and Hendy, M.D. (2000) Can fast early rates reconcile molecular dates with the Cambrian explosion?. *Proc. Roy. Soc. Biol. Sci. Series B* **267**: 1041–1047.

Chaw, S.M., Sung, H.M., Long, H. et al. (1995) The phylogenetic positions of the conifer genera *Amentotaxus, Phyllocladus,* and *Nageia* inferred from 18S rRNA sequences *J. Mol. Evol.* **41**: 224–230

Chaw, S.M., Zharkikh, A., Sung, H.M. et al. (1997) Molecular phylogeny of extant gymnosperms and seed plant evolution: analysis of nuclear 18S rRNA sequences. *Mol. Biol. Evol.* **14**: 56–68

Cutler, D.J. (2000) Estimating divergence times in the presence of an over-dispersed molecular clock. *Molec. Biol. Evol.* **17**: 1647–1660.

Donoghue, M.J. and Alverson, W.S. (2000) A new age of discovery. *Ann. Missouri Bot. Garden* **87**: 110–126.

Donoghue, M.J. and Doyle, J.A. (2000) Seed plant phylogeny: Demise of the anthophyte hypothesis? *Curr. Biol.* **10**: R106–R109.

Doolittle, R.F., Feng, D.F., Tsang, S. et al. (1996) Determining divergence times of the major kingdoms of living organisms with a protein clock. *Science* **271**: 470–477.

Doyle, J.A., Donoghue, M.J. and Zimmer, E.A. (1994) Integration of morphologic and ribosomal RNA data on the origin of angiosperms. *Ann. Missouri Bot. Garden* **81**: 419–450.

Drouin, G., Moniz de Sa, M. and Zuker, M. (1995) The *Giardia lamblia* actin gene and the phylogeny of eukaryotes. *J. Molec. Evol.* **41**: 841–849.

Erwin, D.H. and Valentine, J.W. (1984) "Hopeful monsters", transposons, and metazoan radiation. *Proc. Natl Acad. Sci. USA* **81**: 5482–5483.

Fitch, W.M. and Margoliash, E. (1967) A method for estimating the number of invariant amino acid coding positions in a gene using cytochrome c as a model case. *Biochem. Genetics* **1**: 65–71.

Gu, X. (1998) Early metazoan divergence was about 830 million years ago. *J. Molec. Evol.* **47**: 369–371.

Haeckel, E. (1866) *General Morphology of Organisms*, reprinted Walter de Gruyter, Berlin, 1988.

Jeppsson, L. (1986) A possible mechanism in convergent evolution. *Paleobiology* **12**: 337–344.

Jin, L. and Nei, M. (1990) Limitations of the evolutionary parsimony method of phylogenetic analysis. *Molec. Biol. Evol.* **7**: 82–102.

Jukes, T.H. and Cantor, C.R. (1969) Evolution of protein molecules. In *Mammalian Protein Metabolism* (ed., H.N. Munro), vol. III, pp. 21–132, Academic Press, New York.

Krassilov, V.A. (1977) The origin of angiosperms. *Bot. Rev.* **43**: 143–176.

Krassilov, V.A. (1998) Character parallelism and reticulation in the origin of the Angiosperms. In *Horizontal Gene Transfer* (eds, M. Syvanen. and C. Kado), pp. 409–424, Chapman & Hall, London.

Kumar, S. and Rzhetsky, A. (1996) Evolutionary relationships of eukaryotic kingdoms. *J. Molec. Evol.* **42**: 183–193.

Lankester, E.R. (1870) On the use of the term homology in modern zoology and the distinction between homogenetic and homoplastic agreements. *Ann. Mag. Natl Hist.* **6**: 34–43.

Lee, M.S.Y. (1999) Molecular clock calibrations and metazoan divergence dates. *J. Molec. Evol.* **49**: 385–391.

Mathews, S. and Donoghue, M.J. (1999) The root of angiosperm phylogeny inferred from duplicate phytochrome genes. *Science (Washington DC)* **286**: 947–950.

Mohri, H., Kubo-Irie, M. and Irie, M. (1995) Outer arm dynein of sperm flagella and cilia in the animal kingdom. *Mem. Mus. Natl d'Hist. Naturelle* **166**: 15–22.

Nikoh, N., Hayase, N., Iwabe, N. et al. (1994) Phylogenetic relationship of the kingdoms Animalia, Plantae, and Fungi, inferred from 23 different protein species. *Molec. Biol. Evol.* **11**: 762–768.

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

Reanney, D. (1976) Extrachromosomal elements as possible agents of adaptation and development. *Bacteriolog. Rev.* **40**: 552–590.

Romano, S.L. and Palumbi, S.R. (1996) Evolution of scleractinian corals inferred from molecular systematics. *Science (Washington DC)* **271**: 640–642.

Sarich, V.M. and Wilson, A.C. (1973) Generation time and genomic evolution in primates. *Science* **179**: 1144–1147.

Savard, L., Li, P., Strauss, S.H. et al. (1994) Chloroplast and nuclear gene sequences indicate late Pennsylvanian time for the last common ancestor of extant seed plants. *Proc. Natl Acad. Sci. USA* **91**: 5163–5167.

Sidow, A. and Thomas, W.K. (1994) A molecular evolutionary framework for eukaryotic model organisms. *Curr. Biol.* **4**: 596–603.

Sogin, M.L., Morrison, H.G., Hinkle, G. and Silberman, J.D. (1996) Ancestral relationships of the major eukaryotic lineages. *Microbiologia* **12**: 17–28.

Soltis, P.S., Soltis, D.E., Wolf, P.G. et al. (1999) The phylogeny of land plants inferred from 18S rDNA se- quences: Pushing the limits of rDNA signal? *Molec. Biol. Evol.* **16**: 1774–1784.

Soltis, P.S. and Soltis, D.E. (1997) Phylogenetic analysis of large molecular data sets. *Bol. Soc. Bot. Mexico* **59**: 99–113.

Soltis, D.E., Soltis, P.S., Nickrent, D.L. et al. (1997) Angiosperm phylogeny inferred from 18S ribosomal DNA sequences. *Ann. Missouri Bot. Garden* **84**: 1–49.

Soltis, P.S., Soltis, D.E. and Chase, M.W. (1999) Angiosperm phylogeny inferred from multiple genes as a tool for comparative biology. *Nature (London)* **402**: 402–404.

Sorhannus, U. (1996) Higher ribosomal RNA substitution rates in Bacillariophyceae and Dasycladales than in Mollusca, Echinodermata, and Actinistia-Tetrapoda. *Molec. Biol. Evol.* **13**: 1032–1038.

Stevens, P.F. (1984) Metaphors and typology in the development of botanical systematics 1690–1960 or the art of putting new wine in old bottles. *Taxon* **33**: 169–211.

Syvanen, M. (1985) Cross-species gene transfer: implications for a new theory of evolution. *J. Theoret. Biol.* **112**: 333–343.

Syvanen, M., Hartman, H. and Stevens, P.F. (1989) Classical plant taxonomic ambiguities extend to the molecular level. *J. Molec. Evol.* **28**: 536–544.

Syvanen, M. (1994) Horizontal gene transfer: Evidence and possible consequences. *Annu. Rev. Genetics* **28**: 237–261.

Valentine, J.W., Jablonski, D. and Erwin, D.H. (1999) Fossils, molecules and embryos: New perspectives on the Cambrian explosion. *Development (Cambridge)* **126**: 851–859.

Vavilov, N. (1922) Law of homologous series in variation. *J. Genet.* **12**: 47–89.

Van De Peer, Y.A.D. and De Wachter, R. (1997) Evolutionary relationships among the eukaryotic crown taxa taking into account site-to-site rate variation in 18S rRNA. *J. Molec. Evol.* **45**: 619–630.

Went, F.W. (1972) Parallel evolution. *Taxon* **20**: 197–226.

West, L. and Powers, D. (1993) Molecular phylogenetic position of hexactinellid sponges in relation to the Protista and Demospongiae. *Mol. Marine Biol. Biotechnol.* **2**: 71–75.

Wolfe, K.H., Gouy, M., Yang, Y.W. et al. (1989) Date of the monocot-dicot divergence estimated from chloroplast DNA sequence data. *Proc. Natl Acad. Sci. USA* **86**: 6201–6205.

Wray, G.A., Levinton, J.S. and Shapiro, L.H. (1996) Molecular evidence for deep Precambrian divergences among metazoan phyla. *Science (Washington DC)* **274**: 568–573.

Yokoyama, S. and Harry, D.E. (1993) Molecular phylogeny and evolutionary rates of alcohol dehydrogenases in vertebratas and plants. *Mol. Biol. Evol.* **10**: 1215–1226.