INTEGRATION OF MORPHOLOGICAL AND RIBOSOMAL RNA DATA ON THE ORIGIN OF ANGIOSPERMS[1]

James A. Doyle,[2]
Michael J. Donoghue,[3]
and Elizabeth A. Zimmer[4]

ABSTRACT

Previous phylogenetic analyses of morphological and rRNA data indicated that Gnetales are the closest living relatives of angiosperms but gave different basal angiosperm relationships. A two-step morphological analysis of seed plants (including fossils) and angiosperms rooted the latter near Magnoliales, with tricolpate eudicots and paleoherbs (herbaceous magnoliids and monocots) forming a clade, whereas analyses of rRNA sequences rooted angiosperms among paleoherbs, with eudicots and woody magnoliids forming a clade. Experiments with a revised seed plant morphological data set raise further questions: when angiosperms are scored like different angiosperm subgroups, they associate with different outgroups, although Gnetales are their closest living relatives. To test whether morphological and rRNA data are seriously contradictory or rather complementary, with inconsistencies being a function of better resolution in different parts of the tree, we experimented with morphological and rRNA data sets including the same six extant "gymnosperm" and 12 angiosperm taxa. Both analyses again associate angiosperms and Gnetales. The morphological analysis differs from previous ones in placing Nymphaeales and monocots at the base of the angiosperms, but trees rooted next to Magnoliales are only one step less parsimonious. As in previous studies, the rRNA analysis roots angiosperms next to Nymphaeales and breaks up the eudicots. Bootstrap and decay analyses of the rRNA data show strong support for the monophyly of angiosperms and Gnetales and their sister group relationship, but low support for groupings within angiosperms. However, one or another group of paleoherbs is basal in most bootstrap trees. A combined analysis favors a paleoherb rooting, but other relationships agree with the morphological results; in particular, eudicots form a clade. The conclusion that Gnetales are the closest living relatives of angiosperms permits a wide range of morphological scenarios, depending on the arrangement of fossil outgroups. Discovery of fossils on the long branch leading to angiosperms, methods of factoring out artifacts in rooting, or molecular data on the control of floral morphogenesis in angiosperms and Gnetales may be required for further progress in unraveling the origin of angiosperms.

Recent cladistic analyses of morphological and ribosomal RNA data have led to apparently conflicting results on the position of angiosperms among seed plants, basal relationships among angiosperms, and resulting scenarios for the origin of angiosperms (Crane, 1985; Doyle & Donoghue, 1986, 1992; Donoghue & Doyle, 1989a; Zimmer et al., 1989; Loconte & Stevenson, 1990, 1991; Taylor & Hickey, 1992; Hamby & Zimmer, 1992). The purpose of this paper is to investigate the causes and significance of these conflicts, and what if any robust conclusions concerning the origin of angiosperms can be drawn from these data.

Our initial morphological analysis of seed plants (Doyle & Donoghue, 1986), which included both living and fossil taxa, treated angiosperms as a single taxon, modeled on Magnoliales and Winteraceae, following the consensus at that time. This study was designed especially to test the previous analysis of Crane (1985) and other current ideas on seed plant phylogeny, by including both characters used by Crane and conflicting similarities that he had omitted. Contrary to our expectations, but as Crane had found, this analysis indicated that seed plants are a monophyletic group, with coniferopsids (cordaites, conifers, ginkgos) derived from

advanced "seed ferns" with platyspermic seeds and saccate pollen (roughly as proposed by Rothwell, 1982), and that angiosperms belong to an "anthophyte" clade also including Bennettitales, *Pentoxylon*, and Gnetales, which is in turn nested among so-called Mesozoic seed ferns (corystosperms, glossopterids, *Caytonia*). In response to a cladistic analysis involving only extant seed plants by Loconte & Stevenson (1990), we revised this analysis with generally similar results (Doyle & Donoghue, 1992), except for more definite placement of cycads among platysperms and more uncertainty on the position of ginkgos, which may be associated with the Permian–Triassic seed fern *Peltaspermum* instead of conifers (cf. Meyen, 1984).

A subsequent analysis of extant angiosperms (Donoghue & Doyle, 1989a), using the results of the seed plant study to polarize characters within the group, yielded trees rooted in or next to the "woody magnoliid" order Magnoliales (specifically families with granular exine structure, which excludes Winteraceae and *Austrobaileya*). This result is generally consistent with conventional views on angiosperm evolution (e.g., Takhtajan, 1969; Cronquist, 1981, 1988). The remaining angiosperms form four major clades: two other groups of woody magnoliids, (1) Laurales, including Chloranthaceae, widely discussed because of their unusually simple flowers, and (2) winteroids, including Winteraceae, Illiciales, and possibly Canellaceae; (3) dicots with tricolpate and derived pollen, later designated eudicots (Doyle & Hotton, 1991); and (4) "paleoherbs," consisting of herbaceous or semiherbaceous magnoliids (Aristolochiaceae, *Lactoris*, Piperales, Nymphaeales) and monocots. Somewhat similar results were obtained by Loconte & Stevenson (1991); the most important difference, placement of Calycanthaceae and Idiospermataceae at the base of the angiosperms, may be due to their use of Recent plants only as outgroups, since the closest outgroup, Gnetales, shares features such as opposite leaves and two-trace nodes with Calycanthales.

In contrast, analyses of partial 18S and 26S rRNA sequences from angiosperms and other extant seed plants (Hamby & Zimmer, 1992), chosen to provide a test of current morphological hypotheses, placed a series of paleoherb taxa at the base of the angiosperms (Nymphaeales, Piperales, and monocots). Woody magnoliids and eudicots formed a more derived clade, within which detailed relationships are poorly resolved. These results give a different picture of the first angiosperms: they would be herbaceous or nearly so, with palmately veined leaves and anomocytic stomata, rather than woody with pinnately veined leaves and paracytic stomata. This recalls the views of Burger (1977, 1981) on the primitive status of Piperales and/or monocots, although not in detail.

More recently, the view that the first angiosperms were "paleoherbs" was elaborated by Taylor & Hickey (1990, 1992), based on recognition of an Early Cretaceous (Aptian) paleoherb fossil from Australia and their own morphological analysis. Their tree differs in that Chloranthaceae are basal, followed by Piperaceae, which contrasts with the position of Chloranthaceae in Laurales in the analysis of Donoghue & Doyle (1989a), well removed from Piperales, and in the woody magnoliid-eudicot clade in Hamby & Zimmer (1992). This difference may be partly a result of treating correlated ovule features related to orthotropy (which was assumed to be primitive) as three separate characters, and partly a result of omitting taxa that linked groups differently in the Donoghue & Doyle analysis (e.g., Trimeniaceae, which tend to link Chloranthaceae with other Laurales: cf. Endress, 1987).

The apparent conflicts between the morphological and molecular data, especially regarding "rooting" of the angiosperms, were discussed by Donoghue & Doyle (1989b), who suggested that they may illustrate a general conclusion drawn by Hillis (1987): that conflicts between morphological and molecular results rarely reflect serious contradictions, but rather different levels of resolution of the two sorts of data in different parts of the tree. This seemed to be supported by the fact that experiments with alternative trees in the seed plant studies (Doyle & Donoghue, 1986, 1992) had shown that other positions of anthophytes within seed plants, and of angiosperms within anthophytes, were only slightly less parsimonious than the most parsimonious arrangements. For example, it was only one step less parsimonious to link angiosperms with *Caytonia* and/or glossopterids, thus breaking up the anthophytes, or to link anthophytes with coniferopsids, with Gnetales basal in anthophytes—what we called a neo-englerian arrangement. Similarly, in our morphological analysis of angiosperms (Donoghue & Doyle, 1989a), we had found trees with angiosperms rooted among paleoherbs that were only one step less parsimonious than those rooted near Magnoliales. This may therefore be a case where the morphological data are ambiguous—they favor a woody magnoliid prototype, but only slightly—whereas the molecular data point strongly in one direction, toward a paleoherb rooting, and therefore provide better evidence on relationships. In other cases, it may be the molecular data

that are ambiguous and the morphological data that are clearcut. Thus the two sorts of data may be more complementary than contradictory.

In the present study, we have attempted to address these issues by comparing results derived from morphological and molecular data for the same set of seed plant taxa, probing the strengths and weaknesses of the results with methods such as bootstrap and decay analysis, and analyzing a combined data set. The question of analyzing morphological and molecular data separately and comparing the results or combining them at the outset is a topic of ongoing debate (Kluge, 1989; Barrett et al., 1991; Donoghue & Sanderson, 1992; Bull et al., 1993; de Queiroz, 1993). One argument against combining data sets is that the greater number of molecular characters will simply overwhelm the morphological characters. However, this does not necessarily hold: if the molecular results are poorly resolved, as they often are, even a small number of morphological characters can have a decisive effect (Donoghue & Sanderson, 1992). In any case, it is possible both to analyze data sets separately and to combine them, and our results imply that this approach can give instructive results.

This study is not intended to be a comprehensive examination of "morphological" versus "molecular" data on this topic. Y. Suh (pers. comm.) has obtained different results from another part of the 26S subunit of rDNA, which roots the angiosperms between a clade including most Magnoliales and Laurales and other angiosperms. Equally different trees, with the aquatic genus *Ceratophyllum* basal and remaining angiosperms divided into tricolpate-derived eudicots and monosulcate magnoliids and monocots, have been obtained from *rbc*L sequences (Les et al., 1991; Chase et al., 1993; Qiu et al., 1993). Analyses of shorter rRNA sequences by Troitsky et al. (1991) and a smaller *rbc*L data set by Hasebe et al. (1992) have given different trees, most notably with gymnosperms as a monophyletic group. Instead of considering all these data sets, we have chosen to address the two that we know best, although we will mention briefly some preliminary analyses including *rbc*L.

COMBINING PREVIOUS MORPHOLOGICAL ANALYSES OF SEED PLANTS AND ANGIOSPERMS

A major concern that we wish to address first is the possibility that previous inferences concerning the origin of angiosperms were compromised by circular reasoning. Perhaps the position of angiosperms in the seed plant analysis (Doyle & Don-

oghue, 1986) was a function of the assumption that the first angiosperms were like Magnoliales and Winteraceae. Perhaps then the basal position of Magnoliales in the angiosperm analysis (Donoghue & Doyle, 1989a) was a consequence of an incorrect identification of outgroups based on this initial assumption. For the sake of rapid progress, we split the seed plant and angiosperm problems in two. This seemed reasonable at the time: there was already strong evidence that the angiosperms were monophyletic, and there was some consensus on basic states within the group. However, we realized from the beginning that it would eventually be necessary to carry out additional analyses designed to resolve simultaneously relationships at the point where the two analyses intersect. For example, we noted that Chloranthaceae share many features with Gnetales, and that angiosperms might therefore be directly associated with Gnetales if Chloranthaceae were assumed to be primitive.

It should be recognized that conclusions derived from this sort of two-step procedure are not necessarily incorrect: there might be only one most parsimonious position for angiosperms no matter what internal relationships or basal states in angiosperms are assumed. It should also be noted that the problems are not unique to our study: ours is simply one example of a general method that Mishler (1994) calls compartmentalization. This method was also used within our angiosperm analysis (Donoghue & Doyle, 1989a), in which we did a preliminary analysis of Laurales to determine basic states in a derived subgroup referred to as "core Laurales."

These problems were recognized and addressed in a series of experiments reported by Doyle & Donoghue (1990), which combined nine angiosperm taxa with the 17 nonangiospermous groups used in Doyle & Donoghue (1992). Here we present an updated version of these experiments, which illustrate the nature of the problem and the potential value of considering both morphological and molecular evidence.

TAXA, CHARACTERS, AND ANALYSES

The revised seed plant and angiosperm matrix, henceforth designated the nine-angiosperm analysis, is presented in the Appendix (Table 1). The angiosperm taxa were selected to represent the major clades found by Donoghue & Doyle (1989a) in trees rooted both near Magnoliales and among paleoherbs, which necessitated dividing the paleoherbs into four groups. In five cases these clades are represented by individual taxa used in Dono-

ghue & Doyle (1989a), in four by composite taxa: Magnoliales, based on *Degeneria*, Myristicaceae, Annonaceae, and Magnoliaceae ("core Magnoliales" of Donoghue & Doyle, 1989a); Piperales (Piperaceae, Saururaceae); Nymphaeales (Nymphaeaceae, Cabombaceae); and eudicots (Ranunculidae, *Nelumbo*, Trochodendrales, Hamamelidales). These are scored in terms of estimated ancestral states for the whole taxon (theoretical and practical problems in this procedure are discussed further below). Thus, where the taxon consists of two taxa included in the previous analysis and these have different stages, the group was scored as uncertain (e.g., 0/1). In the case of binary characters, where uncertain (0/1) and unknown (?) are equivalent in tree construction, we usually distinguished between the two scorings in the matrix to indicate the nature of the uncertainty, but we did not try to weed out all cases where "?" had been used for uncertainty in the previous analyses. In Magnoliales, basic states were estimated based on the previous result that *Degeneria* is the sister group of the other three taxa, which themselves form an unresolved trichotomy. Eudicots were assumed to consist of two sister clades, Ranunculidae plus *Nelumbo* and Trochodendrales plus Hamamelidales. Laurales were represented by *Austrobaileya*, which was at or near the base of the order in our previous trees, and Chloranthaceae, which are of special interest because they have been widely discussed as possible primitive angiosperms. Use of *Austrobaileya* to represent Laurales might be questioned, since it lacks many features commonly associated with the order. However, our results bear out its use in this way, since *Austrobaileya* is associated with Chloranthaceae in the trees obtained, as it was with the larger data set. Trees were rooted by including a taxon based on Devonian "progymnosperms" (*Aneurophyton, Archaeopteris*) as outgroup.

Characters are primarily a combination of those used in the seed plant analysis of Doyle & Donoghue (1992) and those in the angiosperm analysis of Donoghue & Doyle (1989a) that are potentially informative for the taxa under consideration. To these we added several apomorphies that potentially hold angiosperms together as a monophyletic group, either as new characters or as additional states of existing seed plant characters (e.g., three-nucleate microgametophyte, complete loss of the megaspore wall). In some cases, features that vary within angiosperms are expressed as additional states of characters recognized in seed plants as a whole (e.g., palmate leaf venation). For angiosperm characters in which there are no clearly comparable

states in other seed plants, we scored the latter as unknown; this is especially true of floral characters, where we hoped to avoid biasing the results by assuming questionable homologies of parts between angiosperms and other groups. These characters also form the basis for the morphological analysis of Recent taxa presented below as a counterpart of the rRNA analysis; for simplicity we retained nine states that were potentially informative in that matrix but autapomorphic in the present one (e.g., tetracytic stomates in Piperales, tricolpate pollen in eudicots).

Multistate characters were unordered, except for two easily ordered quantitative characters (pollen size, megaspore wall thickness). Two multistate characters involving exine structure deserve special consideration, since our decision not to order them had a significant impact on the results presented below.

In the angiosperm study (Donoghue & Doyle, 1989a), we recognized an infratectal structure character with granular and columellar states and assumed that granular was primitive, based on outgroup comparison with other anthophytes. In the revised seed plant study (Doyle & Donoghue, 1992), we recognized spongy-alveolar, honeycomb-alveolar, and granular states and scored angiosperms as granular, assuming that columellar structure evolved within the group. In combining the two data sets, our previous polarization of states within angiosperms could be preserved by ordering the character (spongy–honeycomb–granular–columellar). By placing two steps between alveolar and columellar, this ordering would bias against trees in which angiosperms are linked with alveolar outgroups (e.g., *Caytonia*, glossopterids), columellar groups are basal in angiosperms, and granular structure in groups like Magnoliales is a convergence with Bennettitales, *Pentoxylon*, and Gnetales. This scenario involves four steps if the exine character is ordered but three if it is not. Such a bias might be defended based on the coexistence of columellar and granular structure (and transitional states) within angiosperms (Walker, 1976; Le Thomas, 1980–1981) and of alveolar and granular structure within conifers (Van Campo & Lugardon, 1973). However, it seems unwarranted to assume that a direct transition from alveolar to columellar could not have occurred in other cases. Trees with angiosperms linked with *Caytonia* were only one step longer than the shortest trees in terms of the Doyle & Donoghue (1986, 1992) data sets, and a transition from alveolar to granular is certainly conceivable on structural grounds (e.g., reduction of the side-walls of the alveolae, leaving

Volume 81, Number 3
1994

Doyle et al.
Integration of Morphological and rRNA Data

423

the junctions between the walls as columellae: E. Masure, pers. comm.).

Endexine structure was not included as a character in the seed plant analyses (Doyle & Donoghue, 1986, 1992), because all groups except angiosperms have a uniformly laminated endexine, making the character uninformative. Of the two states in angiosperms, endexine present or absent in the extra-apertural areas, absence of endexine was assumed to be ancestral (Donoghue & Doyle, 1989a). This was based on the hypothesis that the laminated endexine of other seed plants is homologous with the footlayer of angiosperms, since both develop from similar tangential lamellae, and that the endexine of angiosperms, which is nonlaminated except sometimes under the apertures, is a new layer (Zavada, 1984). Again, the original polarization could be preserved in the combined data set by ordering the character (endexine laminated–absent–nonlaminated). However, because the homologies involved are rather speculative (cf. Gabarayeva, 1991), it seems preferable to treat the three states as unordered.

We also made a few substantive changes based on new data, such as recognition that pollen of Piperaceae (Piperales) has supratectal spinules and a sculptured sulcus (Bornstein, 1989), Cabombaceae (Nymphaeales) have a columellar exine structure (Osborn et al., 1991), Myristicaceae (Magnoliales) have both S and PI type sieve-tube plastids, and Aristolochiaceae have basically PII type plastids (as seen in *Saruma* and *Asarum*), like monocots (Behnke, 1988). We added one new potential synapomorphy of angiosperms and Gnetales, double fertilization, in the sense of regular fusion of the second sperm nucleus with a second megagametophyte nucleus. This has been confirmed in *Ephedra* and seems independent of (although probably a prerequisite for) the uniquely angiospermous feature of endosperm formation (Friedman, 1990, 1992; Donoghue & Scheiner, 1992).

Several sets of analyses were performed using the new nine-angiosperm matrix. In one set, we analyzed the matrix with each of the nine angiosperm groups substituted individually for angiosperms. In removing the eight remaining taxa, we retained characters that then became autapomorphies; it is often desirable to remove such characters because they are uninformative and distort measures of homoplasy, but this is not a problem in the present case, where we are not concerned with relative levels of homoplasy. However, we did remove invariant characters with MacClade (Maddison & Maddison, 1992). All these data sets were analyzed with PAUP (version 3.0L, Swofford,

1991). Since there are too many taxa for branch-and-bound analysis, which guarantees finding all most parsimonious trees, we used the heuristic search algorithm, with 10 replicates of stepwise random addition of taxa and TBR branch swapping. This increases the probability of finding most parsimonious trees that belong to different "islands" (Maddison, 1991), which were in fact found in some experiments (see results). Alternative arrangements were investigated using MacClade and the constraints option in PAUP.

SINGLE-ANGIOSPERM ANALYSES

Experiments with single angiosperm taxa resulted in several different positions of the angiosperms relative to other anthophytes (*Pentoxylon*, Bennettitales, Gnetales), and of anthophytes within seed plants.

As expected, when Magnoliales are substituted for angiosperms (Fig. 1), angiosperms are the sister group of other anthophytes, and anthophytes are associated with one or another combination of *Caytonia*, glossopterids, and corystosperms, as in Doyle & Donoghue (1986, 1992). Relationships among other groups also parallel those found by Doyle & Donoghue (1992), including some with the arrangement of extant cycads, *Ginkgo*, and conifers found by Loconte & Stevenson (1990). As shown in Figure 1, one of the characters that supports this position of Magnoliales is the presence of granular exine structure.

In contrast, when Winteraceae, *Austrobaileya*, eudicots, Aristolochiaceae, or monocots are substituted for angiosperms, the angiosperm taxon connects with *Caytonia* or *Caytonia* plus glossopterids. This is also true of most trees found when Nymphaeales are substituted for angiosperms (Fig. 2), although in two such trees the Bennettitales–*Pentoxylon*–Gnetales clade, Nymphaeales, and *Caytonia* form a paraphyletic group at the base of the platysperms (a stratigraphically very unparsimonious arrangement). These trees break up the anthophytes, although the Bennettitales–*Pentoxylon*–Gnetales clade is still the next-closest group to angiosperms. This result might be questioned because it implicitly assumes that *Caytonia* and/or glossopterids had angiosperm and gnetalian states for several characters that were scored as unknown because they are not preserved or not yet established in fossils, such as lignin chemistry, a tunica layer in the apical meristem, and siphonogamy (the same is also true for Bennettitales and *Pentoxylon* in trees of the sort shown in Fig. 1). One reason for the new position of these angiosperm taxa is
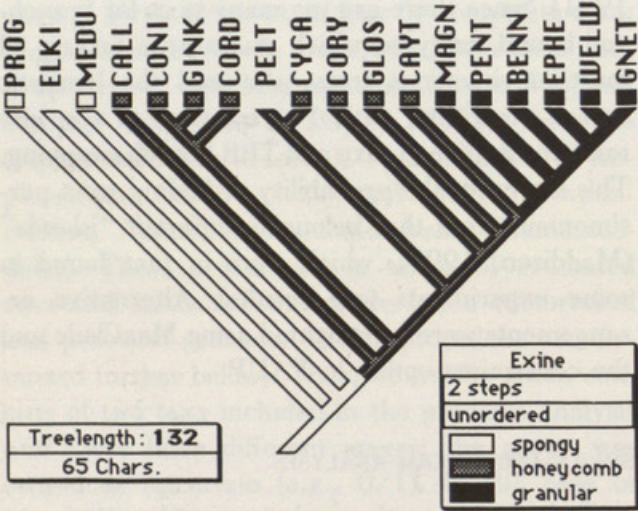
FIGURE 1. Representative most parsimonious tree (of 94) found when "core" Magnoliales (MAGN) are substituted for angiosperms as a whole in the nine-angiosperm analysis of extant and fossil seed plants. Shading of branches shows distribution of the exine structure character, which tends to link Magnoliales with other anthophytes (*Pentoxylon*, Bennettitales, Gnetales). PROG = "progymnosperms"; ELKI = *Elkinsia*; MEDU = Medullosaceae; CALL = *Callistophyton*; CONI = Coniferales; GINK = Ginkgoales; CORD = Cordaitales; PELT = *Peltaspermum*; CYCA = Cycadales; CORY = Corystospermaceae; GLOS = Glossopteridales; CAYT = *Caytonia*; PENT = *Pentoxylon*; BENN = Bennettitales; EPHE = *Ephedra*; WELW = *Welwitschia*; GNET = *Gnetum*.

FIGURE 2. Representative most parsimonious tree (of 31) found when Nymphaeales (NYMP) are substituted for angiosperms, showing distribution of the exine structure character. Generally similar trees (some with angiosperms linked with *Caytonia* alone) are also found when Winteraceae, *Austrobaileya*, eudicots, Aristolochiaceae, and monocots are substituted for angiosperms. Arrows indicate possible exine states on branches where the state is equivocal. Other abbreviations as in Figure 1.

presumably that they have columellar rather than granular exine structure (or both columellar and granular structure in the case of Nymphaeales, which were therefore scored as uncertain), so they are not as strongly "attracted" to other granular anthophytes as Magnoliales are. The exine structure character would have tended to associate angiosperms with other anthophytes if it had been ordered, with columellar implicitly derived from granular, but it does not when it is treated as unordered. Competing characters attracting angiosperms in general to *Caytonia* are reticulate venation, flat guard cells, and anatropous cupules (scored like anatropous bitegmic ovules).

As anticipated, in most of the trees found when Chloranthaceae are substituted for angiosperms (Fig. 3a), Chloranthaceae are linked directly with Gnetales, with which they share such features as opposite leaves, two-trace nodes, spicate inflorescences (scored as compound strobili), and orthotropous ovules. In these trees, the position of anthophytes is highly unstable: they may be the sister group of other platysperms or variously associated with cycads, glossopterids, *Caytonia*, or a corystosperm–glossopterid–*Caytonia* clade. The exceptions are neo-englerian trees in which Chloranthaceae are linked with *Pentoxylon* and Ben-

nettitales (Fig. 3b), which maintain the putative homologies between Chloranthaceae and Gnetales, such as compound strobili, as symplesiomorphies. Anthophytes are not broken up in any of these trees.

The most varied trees are found when Piperales are substituted for angiosperms. As when Chloranthaceae are the single angiosperm group, these include neo-englerian trees where Piperales are linked with *Pentoxylon* and Bennettitales and trees where anthophytes are the sister group of other platysperms. However, in trees of the latter sort Piperales are not linked with Gnetales but rather with *Pentoxylon* and Bennettitales (Fig. 4). This is presumably because Piperales have features like orthotropous and bitegmic ovules that allow them to be nested between Bennettitales and Gnetales, but not features like opposite leaves and two-trace nodes that unite Gnetales and Chloranthaceae. However, Piperales are linked with Gnetales in a few stratigraphically unparsimonious trees where anthophytes form a paraphyletic group at the base of platysperms. Finally, there are numerous trees in which Piperales are basal in anthophytes and anthophytes are nested among *Caytonia*, glossopterids, and corystosperms, analogous to trees of Doyle & Donoghue (1992) and trees found when Magnoliales are substituted for angiosperms (Fig. 1), and a few trees where Piperales are linked

Volume 81, Number 3
1994

Doyle et al.
Integration of Morphological and rRNA Data

425

FIGURE 3. Two representative most parsimonious trees (of 166) found when Chloranthaceae (CHLO) are substituted for angiosperms, showing distribution of the phyllotaxy (a) and strobilus (b) characters; (b) is a neo-englerian tree, with anthophytes nested among "coniferopsids" (ginkgos, conifers, cordaites). Other abbreviations as in Figure 1.
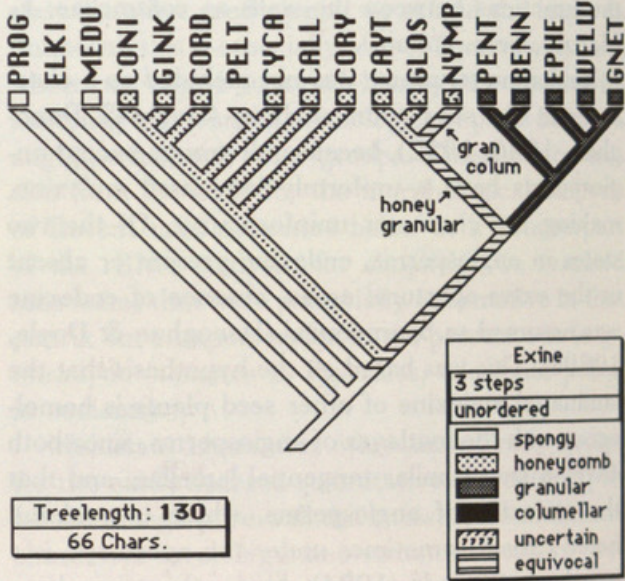
FIGURE 4. Representative most parsimonious tree (of 155) found when Piperales (PIPE) are substituted for angiosperms, showing distribution of the cupule character. Other equally parsimonious trees are more comparable to those in Figures 1, 2, and 3b. Other abbreviations as in Figure 1.

are linked with *Caytonia* and/or glossopterids imply that flowerlike reproductive structures originated independently in angiosperms and the clade consisting of Bennettitales, *Pentoxylon*, and Gnetales. Trees in which angiosperms are linked with Gnetales, and/or neo-englerian trees in which anthophytes are linked with coniferopsids and Gnetales are basal, imply that angiosperm flowers and floral parts, especially carpels containing several ovules with two integuments, were elaborated from simpler structures like those of Gnetales, or derived by aggregation of several gnetalian "flowers" (i.e., a pseudanthial interpretation).

These results bear out the concern that the position of angiosperms in previous analyses (Doyle & Donoghue, 1986, 1992) may have been incorrect because of initial assumptions about basal states in angiosperms. They also raise the possibility that angiosperms are polyphyletic, with different "angiosperm" groups related to different "gymnosperms." However, neither conclusion necessarily follows, since they depend on whether and how the various angiosperm groups link up with each other.

NINE-ANGIOSPERM ANALYSES

To assess the possibilities just raised, we included all nine angiosperm groups and analyzed the resulting matrix (35 replicates, stepwise random addition of taxa, TBR branch swapping). This analysis yielded 11 most parsimonious trees of 192 steps, which differ only in arrangements within a clade consisting of *Callistophyton*, coniferopsids, cory-

directly with *Caytonia*, thus breaking up the anthophytes. Presumably, Piperales have fewer characters that support any one of these arrangements over another.

The instability of angiosperm relationships seen in these experiments is not evident when extant taxa alone are considered, since in almost all trees Gnetales are the closest living relatives of angiosperms (the exceptions are the few trees in which Nymphaeales and Gnetales form a paraphyletic group). The variations are a function of different relationships between angiosperms and various fossil taxa. However, the different trees would have very different implications for basic states and character evolution within the angiosperms, and this illustrates the severe limitations of trees based on extant taxa alone in evaluation of evolutionary scenarios, because of the gaps between extant groups. For example, trees in which angiosperms
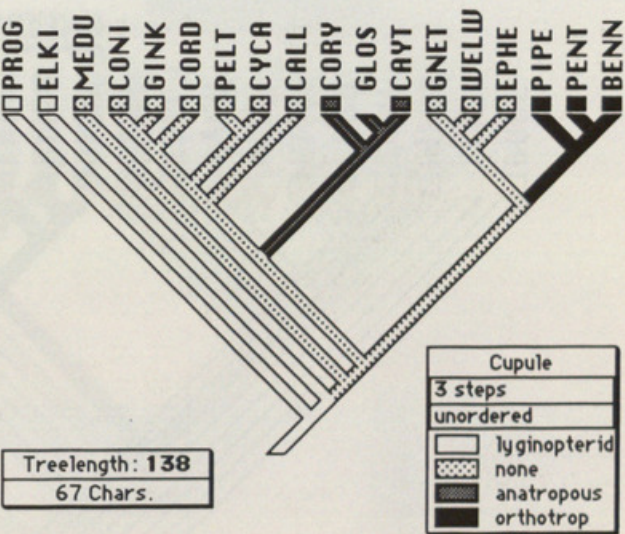
FIGURE 5. Representative most parsimonious tree (of 11) found when all nine angiosperm taxa are included in the analysis, showing distribution of the exine structure character. PIPE = Piperales; ARIS = Aristolochiaceae; NYMP = Nymphaeales; MONO = monocots; MAGN = "core" Magnoliales; WINT = Winteraceae; EUDI = eudicots (= groups with tricolpate and derived pollen); AUST = *Austrobaileya*; CHLO = Chloranthaceae; other abbreviations as in Figure 1.

stosperms, *Peltaspermum*, and cycads (e.g., Fig. 5). In these trees, angiosperms form a monophyletic group, and Gnetales are their closest living relatives. However, in contrast to trees of Doyle & Donoghue (1986), where angiosperms are the sister group of other anthophytes, angiosperms are linked with *Caytonia* and glossopterids, as when Winteraceae, *Austrobaileya*, eudicots, and some paleoherbs were substituted for angiosperms (Fig. 2). Furthermore, the arrangement within angiosperms differs from any seen previously: they split into one clade consisting of paleoherbs (including monocots) and another consisting of woody magnoliids and eudicots. As with trees where single angiosperm taxa were associated with *Caytonia* and/or glossopterids, this rearrangement is presumably influenced by treatment of the exine structure and endexine characters as unordered. This change in character analysis weakens the tendency of angiosperms to associate with the granular Bennettitales–*Pentoxylon*–Gnetales clade, with granular Magnoliales attached between the latter and columellar angiosperms.

Although these results cast doubt on previous inferences regarding the origin of angiosperms, the change from the previous situation is less radical than it seems. Essentially the same alternative relationships recognized as almost equally parsimo-

nious by Doyle & Donoghue (1986, 1992) and Donoghue & Doyle (1989a) are seen among one step less parsimonious ("one-off") trees (193 steps) (Fig. 6). Some of these trees show the sorts of relationships that were most parsimonious in the previous analyses, in which anthophytes are associated with *Caytonia* and glossopterids, angiosperms are the sister group of other anthophytes, and Magnoliales are basal in angiosperms (Fig. 6a). As in the previous analyses, other one-off trees are of the neo-englerian type (Fig. 6b), in which anthophytes are linked with coniferopsids, and Gnetales, with linear leaves and simple sporophylls, are basal and relatively plesiomorphic in anthophytes. In other 193-step trees, Nymphaeales are basal in angiosperms, as inferred from rRNA data (Hamby & Zimmer, 1992); in fact, some of these have the same arrangement of paleoherb taxa found in the rRNA analyses presented below, with Piperales next above Nymphaeales and monocots linked with Aristolochiaceae (Fig. 6c). Given the minimal differences in parsimony, it would be unwarranted to conclude that the tree in Figure 5 should now be strongly preferred over those in Donoghue & Doyle (1989a) or Doyle & Donoghue (1992). Although the shift away from a basal position of Magnoliales may represent real progress, reflecting as it does the removal of previous biases in exine characters,
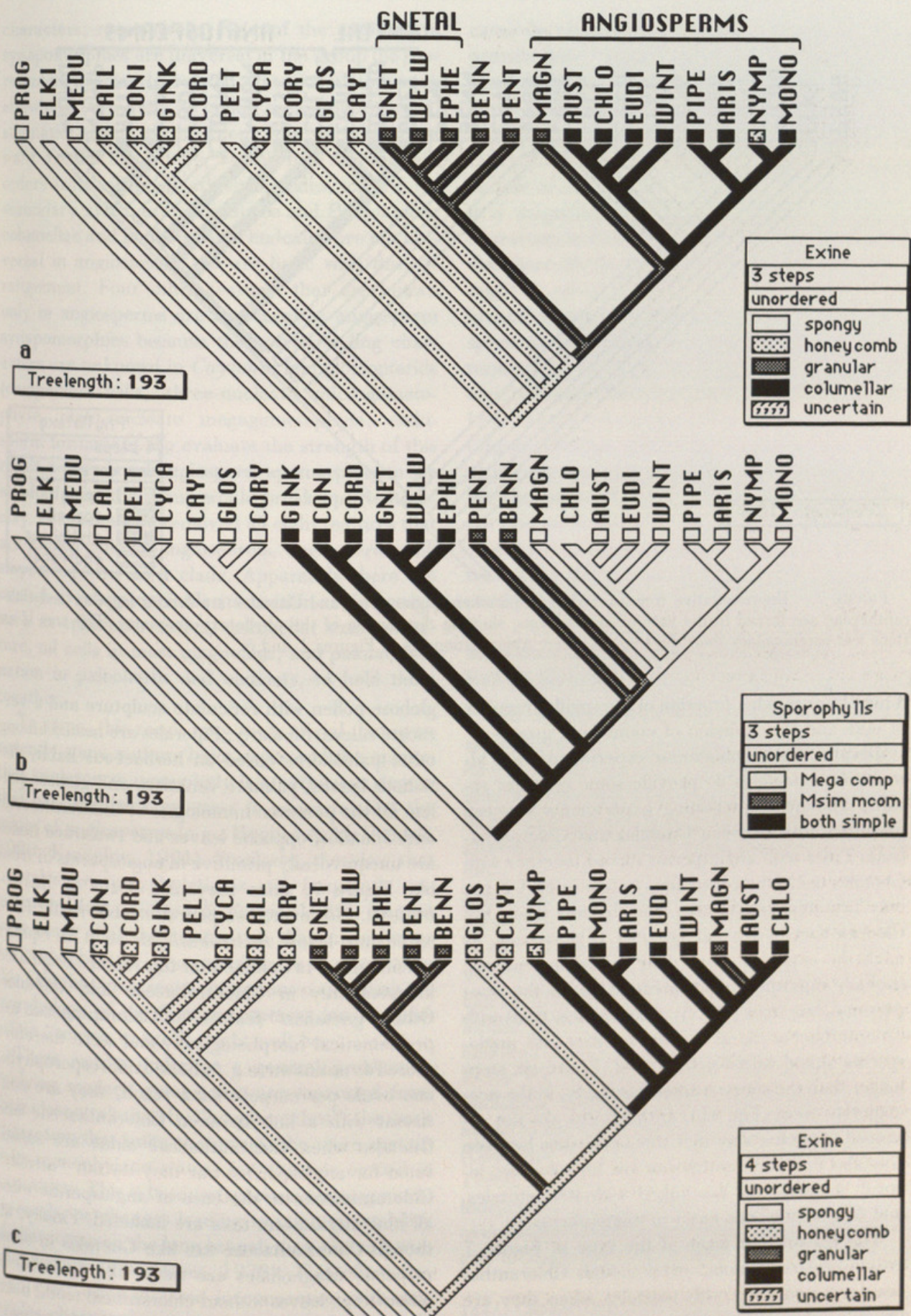
FIGURE 6. Representative one step less parsimonious (one-off) trees found with the nine-angiosperm data set: (a) with angiosperms the sister group of other anthophytes and Magnoliales basal in angiosperms, as in Doyle & Donoghue (1986, 1992) and Donoghue & Doyle (1989a), showing distribution of the exine structure character; (b) neo-englerian tree, with anthophytes nested among "coniferopsids," showing distribution of the sporophyll character; (c) with angiosperms linked with *Caytonia* and glossopterids and Nymphaeales basal in angiosperms, as in trees based on rRNA data, showing distribution of the exine structure character. Abbreviations as in Figures 1 and 5.

GNETAL   ANGIOSPERMS

PROG ELKI MEDU CONI CORD GINK PELT CYCA CALL CORY GLOS CAYT BENN PENT GNET WELW EPHE CHLO AUST MAGN WINT EUDI PIPE ARIS NYMP MONO

Phyllotaxy
4 steps
unordered
☐ spiral
▨ opposite
■ distichous
▤ equivocal

Treelength: 198
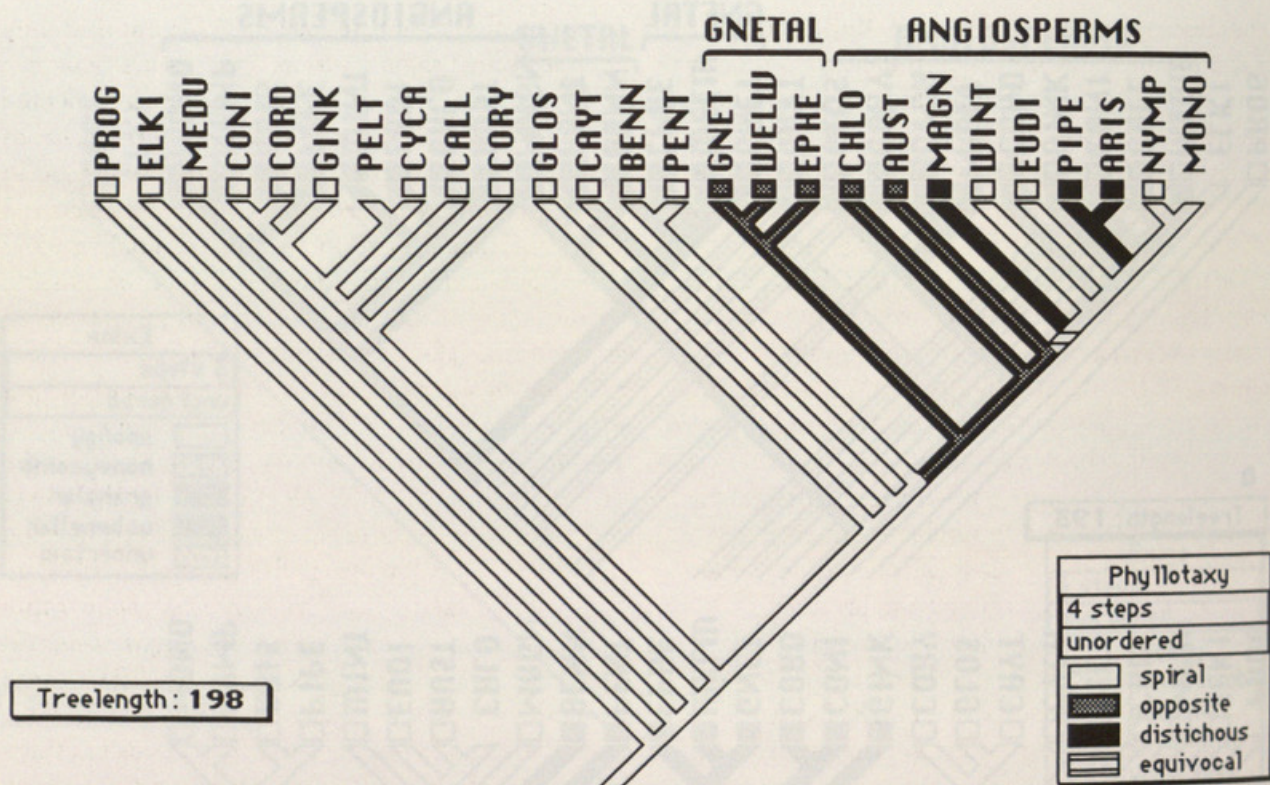
FIGURE 7. Representative tree (of 1675) found when angiosperms and Gnetales are forced together and Chloranthaceae are forced to the base of angiosperms, showing distribution of the phyllotaxy character. This tree is six steps less parsimonious than the shortest trees. Abbreviations as in Figures 1 and 5.

it may be as much a function of the smaller number of taxa and the omission of connecting groups.

Despite these ambiguities, experiments with alternative topologies do provide some stronger results. The most interesting experiments concern proposed links between Gnetales and Chloranthaceae. Trees with angiosperms forced together with Gnetales using the constraints option in PAUP are only two steps longer than the shortest trees, but Chloranthaceae are not basal in angiosperms, as might be expected from their gnetalian features; instead, angiosperms are arranged as in the most parsimonious trees (Fig. 5). The shortest trees with Chloranthaceae basal in angiosperms and angiosperms linked with Gnetales (Fig. 7) are six steps longer than the shortest trees. Similarly, in the neo-englerian trees (Fig. 6b), anthophytes are not arranged in such a way that the similarities between Gnetales and Chloranthaceae are homologous; instead, angiosperms are linked with Bennettitales, and Magnoliales are basal in angiosperms.

The reason that trees of the type in Figure 7 are so unparsimonious, even though Chloranthaceae are associated with Gnetales when they are the only angiosperms in the analysis, is presumably that Chloranthaceae are "screened off" from having an influence on the position of angiosperms by being linked with *Austrobaileya* and other groups, based on such features as laterocytic stomates and

globose pollen with reticulate sculpture and a verrucate sulcus. Similarly, Piperales are nested among other paleoherbs based on herbaceous habit, distichous leaves, palmate venation, and globose pollen. Of the potential homologies of Chloranthaceae and Gnetales, opposite leaves and two-trace nodes are unequivocally primitive in angiosperms in trees like Figure 7, since they also occur in *Austrobaileya*. However, compound strobili would have to be lost below *Austrobaileya*, so it is equally parsimonious to assume that this feature originated independently in Chloranthaceae and Gnetales. Other "gnetalian" features of Chloranthaceae are problematical morphologically and were therefore scored as unknown (e.g., whorled microsporophylls, one ovule per carpel). As a result, they are consistent with a link between Chloranthaceae and Gnetales when Chloranthaceae alone are substituted for angiosperms, but they do not "attract" Chloranthaceae to the base of angiosperms when all nine angiosperm taxa are included. Finally, although Chloranthaceae are like Gnetales in some respects, Magnoliales are more like Gnetales in others (e.g., leaves without chloranthoid teeth; boat-shaped pollen; tectate, granular exine structure).

The strongest conclusion of these analyses is that both angiosperms and Gnetales are monophyletic groups. In the tree shown in Figure 5, they are unambiguously supported by nine and eight

characters, respectively. Four of the angiosperm synapomorphies are universal in the group (or represented by clearly related states) and unknown elsewhere (two pairs of pollen sacs, endothecium, stigmatic pollen germination, loss of megaspore wall). Of the five others, vessels and several vein orders arose independently in Gnetales, scalariform secondary xylem in Bennettitales and *Pentoxylon*; columellae and nonlaminated endexine are not universal in angiosperms but are basic with this arrangement. Four other features that are known only in angiosperms are equivocal as angiosperm synapomorphies because the corresponding characters are unknown in *Caytonia* and glossopterids (companion cells, three-nucleate microgametophyte, eight-nucleate megagametophyte, endosperm formation). To evaluate the strength of this result, we removed angiosperm apomorphies to see at what point the group would break up. Remarkably, even when we removed all eight features that are known only in angiosperms, angiosperms still stayed together as a clade. Apparently there are enough overlapping similarities within angiosperms, such as trilacunar nodes, columellar exine structure, oil cells in most magnoliids, and palmate venation in paleoherbs and eudicots, to hold them together.

To some, this result may seem trivial, but even recently some authors have expressed the opinion that angiosperm monophyly is a pernicious dogma that has held back progress in understanding the origin of angiosperms (e.g., Hughes & McDougall, 1990; Krassilov, 1991). Similarly, the view that Gnetales are polyphyletic is still frequently encountered (e.g., Gifford & Foster, 1989). Our results imply that the assumption that angiosperms are monophyletic is not an obstacle to progress in this field, but incorrect assumptions concerning the morphology of the first angiosperms (an inappropriate "search image") might well be.

These results underline a general problem in studying modern groups that are separated from their closest relatives by "long branches": the more characters that unite the group, the more certain is its monophyletic status, but the less certain is its position. This is because spurious convergences, reversals, or changes leading to uninterpretability of characters on the long branch may obscure true relationships (Felsenstein, 1978). This effect has been most often stressed in connection with molecular characters, where there are only three alternative states to which a base at any position can change, but it may also apply to morphological characters (cf. Wake, 1991). For the same reasons, rooting of the group may be ambiguous, be-

cause the closest outgroups are so distant that they provide little "signal" as to which subgroups are basal (Wheeler, 1990). This problem is reflected in the large number of angiosperm characters that could not be polarized by outgroup comparison in Donoghue & Doyle (1989a), or could not be scored outside of angiosperms in the present analysis, and it is magnified by the large number of missing characters in fossils. As discussed in greater detail elsewhere (Doyle & Donoghue, 1993), the problem might be solved by discovery of fossils on the long branch leading to angiosperms (i.e., non-angiospermous angiophytes, or stem angiophytes, in the terminology of Doyle & Donoghue, 1993). However, although there are a few candidates, like the Late Triassic Crinopolles pollen group described by Cornet (1989a), which has angiospermlike reticulate sculpture and columellae but a gymnospermlike endexine (cf. Doyle & Hotton, 1991), there are still no fossils with angiosperm states in some characters and more plesiomorphic states in others that can be placed with certainty on the angiosperm stem lineage.

MORPHOLOGICAL AND rRNA ANALYSES OF EXTANT GROUPS

The relationship of these results to those obtained from rRNA sequences is addressed more directly by our analyses of morphological and rRNA data from the same taxa. Since position and rooting of the angiosperms are ambiguous with the morphological data set just presented, and since still other results have been obtained with other interpretations of angiosperm characters (Loconte & Stevenson, 1991; Taylor & Hickey, 1992), one of the motivations for this study was to determine what if any additional insights into these questions can be extracted from rRNA data. Although molecular data have the disadvantage of being available only from living groups (except for a few recent fossils: e.g., Golenberg et al., 1990), whereas morphological data may exist for key fossil taxa that attach to the long branches separating extant groups (Donoghue & Doyle, 1989b; Donoghue et al., 1989), molecular data have the advantage of being independent of the seemingly endless controversies over interpretation of the morphological homologies of angiosperm structures.

TAXA, CHARACTERS, AND ANALYSES

The starting point for our rRNA analyses was a successor to the 60-taxon data set of Hamby & Zimmer (1992), enlarged to include 71 taxa. Complete sequences are available from GenBank (Ac-

cession Nos. M81965-M82800) and the NMNH Gopher Server under "LMS." Our morphological data set is derived from the nine-angiosperm analysis described above, modified by inclusion of somewhat different angiosperm taxa and removal of fossil groups (Appendix).

We encountered a variety of problems in obtaining comparable taxa for the morphological and rRNA analyses and combining the two data sets. This required many compromises and approximations. All of these involve a certain risk of error, but we feel they are unavoidable if progress is to be made at this point (cf. Maddison & Maddison, 1992). The important thing is to spell out the assumptions involved so that they can be scrutinized and tested in future analyses.

In reducing the morphological and molecular data sets to a comparable set of taxa, our goal was to include an adequate sampling of critical taxa (as judged from results of previous studies) while keeping the number of taxa small enough for the more time-consuming analyses. We had to omit taxa for which rRNA data are still lacking; this is often unfortunate, since current evidence suggests that some such groups constitute important links. Examples are *Lactoris*, which recent authors have linked with Piperales (Carlquist, 1990) or Aristolochiaceae (Donoghue & Doyle, 1989a; Qiu et al., 1993), and *Austrobaileya*, Trimeniaceae, and *Amborella*, which may help to tie together Laurales and strengthen the position of Chloranthaceae in this group. We did not include *Ceratophyllum* (also omitted by Donoghue & Doyle, 1989a), because its position was highly unstable in the complete rRNA analyses (Hamby & Zimmer, 1992) and so many of its morphological characters are difficult to interpret. Although *Ceratophyllum* occupies a key position at the base of the angiosperms in *rbc*L analyses of seed plants as a whole (Les et al., 1991; Chase et al., 1993), its·position is unstable in unrooted *rbc*L analyses of angiosperms alone (Qiu et al., 1993), suggesting that its basal position may be an artifact of long branch attraction (cf. Donoghue, 1994).

Similarly, we did not include any non-seed plants as outgroups, since outgroups in the original data sets were different. In the morphological analyses of extant seed plants (Doyle & Donoghue, 1987, 1992), we assumed that ferns and *Equisetum* are closer to extant seed plants and lycopsids more distant, but the outgroups included in the rRNA analysis were *Equisetum* and *Psilotum*. In addition, all extant outgroups are very distant from seed plants; the really appropriate outgroups are Devonian "progymnosperms" and Carboniferous "seed

ferns." There is therefore reason to fear that any rooting obtained from extant outgroups might be an artifact of spurious long branch attraction (Wheeler, 1990; Maddison et al., 1992). The resulting trees are therefore unrooted networks, with the root arbitrarily placed along the branch leading to cycads. However, it will be seen that this procedure does provide instructive contrasts, because trees derived from the morphological and rRNA data sets are topologically different.

In combining original taxa into larger groups, we generally accepted clades that appeared in both the morphological analysis of Donoghue & Doyle (1989a) and the consensus of most parsimonious and one-off rRNA trees, with a few exceptions motivated by a desire to test current hypotheses. Thus we retained Piperaceae and Saururaceae as separate taxa, even though they were strongly linked in the trees of Donoghue & Doyle (1989a) and associated in some one-off rRNA trees, because they are separated in the tree of Taylor & Hickey (1992). Conversely, we combined *Hedycarya* (Monimiaceae) and *Persea* (Lauraceae) as "core Laurales" in the rRNA analysis, even though they are not associated in all most parsimonious rRNA trees, because their relationship is strongly supported and uncontroversial on morphological grounds.

Despite our efforts, taxa in the morphological and rRNA data sets are not perfectly comparable, except perhaps when they are monotypic (i.e., *Ginkgo*, *Welwitschia*). In general, clades in the morphological data set are represented in the molecular data set by a few species that show only part of the variation in the whole clade (e.g., Magnoliales by *Asimina*, *Magnolia*, and *Liriodendron*; core Laurales by *Persea* and *Hedycarya*), or by single "exemplar" species (Chloranthaceae by *Chloranthus*; Ranunculidae by *Ranunculus*; Trochodendrales by *Trochodendron*). One solution would be to rescore taxa in the morphological data set to correspond exactly to the species in the rRNA analysis. However, we opted instead to score clades as in the earlier morphological analysis and to assume that they are adequately represented in the rRNA analysis by the exemplars, since it seemed unlikely that rescoring them would lead to significantly different results in the new morphological analysis. For example, *Ranunculus* and *Trochodendron* have most of the characters that helped link Ranunculidae, Trochodendrales, and other eudicots in the previous analysis (tricolpate pollen, sculptured aperture membranes, chloranthoid teeth, lack of oil cells), plus one that would be ambiguous if Ranunculidae as a whole were considered, be-

Volume 81, Number 3
1994

Doyle et al.
Integration of Morphological and rRNA Data

431
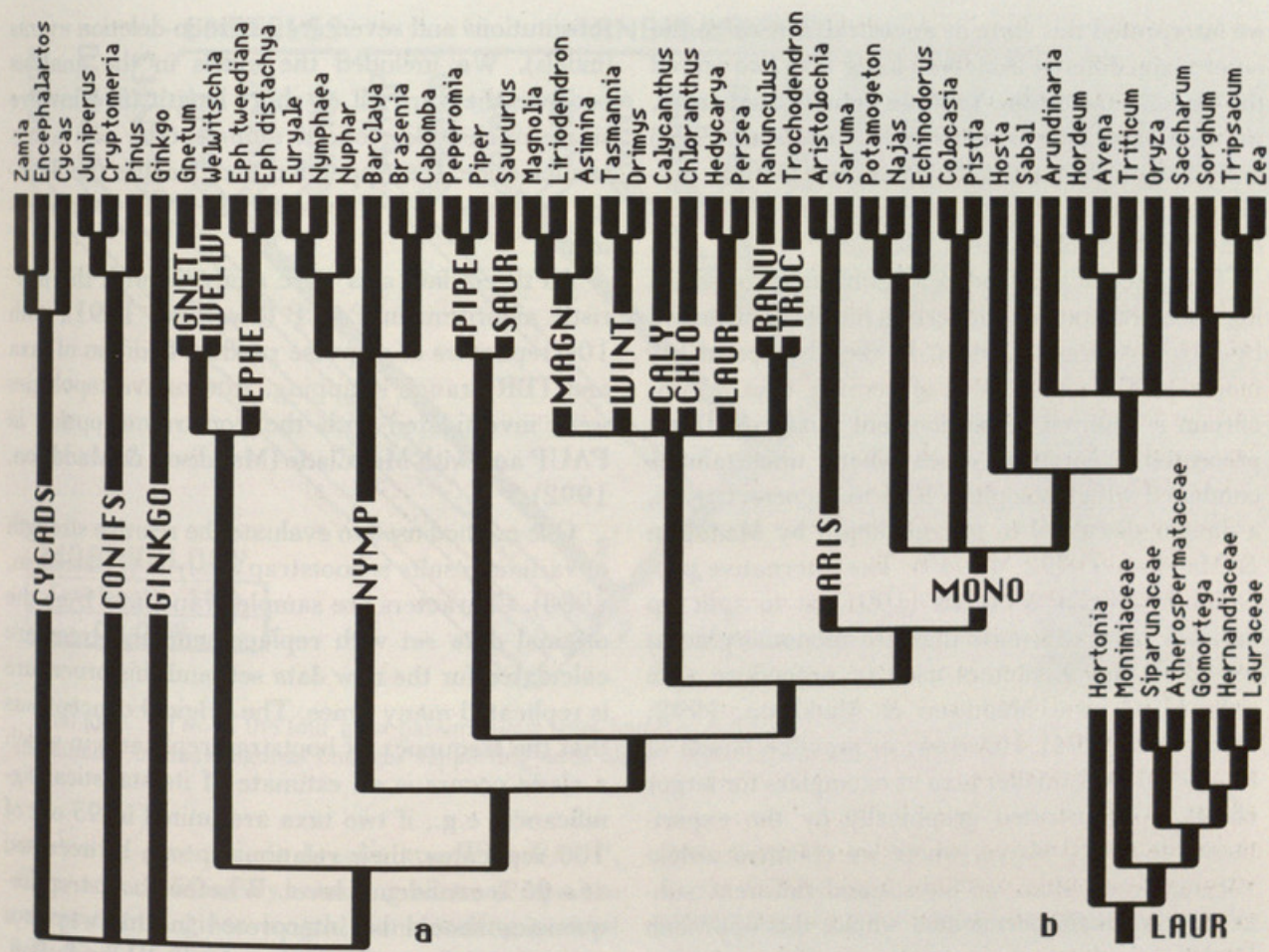
FIGURE 8. (a) Relationships assumed in reducing taxa in the original 71-taxon rRNA data set to the 18 taxa used in the present rRNA anlayses (see text for discussion). CYCADS = Cycadales; CONIFS = Coniferales; GINKGO = *Ginkgo*; GNET = *Gnetum*; WELW = *Welwitschia*; EPHE = *Ephedra*; NYMP = Nymphaeales; PIPE = Piperaceae; SAUR = Saururaceae; MAGN = "core" Magnoliales; WINT = Winteraceae; CALY = Calycanthaceae; CHLO = Chloranthaceae; LAUR = "core" Laurales; RANU = Ranunculidae; TROC = Trochodendrales; ARIS = Aristolochiaceae; MONO = monocots. (b) Relationships within core Laurales assumed in the morphological analyses.

cause of variation within that group (stamens with well-differentiated filaments).

In scoring taxa that vary for characters included in the matrix, our goal was to obtain the best estimate of ancestral states for the whole taxon. In the rRNA data set, we accepted internal relationships that were consistent in the original 71-taxon rRNA analysis and in morphological analyses by ourselves and others (Fig. 8a). For example, considering taxa included in the analysis, we assumed that *Cycas* is the sister group of *Zamia* and *Encephalartos*, and *Pinus* is the sister group of *Juniperus* and *Cryptomeria*, since these relations are found in the whole rRNA analysis and in the morphological analyses of Hart (1987), Crane (1988), and Stevenson (1990). In contrast, in Nymphaeales the rRNA data indicate that *Barclaya* is the sister group of the remaining taxa, but morphological data (Ito, 1987) imply that Cabombaceae occupy this position; therefore we treated the three groups as a trichotomy. In estimating

basic states for monocots, we accepted relationships among grasses that are consistent with rRNA and morphological analyses (Kellogg & Campbell, 1987), treated *Hosta*, *Sabal*, and grasses as a trichotomy, and treated alismids, aroids, and the *Hosta*–*Sabal*-grass clade as a trichotomy. In the morphological data set, we inferred states for core Magnoliales based on the assumption that *Degeneria* is the sister group of Myristicaceae, Annonaceae, and Magnoliaceae (cf. above); in core Laurales, we assumed relationships shown in Figure 8b, derived from an unpublished analysis of Laurales used for the same purpose by Donoghue & Doyle (1989a).

The use of parsimony in optimization of ancestral states on trees is discussed by Swofford & Maddison (1987). As in the nine-angiosperm analysis, when clades consisted of two taxa that varied at a given site, we coded them as uncertain. When there were three taxa, and the "outer" taxon and one of the two "inner" taxa had the same state,

we interpreted this state as ancestral; but when the outer taxon differed from both inner taxa, we scored the clade as uncertain. In more complicated cases, we used MacClade (Maddison & Maddison, 1992) to find the most parsimonious ancestral state, taking into account all possible resolutions of trichotomies and polychotomies.

The practice of including variable taxa and scoring ancestral states as uncertain has been criticized by Nixon & Davis (1991). In the absence of homoplasy, the main effect of scoring taxa as uncertain is lowered resolution, but Nixon & Davis presented theoretical cases where uncertainties combined with homoplasy lead to incorrect trees, a danger discussed in greater depth by Maddison & Maddison (1992: 47–49). The alternative proposed by Nixon & Davis (1991) is to split up variable taxa into units that are monomorphic in terms of the characters used (a procedure with risks of its own: Maddison & Maddison, 1992; Donoghue, 1994). However, in practice Nixon et al. (1994) used smaller taxa as exemplars for larger clades. As illustrated graphically by the experiments described above, where we obtained widely varying trees when we substituted different subgroups for angiosperms as a whole, this approach is ridden with as many implicit assumptions and potentials for error (e.g., that convergences and reversals in the exemplar will not affect its position) as attempts to reconstruct basal states (Donoghue, 1994). Splitting up variable taxa into all potentially relevant monophyletic taxa might be the ideal solution, but if carried to its logical conclusion this would quickly lead to computational paralysis. In the meantime, we prefer to make explicit assumptions about basal states of the sort described, while emphasizing that these assumptions can and should be tested in the future.

Our morphological data set is presented in Table 2 (Appendix). Characters are the same as those in the nine-angiosperm analysis; changes in the number and/or definition of states as a result of removal of fossils are indicated in the character definitions. Deletion of characters that became uninformative after removal of fossil taxa left a total of 69 characters. This data set therefore parallels the extant seed plant matrix of Doyle & Donoghue (1992).

The rRNA data set is presented in Table 3 (Appendix). Characters from the 18S subunit are keyed to the corresponding positions in soy, characters from the 26S subunit to positions in rice. As a result of reducing the number of taxa from 71 to 18 and eliminating characters that became uninformative, the number of characters was reduced from 411 to 174, of which 167 are base

substitutions and seven are insertion-deletion event (indels). We included the indels in the analysi because they are all of short length (five involv single nucleotides, one a dinucleotide, one a tet ranucleotide: see Appendix) and because the flank ing sequences leave no ambiguity as to their align ment.

All three data sets were analyzed with the heu ristic algorithm in PAUP (Swofford, 1991), with 100 replicates of stepwise random addition of taxa and TBR branch swapping. Alternative topologie were investigated with the constraints option in PAUP and with MacClade (Maddison & Maddison 1992).

One method used to evaluate the relative strength of various results is bootstrap analysis (Felsenstein 1985). Characters are sampled randomly from the original data set with replacement and trees are calculated for the new data set, and this procedure is replicated many times. The original concept was that the frequency of bootstrap replicates in which a clade occurs is an estimate of its statistical sig nificance: e.g., if two taxa are united in 95 out of 100 replicates, their relationship can be accepted at a 95% confidence level. Whether bootstrap fre quencies should be interpreted in this way is a matter of debate (Carpenter, 1992; Hillis & Bull, 1993; Felsenstein & Kishino, 1993), but the crit icisms made do not call into question the value of bootstrap analysis as a means of evaluating the relative robustness of clades. Actually, simulation experiments suggest that the bootstrap errs on the side of being conservative; under many circum stances, clades seen at bootstrap levels lower than the conventional limit of 95% are more accurate than the bootstrap numbers would imply (Hillis & Bull, 1993). In addition, bootstrap analysis may be useful in uncovering possible alternative rela tionships, as seen in the dot-plots of "frequency of occurrence" of groups provided by PAUP. If a link seen in the most parsimonious trees is actually due to convergence, there should also be minority characters that reflect the true relationship, and these should be sampled and amplified in a rela tively high frequency of bootstrap replicates. For each data set, we did 1000 bootstrap replicates. To increase the probability of finding most parsi monious trees in each replicate, we did 10 heuristic analyses with stepwise random addition of taxa and TBR branch swapping.

The other method we employed to evaluate ro bustness is decay analysis (Bremer, 1988; Dono ghue et al., 1992). This method determines how much longer trees have to be—how much parsi mony has to be relaxed—before trees are found

Volume 81, Number 3
1994

Doyle et al.
Integration of Morphological and rRNA Data

433

FIGURE 9. Two of the four most parsimonious trees found in analysis of the extant morphological data set, showing the number of unambiguous changes supporting each clade. Abbreviations as in Figure 8.

in which a given clade breaks up. With PAUP, this is accomplished by retaining all trees equal to or less than a given length, constructing a strict consensus of the resulting trees, and observing which clades remain in the consensus. Ideally, this is done with a branch-and-bound algorithm, which guarantees finding all trees of a given length; because this was not possible with 18 taxa, we used a heuristic search with 100 replicates of stepwise random addition and TBR branch swapping.

RESULTS OF MORPHOLOGICAL ANALYSES

Analysis of the morphological data set yields four most parsimonious trees of 150 steps, two illustrated in Figure 9. The variations concern whether Magnoliales are the sister group of the three lauralian taxa or nested among them. Figure 9 also shows the number of characters that unequivocally support clades; angiosperms are united by at least fifteen characters, Gnetales by eight, and the two groups by seven. This gives some indication of the level of support, but it is potentially misleading because it says nothing about the distribution of homoplasy—whether these changes are unique or duplicated elsewhere on the tree. This problem is addressed by the bootstrap and decay analyses. The consistency index is 0.58, about average for this number of taxa (Sanderson & Donoghue, 1989); the retention index is 0.73. These trees show the arrangement of non-angiospermous groups found in analyses of extant

taxa alone by Loconte & Stevenson (1990) and Doyle & Donoghue (1992). Doyle & Donoghue (1992) argued that this result may be an artifact of omitting fossils: when fossils are included in the analysis, this arrangement of living groups is only one of several that are equally parsimonious. In addition, several of the characters that apparently unite angiosperms are not unique to the group when fossil taxa are considered, since they also occur in *Caytonia* (flat stomata, anatropous cupules), Bennettitales (scalariform metaxylem), or both (pinnate sporophyll organization, integument free from nucellus). The whorled microsporophylls and tubular micropyle that apparently unite Gnetales are shared with Bennettitales. Of the characters uniting angiosperms and Gnetales, opposite phyllotaxy and vessels are not synapomorphies if Bennettitales and *Pentoxylon* are interpolated between the two groups.

Contrary to Donoghue & Doyle (1989a), the four most parsimonious trees root angiosperms among paleoherbs rather than Magnoliales, with Nymphaeales plus monocots as the sister group of other angiosperms. Plesiomorphic features of Nymphaeales and monocots include boat-shaped pollen and lack of oil cells (although the latter may not be valid if *Acorus*, which has oil cells, is basal in monocots, as inferred from *rbc*L data: Duvall et al., 1993). However, the conflict with the Donoghue & Doyle results is less severe than it appears: if angiosperms are rerooted on the branch to Magnoliales, only one step longer trees are obtained
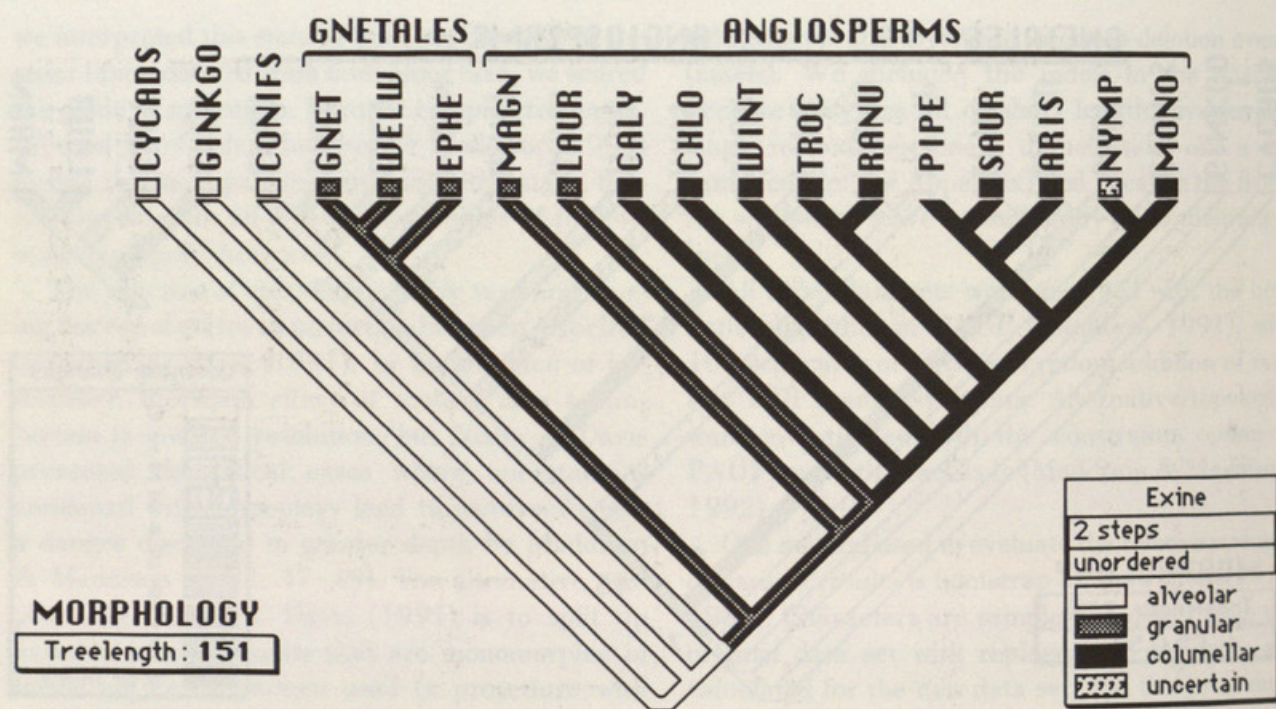
FIGURE 10. A one step less parsimonious tree based on the morphological data set with Magnoliales basal in angiosperms, showing distribution of the exine structure character. Abbreviations as in Figure 8.

that are almost entirely consistent with trees in Donoghue & Doyle (1989a), except that Laurales are paraphyletic rather than monophyletic (Fig. 10), and only one additional step is required to make Laurales a clade. The association of Magnoliales with core Laurales, either as a clade or a basal paraphyletic group, is due to possession of PI type sieve-tube plastids, granular exine structure, and a continuous tectum, all features that appear to be convergent when more taxa are included (Donoghue & Doyle, 1989a). Although monocots are associated with Nymphaeales rather than Aristolochiaceae, as in the rRNA analyses presented below, trees in which monocots are linked with Aristolochiaceae are only one step less parsimonious.

This shift in rooting may be partly a function of the smaller sampling of taxa, but as we argued above in connection with the nine-angiosperm analysis, it is also a result of treating the exine structure and endexine characters as unordered rather than implicitly ordered. Although the new morphological results are more consistent with the rRNA results, the situation does not contradict the view that the previous conflict between trees derived from the two sorts of data was a function of lower resolution of the morphological evidence—actually, it strengthens this view. Previously, morphological data favored a magnolialian rooting, but only weakly; now both data sets favor a paleoherb rooting, but the morphological data do so weakly. The fact

that this shift followed from a rather subtle change in interpretation of two characters underlines the ambiguity of the morphological data.

Figure 11 summarizes the bootstrap and decay analyses of the morphological data. The strongest results of the bootstrap analysis are the monophyly of angiosperms, seen in 100% (more precisely 99.9%) of the bootstrap replicates, and, within the angiosperms, the association of Saururaceae and Piperaceae (99%). This contradicts the tree of Taylor & Hickey (1992), in which Saururaceae and Piperaceae are distantly separated. Next strongest is the link between angiosperms and Gnetales (95%). The monophyly of Gnetales is somewhat weaker (92%); examination of lower-frequency groupings indicates that this is because angiosperms are nested within Gnetales in some replicates, presumably due to features that they share with *Welwitschia* and *Gnetum* (reticulate venation, paracytic stomata, cellular embryogeny). Presumably, features of this sort are responsible for the position of angiosperms among Gnetales in some of the trees of Nixon et al. (1994). The Loconte & Stevenson (1990) arrangement of cycads, *Ginkgo*, and conifers occurs at a frequency of 79%. Except for Piperales, groupings within angiosperms that were seen in Donoghue & Doyle (1989a) appear at much lower frequencies, the strongest being the eudicots (43%).

The bootstrap results also bear on the rooting problem, although only indirectly. Insights come
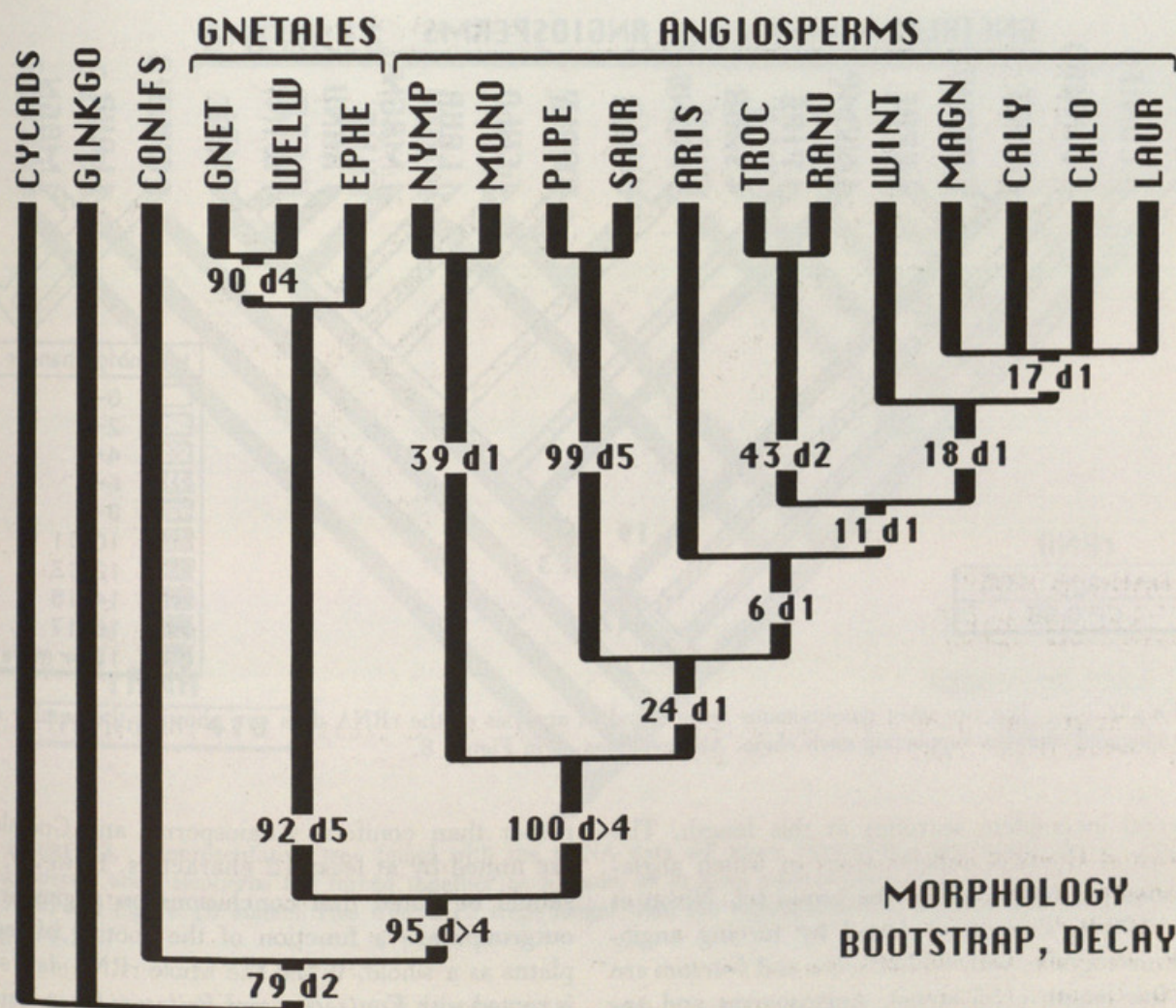
FIGURE 11. Results of bootstrap and decay analyses of the morphological data set. The first number indicates the percentage of bootstrap replicates in which each clade is found; the second number (d1, etc.) indicates how many steps longer trees must be before some are found in which the clade no longer occurs (decays). The search of five-off trees was incomplete; clades that had not decayed in that search are labeled d > 4, because we cannot rule out the possibility that they do decay in "islands" of five-off trees that were not discovered. Abbreviations as in Figure 8.

from examining the frequency of angiosperm clades containing all but one or two taxa, which imply that the taxa not included are basal, without specifying their exact arrangement. As expected, groupings implying that Nymphaeales and monocots are basal are most frequent, but at only 24%, followed by Nymphaeales alone at 17%, and Magnoliales at 11%. These results again illustrate the instability of the root, while favoring the same alternatives inferred from the primary analysis.

The decay numbers in Figure 11 refer to the number of steps that must be added (how many steps longer trees must be) before the group in question no longer forms a clade; e.g., "d2" indicates that it is present in all one-off trees but "decays" in some two-off trees (152 steps). In the consensus of one-off trees, the only clades remaining within angiosperms are Piperales (Piperaceae and Saururaceae) and eudicots (Ranunculidae and Trochodendrales). This result reflects the unstable

position of the root; the one-off trees include not only those rooted among paleoherbs and next to Magnoliales, but also some rooted next to eudicots, next to paleoherbs as a group (as in the nine-angiosperm analysis: Fig. 5), and next to woody magnoliids as a group. It also illustrates the fact that strict consensus trees may underestimate the amount of structure in the data: if a single taxon (or the root) "jumps" from one clade to another, the intervening groups collapse to a polychotomy, even though their other members maintain the same arrangement. Eudicots decay in two-off trees, leaving only Piperales. The arrangement of cycads, *Ginkgo*, and conifers decays in two steps, and the relationship between *Welwitschia* and *Gnetum* in four. However, angiosperms, Gnetales, the relationship between them, and Piperales are still intact in the four-off trees. Five-off trees were not searched exhaustively because of time and memory limitations, but both Piperales and Gnetales decayed in
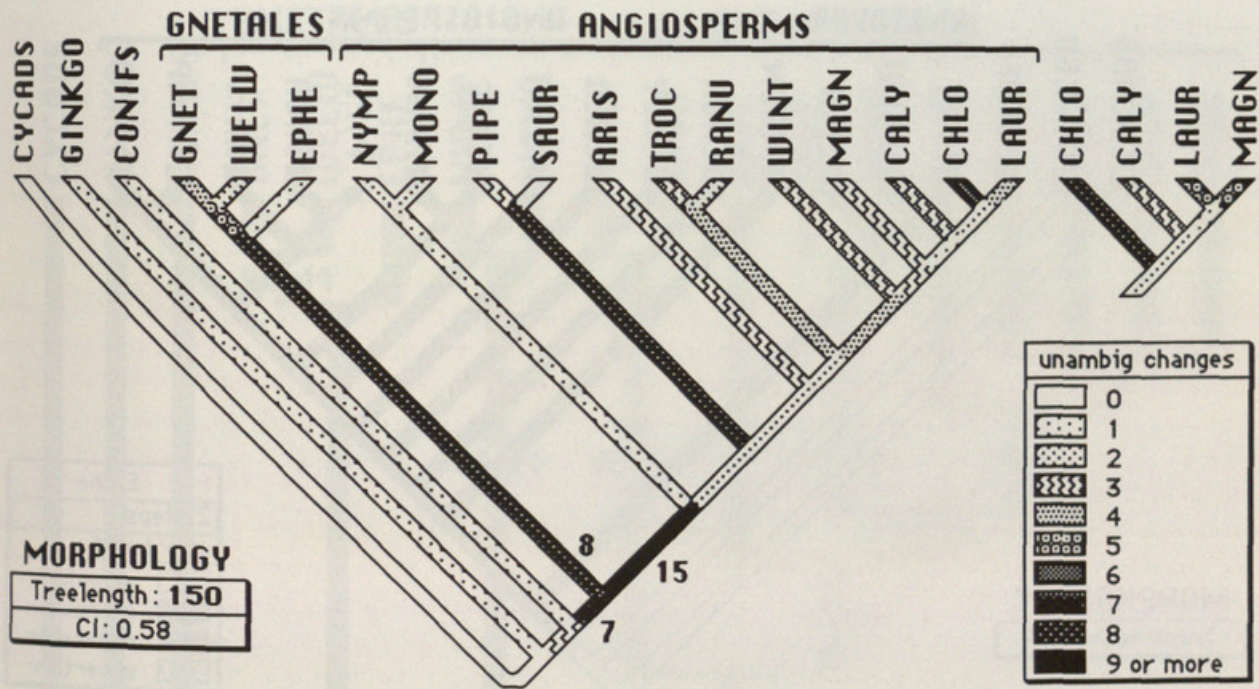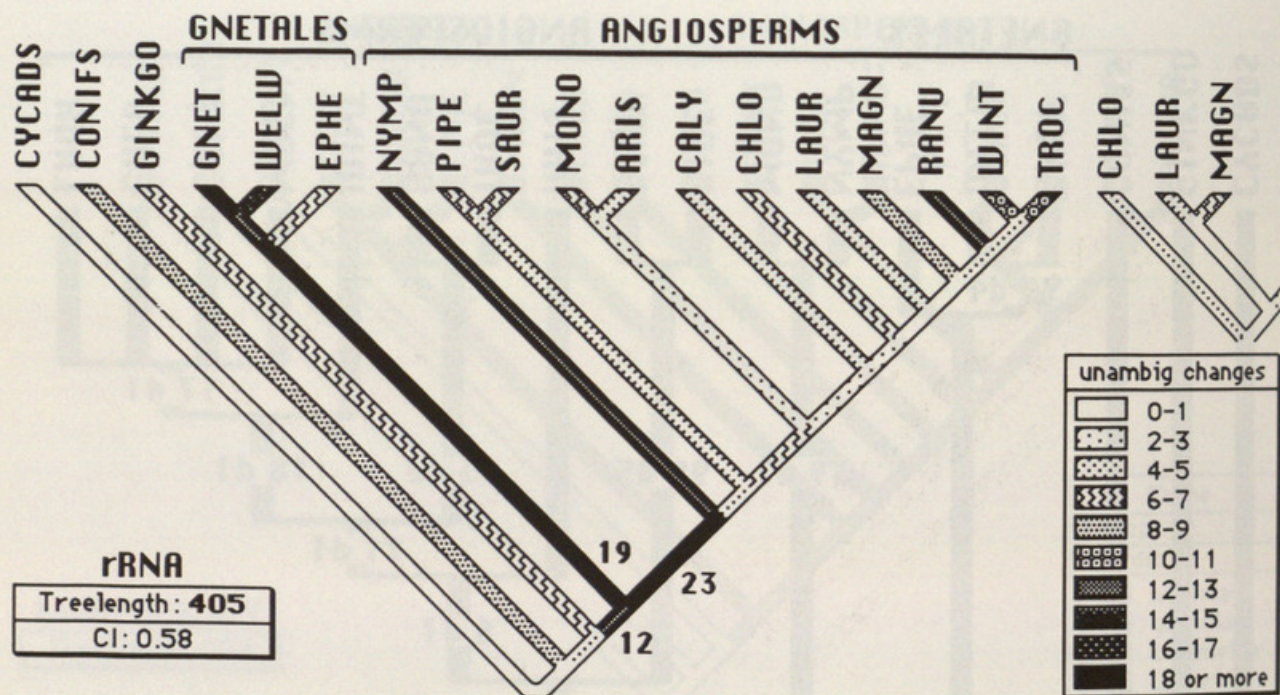
FIGURE 12. The two most parsimonious trees found in analysis of the rRNA data set, showing the number of unambiguous changes supporting each clade. Abbreviations as in Figure 8.

several incomplete searches at this length. The decay of Gnetales reflects trees in which angiosperms are nested within the group (cf. Nixon et al., 1994), since trees found by forcing angiosperms together with *Welwitschia* and *Gnetum* are of this length (155 steps). Angiosperms and anthophytes remain as clades in all five-off trees found, but because we cannot be certain that they do not decay in "islands" of trees that were not searched, they are labeled d > 4. This decay order closely parallels the relative strength of clades inferred from bootstrap analysis.

RESULTS OF rRNA ANALYSES

Analysis of the rRNA data yields two most parsimonious trees of 405 steps (Fig. 12), differing only in the relationship of Magnoliales and core Laurales. These trees are generally consistent with those derived from the whole rRNA data set (cf. Fig. 8), except in the exact arrangement of eudicots and woody magnoliids. The consistency index is 0.58, the same as in the morphological analysis; the retention index is 0.66. Based on this comparison, there is no reason to assume a priori that the rRNA data are any more or less reliable than the morphological data, although strong conclusions on relative consistency would be unwarranted because of differing amounts of missing data.

The arrangement of cycads, conifers, and *Ginkgo* differs from that derived from morphology, in that the group closest to anthophytes is *Ginkgo*

rather than conifers. Angiosperms and Gnetales are united by at least 12 characters. However, it should be noted that conclusions on angiosperm outgroups are a function of the rooting of seed plants as a whole. When the whole rRNA data set is rooted with *Equisetum* and *Psilotum* (as in Hamby & Zimmer, 1992), it is only two steps less parsimonious to associate angiosperms with a clade consisting of *Ginkgo*, cycads, and conifers, rather than with Gnetales. The figure of 12 synapomorphies holds if seed plants are rooted somewhere among cycads, conifers, and *Ginkgo*, which we consider most likely. Certainly this is more consistent with analyses that include fossils (Crane, 1985; Doyle & Donoghue, 1986, 1992) and with the stratigraphic record; cycads, conifers, and ginkgos appear in the Late Carboniferous or Permian, but Gnetales (or forms on the line leading to them) are not known before the Late Triassic (Crane, 1988; Doyle & Donoghue, 1993).

Angiosperms themselves are united by at least 23 characters. As with previous rRNA analyses (Hamby & Zimmer, 1992), they are rooted among the paleoherbs: Nymphaeales are the sister group of other angiosperms, followed by Piperales, then Aristolochiaceae plus monocots. The fact that there are 13 unambiguous changes uniting Nymphaeales and four more between the nodes where Nymphaeales and Piperales are attached may raise suspicion that this rooting is an artifact of long branch attraction between Nymphaeales and the outgroups. However, at least two considerations argue
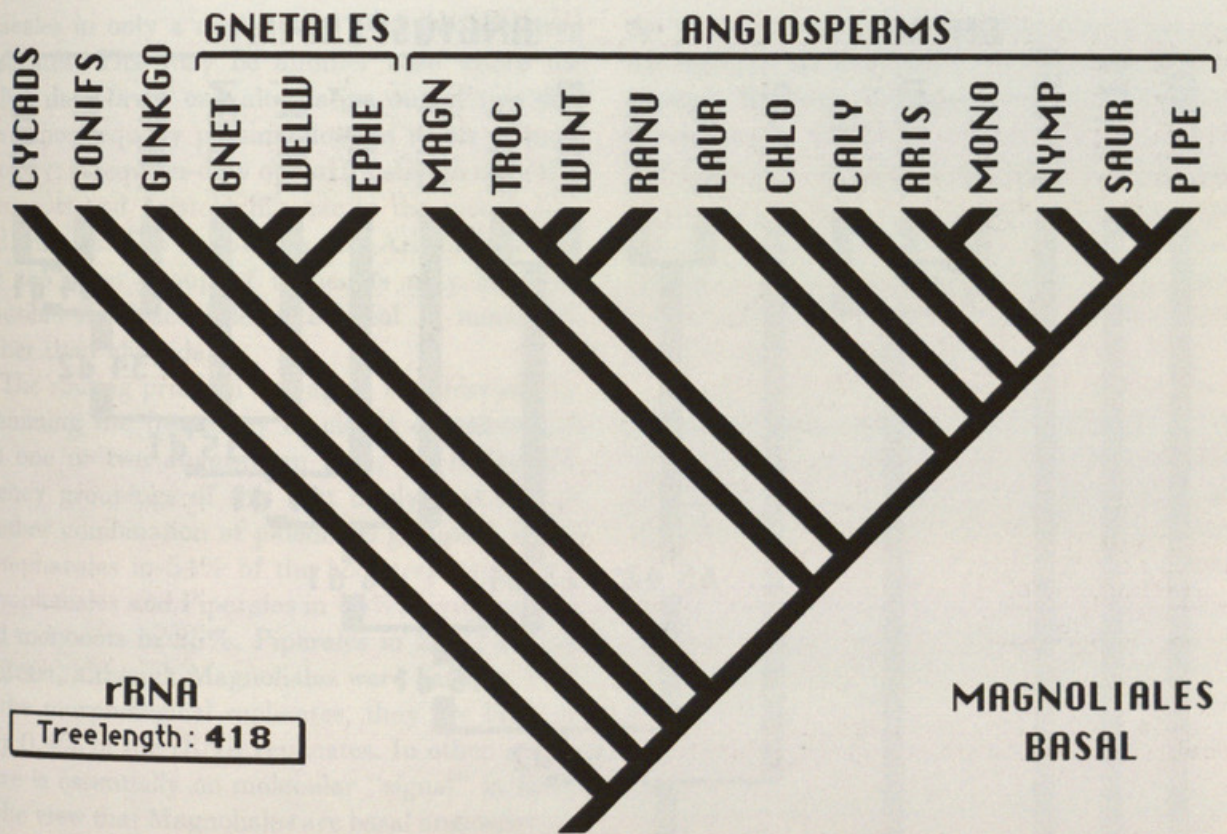
FIGURE 13.   Representative tree found with the rRNA data set when Magnoliales are forced to the base of angiosperms and paleoherbs are forced together as a clade, as in most parsimonious trees of Donoghue & Doyle (1989a) and Figure 10 above. This tree is 13 steps longer than the most parsimonious trees. Abbreviations as in Figure 8.

against this. First, trees with the positions of Piperales and Nymphaeales reversed, which are equally "paleoherb-rooted," are only two steps longer, even though Piperales are a relatively short branch. Second, reanalysis of the data set without Nymphaeales results in two trees otherwise identical to those in Figure 12, with Piperales basal.

Experiments in which alternative groups were forced together with the constraints option in PAUP also support the view that the conflicts between the rRNA and morphological results are not serious. For example, Winteraceae are interpolated between Ranunculidae and Trochodendrales in the rRNA trees, breaking up the eudicots, but forcing the eudicots together adds only two steps (if Magnoliales are linked with core Laurales). On the other hand, whereas trees rooted next to Magnoliales and among paleoherbs are almost equally parsimonious in terms of morphology, the magnolialian rooting is much less parsimonious in terms of rRNA data. When Magnoliales are forced to the base of the angiosperms, the resulting trees are nine steps longer than the shortest trees (414 steps). Furthermore, these trees are not closely analogous to magnolialian-rooted trees based on morphology, because paleoherb groups are interpolated in various ar-

rangements between Magnoliales and other woody magnoliids and eudicots. These are essentially paleoherb-rooted trees with Magnoliales alone pulled to the base. A more analogous tree, obtained by rerooting one of the most parsimonious angiosperm networks on the line leading to Magnoliales (Fig. 13), is 13 steps longer than the most parsimonious trees.

Experiments of this kind fail to support other current hypotheses on the rooting of angiosperms. Trees with Chloranthaceae forced to the base of the angiosperms (cf. Taylor & Hickey, 1992) are 10 steps longer than the shortest trees, roughly the same deficit seen when Magnoliales are basal. Trees with Calycanthaceae basal in angiosperms (cf. Loconte & Stevenson, 1991) are eight steps longer. As when Magnoliales are forced to the base, paleoherbs are interpolated between the basal group and other woody magnoliids and eudicots. The shortest trees obtained by rerooting the most parsimonious angiosperm networks on the line leading to Chloranthaceae and Calycanthaceae are 11 and nine steps longer, respectively; the latter tree is relatively parsimonious because paleoherbs are a next-most-basal clade. Trees with eudicots as the sister group of other (monosulcate) angiosperms,
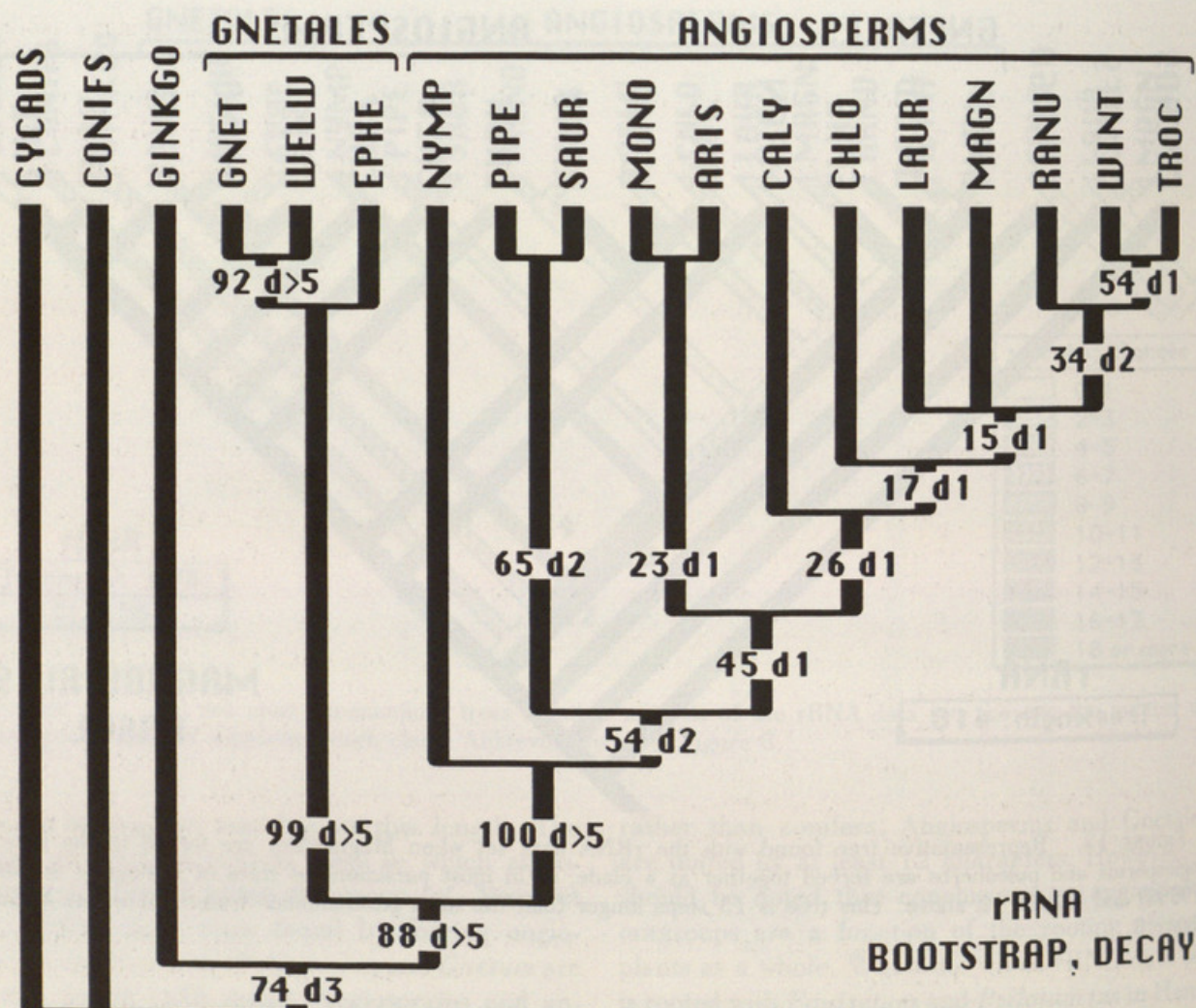
FIGURE 14. Results of bootstrap and decay analyses of the rRNA data set (see Fig. 11 for explanation). Abbreviations as in Figure 8.

analogous to trees based on *rbc*L data (Les et al., 1991; Chase et al., 1993; Qiu et al., 1993), are 14 steps longer than the shortest trees.

Results of bootstrap analysis of the rRNA data (Fig. 14) also complement those based on morphology. Again, the monophyly of the angiosperms is very strongly supported (99.997%). The link between angiosperms and Gnetales is weaker (88%), but the monophyly of Gnetales is stronger (99%). This result supports the view that the weaker morphological support for Gnetales is due strictly to morphological convergences between the subgroup consisting of *Welwitschia* and *Gnetum* and angiosperms, and it argues against trees in which angiosperms are derived from (nested within) Gnetales (Nixon et al., 1994).

It may be objected that molecular evidence for angiosperm monophyly applies only to living groups, leaving open the possibility that different "angiosperm" lines were derived polyphyletically from different fossil "gymnosperm" lines. However, this objection is valid only if all angiosperms and their fossil relatives are more closely related to each other than they are to any living gymnosperm group, not if some angiosperm line is more closely related to any living gymnosperm group—such as Gnetales, as assumed by most polyphyleticists. For example, if some angiosperms were related to *Caytonia* and others to Gnetales, and if molecular data gave the correct relationships among living taxa, angiosperms would form one branch associated with Gnetales and another located one or more nodes below, not a clade.

Within angiosperms, Piperales are less well supported than they were with morphology (65%), but they are the strongest grouping, again contrary to Taylor & Hickey (1992). On the other hand, none of the rRNA links among angiosperms that conflicted with the morphological results are very strong. The grouping of Winteraceae and Trochodendrales, which breaks up the eudicots, appears at a frequency of only 54%. Although the connection between monocots and Aristolochiaceae is weak (23%), monocots are linked with Nym-

Volume 81, Number 3
1994

Doyle et al.
Integration of Morphological and rRNA Data

439

phaeales in only a negligible 1% of the bootstrap replicates. This may be another case where the rRNA data favor one alternative out of two that are almost equally parsimonious in terms of morphology; it requires only one extra step to associate monocots and Aristolochiaceae in the morphological analysis. The hypothesis that Aristolochiaceae are the sister group of monocots may be more plausible if Dioscoreales are basal in monocots, rather than alismids.

The rooting problem can again be addressed by examining the frequency of clades containing all but one or two angiosperm taxa. All higher-frequency groupings of this sort imply that one or another combination of paleoherb groups is basal: Nymphaeales in 54% of the bootstrap replicates, Nymphaeales and Piperales in 45%, Nymphaeales and monocots in 25%, Piperales in 22%, etc. In contrast, although Magnoliales were basal in 11% of the morphological replicates, they are basal in only 0.4% of the rRNA replicates. In other words, there is essentially no molecular "signal" in favor of the view that Magnoliales are basal angiosperms. This analysis also fails to support the concept that Calycanthaceae (Loconte & Stevenson, 1991) or Chloranthaceae (Taylor & Hickey, 1992) are basal: the corresponding groupings are observed at frequencies of less than 0.2%.

In the decay analysis (Fig. 14), Piperales and the group consisting of Winteraceae, Ranunculidae, and Trochodendrales are the only angiosperm clades left in the consensus of one-off trees. Both of these groups decay in two-off trees. The arrangement of cycads, conifers, and *Ginkgo* breaks down in three steps, but angiosperms, Gnetales, the relationship between them, and the association of *Welwitschia* and *Gnetum* are still intact in five-off trees, beyond which the analysis was abandoned. This decay order is generally consistent with the relative strength of clades inferred from the bootstrap analysis, although less precisely than with the morphological data.

RESULTS OF COMBINED ANALYSES

Analysis of the combined data set (Fig. 15) yields only one tree of 563 steps. The consistency index is 0.57, which is almost identical to that in the two separate analyses (0.58); the retention index is 0.67. This refutes one possible argument against combining morphological and molecular data, namely that adding two homoplastic data sets should only result in more total noise.

Examination of Figure 15 and alternative trees demonstrates graphically the complementarity of the two data sets. In non-angiospermous groups, the Loconte & Stevenson (1990) arrangement is favored. However, the conifer–anthophyte link is unequivocally supported by only six characters, and trees with conifers and *Ginkgo* reversed are only one step longer. The strong links inferred from both component data sets are seen among Gnetales (united by at least 26 characters), among angiosperms (40 characters), and between angiosperms and Gnetales (15 characters).

As expected from the ambiguity of the morphological results and the relative strength of the rRNA results, angiosperms are rooted among the paleoherbs, with Nymphaeales basal. Trees with Magnoliales forced to the base of the angiosperms are 13 steps longer than the shortest tree; trees with Chloranthaceae and Calycanthaceae basal are 14 and 15 steps longer, respectively.

Despite the much greater number of rRNA characters, other results are more consistent with the morphological analysis, in keeping with the expectation that even a few morphological characters can be decisive when molecular data are ambiguous. Laurales and Magnoliales form a monophyletic group, as in the morphological trees, rather than a paraphyletic grade, as in the rRNA trees. Monocots are interpolated between Nymphaeales and Piperales, although it costs only one extra step to link them with Aristolochiaceae, as in the rRNA analysis. Most significantly, Ranunculidae and Trochodendrales form a eudicot clade—Winteraceae are dissociated from Trochodendrales and are instead the sister group of Laurales and Magnoliales.

The general result of the bootstrap analysis (Fig. 16) is that the two data sets tend to reinforce each other in cases where they were congruent. Angiosperms (100%), Gnetales (100%), Piperales (99%), and *Welwitschia* plus *Gnetum* (98%) form clades at bootstrap frequencies similar to or higher than those in the separate analyses. Angiosperms and Gnetales are united at the 98% level, rather than 95% in the morphological analysis and 88% in the rRNA analysis. This again belies the fear that conflicting patterns of homoplasy will simply lower overall resolution, and it suggests instead that both analyses are detecting the same real phylogenetic signal. Conversely, basal seed plant relationships, which conflicted but were weakly supported in both analyses, are less resolved: conifers and anthophytes are linked at a frequency of only 54%, rather than 79% in the morphological analysis.

A more subtle effect is that the two data sets also seem to reinforce each other in one case where the most parsimonious trees derived from them
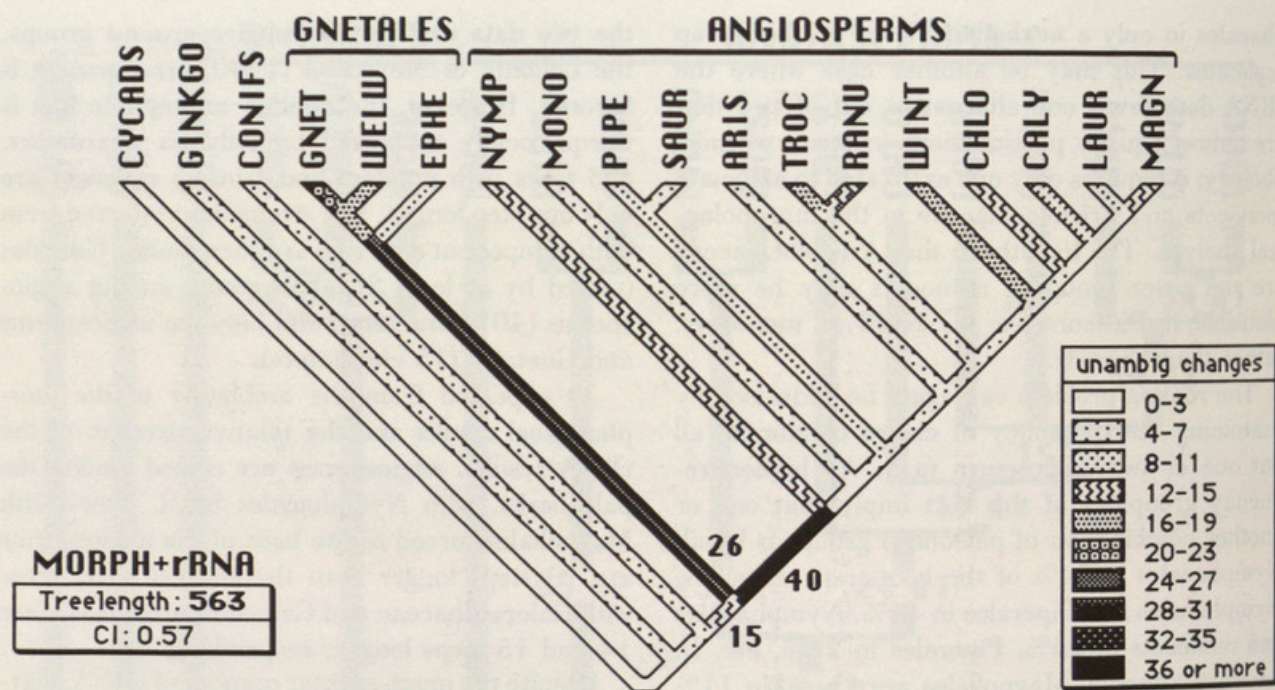
FIGURE 15. Single most parsimonious tree found in analysis of the combined morphological and rRNA data set, showing the number of unambiguous changes supporting each clade. Abbreviations as in Figure 8.

were *not* congruent. This involves the link between the two eudicot taxa, which is seen in 50% of the replicates in the combined analysis, rather than 43% in the morphological analysis and only 12% in the molecular analysis. We suspect that this reflects the existence of "minority" rRNA characters that support the eudicots (also supported by *rbc*L data: Chase et al., 1993). Once Winteraceae are forced outside eudicots by the morphological characters, these molecular characters reinforce those from morphology. The existence of such effects is a general argument for combining data sets (Barrett et al., 1991).

As in the rRNA analysis, the relative support for clades inferred from the decay analysis (Fig. 16) roughly parallels the bootstrap results. All the stronger clades decay more slowly than they did in either individual analysis, in keeping with the larger total number of characters. Eudicots appear to be more robust than implied by the bootstrap; they do not decay until four step less parsimonious trees. All groups retained in the four-off trees are still present in eight-off trees, beyond which the analysis was abandoned.

CONCLUSIONS

These exercises clearly show the utility of combining molecular and morphological data sets as well as analyzing them separately. This procedure has the potential of resolving conflicts between data sets even when one is much larger, presumably

because there are minority characters in each data set that reflect true historical relationships (Barrett et al., 1991).

The strongest results of these analyses are that angiosperms, Gnetales, and Piperales (Piperaceae plus Saururaceae, but not Chloranthaceae) are monophyletic groups, and that Gnetales are the closest living relatives of angiosperms. We suggest that polyphyly of angiosperms can be set aside and other reasons examined for lack of progress in understanding the origin of the group (cf. Donoghue & Doyle, 1991). In contrast, relationships among cycads, conifers, *Ginkgo*, and anthophytes appear to be quite unresolved on present data, even when fossil taxa are considered (Doyle & Donoghue, 1992).

The potentially most significant result of this exercise concerns the rooting of the angiosperms. There appears to be essentially no rRNA support for the conclusion derived from morphology that Magnoliales are basal angiosperms, and a magnolialian rooting is increasingly ambiguous in terms of morphological data. Since there are still apparent conflicts with other molecular data sets, it would be premature to consider the paleoherb rooting established. However, our own preliminary experiments in combining morphological, rRNA, and *rbc*L data also give a paleoherb rooting, specifically between monocots and dicots. These analyses are not strictly comparable to the analyses presented above, since so far we have only used single exemplar species in the *rbc*L data set to represent

Volume 81, Number 3
1994

Doyle et al.
Integration of Morphological and rRNA Data

441

GNETALES

ANGIOSPERMS

CYCADS GINKGO CONIFS GNET WELW EPHE NYMP MONO PIPE SAUR ARIS TROC RANU WINT CHLO CALY LAUR MAGN

98 d>8

30 d1

51 d1

51 d3

99 d>8   50 d4   32 d3

53 d3

33 d1

45 d1

65 d1

100 d>8   100 d>8

98 d>8

54 d1

MORPH+rRNA
BOOTSTRAP, DECAY
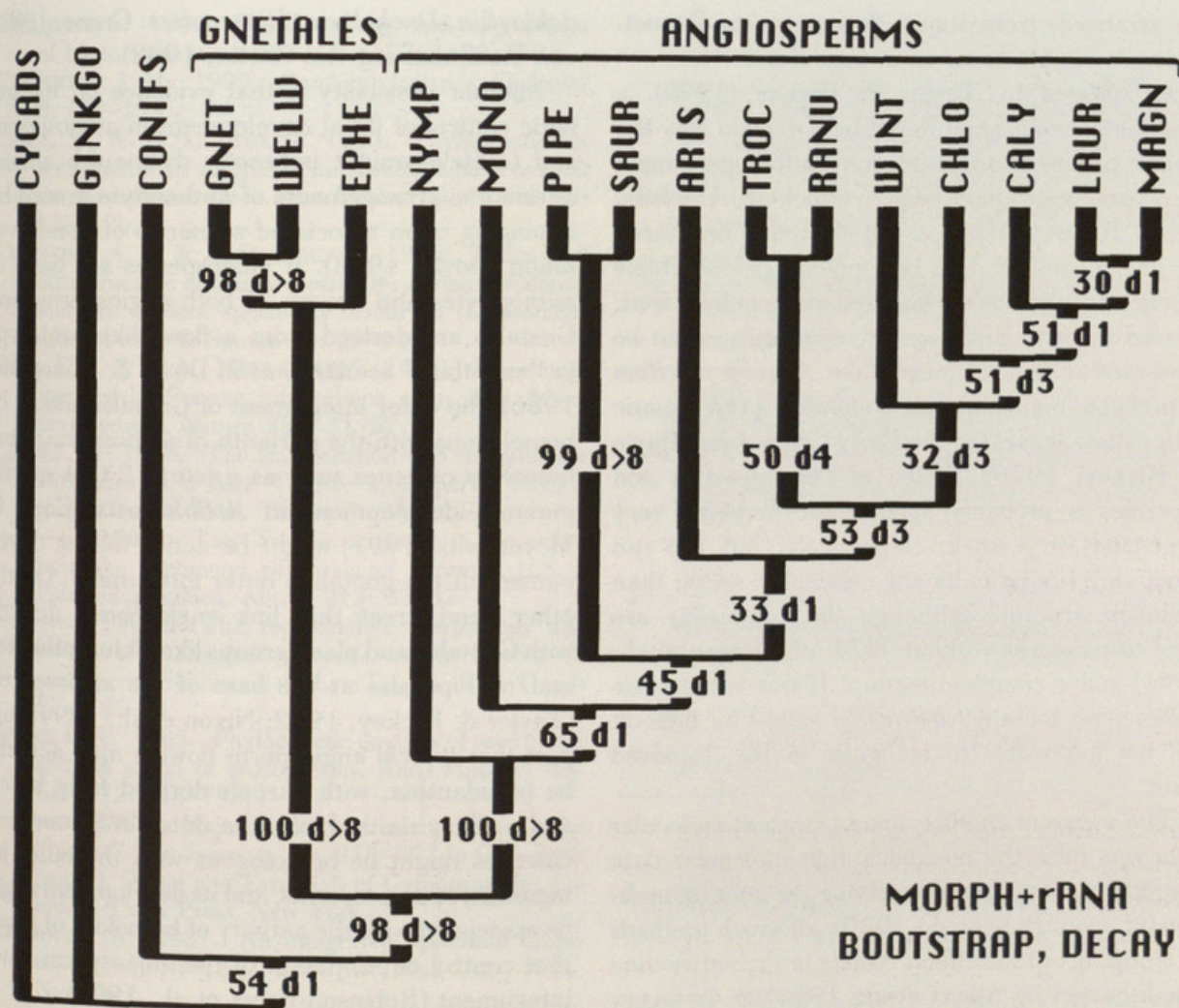
FIGURE 16.   Results of bootstrap and decay analyses of the combined morphological and rRNA data set (see Fig. 11 for explanation). Abbreviations as in Figure 8.

presumed clades. Still, it appears that the *rbc*L data do not support some different arrangement so strongly that they overwhelm the other data, even though there are more potentially informative *rbc*L characters than rRNA characters.

Until the disagreements among molecular data sets are resolved, paleoherbs clearly deserve as much attention from botanists as has been paid to woody magnoliids. To discourage potential misconceptions, we should clarify what a paleoherb rooting would say about primitive conditions (cf. Taylor & Hickey, 1992). Although our trees imply that the first angiosperms would be at least semiherbaceous, the aquatic habit and complete lack of secondary growth in Nymphaeales and monocots may be autapomorphies of these groups. Leaves would have more or less palmate venation and anomocytic stomates; contrary to Doyle & Donoghue (1986), the paracytic stomates of woody magnoliids would be a convergence with Bennettitales. The fact that Piperales are near-basal might seem to support the view that small, crowded flowers with orthotropous

ovules and no perianth are primitive (Burger, 1977; Taylor & Hickey, 1990, 1992). However, if taxa are arranged as in Figures 9, 12, or 15, it is more parsimonious to assume that angiosperms originally had flowers like those of Cabombaceae, *Lactoris*, *Saruma* (Aristolochiaceae), and monocots, with one or two cycles of three perianth parts, a trimerous androecium and gynoecium, and anatropous ovules. The syncarpous gynoecium of Nymphaeaceae and the tubular calyx and inferior ovary of most Aristolochiaceae would be autapomorphies, although the laminar placentation of Nymphaeales as a whole might be primitive. Stamens would be differentiated into filament and anther, not laminar as in woody magnoliids. The basal position of Nymphaeales and Piperales raises the intriguing possibility that the presence of both endosperm and perisperm in seeds of these orders is a primitive transitional state, not derived as usually assumed (D. Haig, pers. comm.). The conclusion that Gnetales are the closest modern relatives of angiosperms permits a wide range of floral prototypes, depending on how fossil taxa

are arranged, from showy flowers, as in Bennet-titales, to simple ones, as in Gnetales.

As stressed by Taylor & Hickey (1990), a (semi)herbaceous habit might help explain why the earliest phases of angiosperm evolution and angiosperm precursors have been overlooked in the fossil record. However, this does not mean that the search for paleobotanical data is hopeless: paleoherblike precursors might be represented in the pollen, fruit, or seed records, and vegetative remains might be preserved in special facies, like *Acaciaephyllum* (a probable monocot) and *Nelumbites* (an aquatic with peltate leaves) in the Early Cretaceous (Doyle & Hickey, 1976). Pollen of Nymphaeales and Piperales is probably apomorphic in being very large and very small, respectively, but the two groups are similar in having columellar rather than granular structure (although the columellae are hard to recognize without TEM: cf. Osborn et al., 1991) and a complete tectum. If this is the basic pollen type for angiosperms, it would be difficult but not impossible to recognize in the dispersed state.

The apparent conflicts among present molecular data sets raise the possibility that molecular data are simply incapable of resolving the rooting problem (cf. Donoghue et al., 1989), although methods of factoring out the effects of long branch attraction (as discussed by Albert et al., 1994) or discovery of genome rearrangements or duplications that occurred early in the angiosperm radiation (cf. Iwabe et al., 1989; Raubeson & Jansen, 1992) might permit firmer inferences. It is possible that significant progress on the origin of angiosperms will require recognition of fossil forms on the long branch leading to the group. Potential examples include *Phyllites* (Seward, 1904), a Jurassic leaf with paleoherblike palmate venation; Triassic Crinopolles pollen (Cornet, 1989a), with monocotlike sculpture; and the still-enigmatic Triassic fossil *Sanmiguelia* (Cornet, 1986, 1989b; Doyle & Hotton, 1991; Doyle & Donoghue, 1993). Better evidence on the morphology of *Caytonia* and glossopterids—whether or not they have anthophyte states in currently unknown characters, as required in trees where they are linked with angiosperms—could also have a decisive effect in choosing among alternative angiosperm relationships. Better reconstructions of primitive members or stem-relatives of Bennettitales and Gnetales could clarify whether flowers are indeed a synapomorphy of anthophytes or arose independently in each anthophyte line, a possibility raised by the existence of less flowerlike Late Triassic–Early Jurassic reproductive structures related to these groups (*Westersheimia*, *Var-*

dekloeftia, Dechellyia, Piroconites: Crane, 1988; van Konijnenburg-van Cittert, 1992).

Another possibility is that evidence on the genetic control of floral development in angiosperms and Gnetales might indirectly distinguish among alternative arrangements of anthophyte groups by impinging upon associated scenarios of floral evolution (Doyle, 1993). If angiosperms are basal in anthophytes and flowers of both angiosperms and Gnetales are derived from a flowerlike prototype (a "euanthial" scenario, as in Doyle & Donoghue, 1986), the outer integument of Gnetales should be homologous with the perianth of angiosperms, and homologs of genes such as *apetala 2* that specify perianth development in *Arabidopsis* (Coen & Meyerowitz, 1991) might be active during development of the gnetalian outer integument. On the other hand, trees that link angiosperms directly with Gnetales and place groups like Chloranthaceae and/or Piperales at the base of the angiosperms (Taylor & Hickey, 1992; Nixon et al., 1994) suggest that typical angiosperm flowers may actually be pseudanthia, with carpels derived from bracts and axillary units. If so, the outer integument of Gnetales might be homologous with the outer integument of angiosperms, and its development might be associated with the activity of homologs of genes that control development of the angiosperm outer integument (Robinson-Beers et al., 1992).

LITERATURE CITED

ALBERT, V. A., A. BACKLUND, K. BREMER, M. W. CHASE, J. R. MANHART, B. D. MISHLER & K. C. NIXON. 1994. Functional constraints and *rbc*L evidence for land plant phylogeny. Ann. Missouri Bot. Gard. 81: 534–567.

BARRETT, M., M. J. DONOGHUE & E. SOBER. 1991. Against consensus. Syst. Zool. 40: 486–493.

BECK, C. B., R. SCHMID & G. W. ROTHWELL. 1982. Stelar morphology and the primary vascular system of seed plants. Bot. Rev. 48: 691–815.

BEHNKE, H.-D. 1988. Sieve-element plastids, phloem protein, and evolution of flowering plants: III. Magnoliidae. Taxon 37: 699–732.

BORNSTEIN, A. J. 1989. Taxonomic studies in the Piperaceae—I. The pedicellate pipers of Mexico and Central America (*Piper* subg. *Arctottonia*). J. Arnold Arbor. 70: 1–55.

BREMER, K. 1988. The limits of amino acid sequence data in angiosperm phylogenetic reconstruction. Evolution 42: 795–803.

BULL, J. J., J. P. HUELSENBECK, C. W. CUNNINGHAM, D. L. SWOFFORD & P. J. WADDELL. 1993. Partitioning and combining data in phylogenetic analysis. Syst. Zool. 42: 384–397.

BURGER, W. C. 1977. The Piperales and the monocots. Alternative hypotheses for the origin of monocotyledonous flowers. Bot. Rev. 43: 345–393.

———. 1981. Heresy revived: The monocot theory of angiosperm origin. Evol. Theory 5: 189–225.

Volume 81, Number 3
1994

Doyle et al.
Integration of Morphological and rRNA Data

443

CARLQUIST, S. 1990. Wood anatomy and relationships of Lactoridaceae. Amer. J. Bot. 77: 1498–1504.

CARPENTER, J. M. 1992. Random cladistics. Cladistics 8: 147–153.

CHASE, M. W. & 41 OTHERS. 1993. Phylogenetics of seed plants: An analysis of nucleotide sequences from the plastid gene *rbc*L. Ann. Missouri Bot. Gard. 80: 528–580.

CHLONOVA, A. F. & T. D. SUROVA. 1988. Pollen wall ultrastructure of *Clavatipollenites incisus* Chlonova and two modern species of *Ascarina* (Chloranthaceae). Pollen & Spores 30: 29–44.

COEN, E. S. & E. M. MEYEROWITZ. 1991. The war of the whorls: Genetic interactions controlling flower development. Nature 353: 31–37.

CORNET, B. 1986. The leaf venation and reproductive structures of a Late Triassic angiosperm, *Sanmiguelia lewisii*. Evol. Theory 7: 231–309.

———. 1989a. Late Triassic angiosperm-like pollen from the Richmond rift basin of Virginia, U.S.A. Palaeontographica, Abt. B 213: 37–87.

———. 1989b. The reproductive morphology and biology of *Sanmiguelia lewisii*, and its bearing on angiosperm evolution in the Late Triassic. Evol. Trends Pl. 3: 25–51.

CRANE, P. R. 1985. Phylogenetic analysis of seed plants and the origin of angiosperms. Ann. Missouri Bot. Gard. 72: 716–793.

———. 1988. Major clades and relationships in the "higher" gymnosperms. Pp. 218–272 *in* C. B. Beck (editor), Origin and Evolution of Gymnosperms. Columbia Univ. Press, New York.

CRONQUIST, A. 1981. An Integrated System of Classification of Flowering Plants. Columbia Univ. Press, New York.

———. 1988. The Evolution and Classification of Flowering Plants, 2nd ed. New York Botanical Garden, Bronx, New York.

DAVIS, G. L. 1966. Systematic Embryology of the Angiosperms. John Wiley & Sons, New York.

DONOGHUE, M. J. 1994. Progress and prospects in reconstructing plant phylogeny. Ann. Missouri Bot. Gard. 81: 405–418.

——— & J. A. DOYLE. 1989a. Phylogenetic analysis of angiosperms and the relationships of Hamamelidae. Pp. 17–45 *in* P. R. Crane & S. Blackmore (editors), Evolution, Systematics, and Fossil History of the Hamamelidae, vol. 1. Clarendon Press, Oxford.

——— & J. A. DOYLE. 1989b. Phylogenetic studies of seed plants and angiosperms based on morphological characters. Pp. 181–193 *in* B. Fernholm, K. Bremer & H. Jörnvall (editors), The Hierarchy of Life. Elsevier, Amsterdam.

——— & ———. 1991. Angiosperm monophyly. Trends Ecol. Evol. 6: 407.

——— & M. J. SANDERSON. 1992. The suitability of molecular and morphological evidence in reconstructing plant phylogeny. Pp. 340–368 *in* P. S. Soltis, D. E. Soltis & J. J. Doyle (editors), Molecular Systematics of Plants. Chapman & Hall, New York.

——— & S. M. SCHEINER. 1992. The evolution of endosperm: A phylogenetic account. Pp. 356–389 *in* R. Wyatt (editor), Ecology and Evolution of Plant Reproduction: New Approaches. Chapman & Hall, New York.

———, R. G. OLMSTEAD, J. F. SMITH & J. D. PALMER. 1992. Phylogenetic relationships of Dipsacales based on *rbc*L sequences. Ann. Missouri Bot. Gard. 79: 333–345.

———, J. A. DOYLE, J. GAUTHIER, A. G. KLUGE & T. ROWE. 1989. The importance of fossils in phylogeny reconstruction. Annual Rev. Ecol. Syst. 20: 431–460.

DOYLE, J. A. 1993. Cladistic and paleobotanical perspectives on the origin of angiosperm organs. J. Cell. Biochem., Suppl. 17B: 8.

——— & M. J. DONOGHUE. 1986. Seed plant phylogeny and the origin of angiosperms: An experimental cladistic approach. Bot. Rev. 52: 321–431.

——— & ———. 1987. The importance of fossils in elucidating seed plant phylogeny and macroevolution. Rev. Palaeobot. Palynol. 50: 63–95.

——— & ———. 1990. Reexamination of seed plant and angiosperm relationships. Amer. J. Bot. 77(6, Suppl.): 84.

——— & ———. 1992. Fossils and seed plant phylogeny reanalyzed. Brittonia 44: 89–106.

——— & ———. 1993. Phylogenies and angiosperm diversification. Paleobiology 19: 141–167.

——— & L. J. HICKEY. 1976. Pollen and leaves from the mid-Cretaceous Potomac Group and their bearing on early angiosperm evolution. Pp. 139–206 *in* C. B. Beck (editor), Origin and Early Evolution of Angiosperms. Columbia Univ. Press, New York.

——— & C. L. HOTTON. 1991. Diversification of early angiosperm pollen in a cladistic context. Pp. 169–195 *in* S. Blackmore & S. H. Barnes (editors), Pollen and Spores: Patterns of Diversification. Clarendon Press, Oxford.

DUVALL, M. R., M. T. CLEGG, M. W. CHASE, W. D. CLARK, W. J. KRESS, H. G. HILLS, L. E. EGUIARTE, J. F. SMITH, B. S. GAUT, E. A. ZIMMER & G. H. LEARN, JR. 1993. Phylogenetic hypotheses for the monocotyledons constructed from *rbc*L sequence data. Ann. Missouri Bot. Gard. 80: 607–619.

EHRENDORFER, F. 1976. Evolutionary significance of chromosomal differentiation patterns in gymnosperms and primitive angiosperms. Pp. 220–240 *in* C. B. Beck (editor), Origin and Early Evolution of Angiosperms. Columbia Univ. Press, New York.

ENDRESS, P. K. 1983. Dispersal and distribution in some small archaic relic angiosperm families (Austrobaileyaceae, Eupomatiaceae, Himantandraceae, Idiospermoideae–Calycanthaceae). Sonderb. Naturwiss. Vereins Hamburg 7: 201–217.

———. 1987. The Chloranthaceae: Reproductive structures and phylogenetic position. Bot. Jahrb. Syst. 109: 153–226.

FELSENSTEIN, J. 1978. Cases in which parsimony or compatibility methods will be positively misleading. Syst. Zool. 27: 401–410.

———. 1985. Confidence limits on phylogenies: An approach using the bootstrap. Evolution 39: 783–791.

——— & H. KISHINO. 1993. Is there something wrong with the bootstrap on phylogenies? A reply to Hillis and Bull. Syst. Zool. 42: 193–200.

FRIEDMAN, W. E. 1990. Sexual reproduction in *Ephedra nevadensis* (Ephedraceae): Further evidence of double fertilization in a nonflowering seed plant. Amer. J. Bot. 77: 1582–1598.

———. 1992. Evidence of a pre-angiosperm origin of endosperm: Implications for the evolution of flowering plants. Science 255: 336–339.

GABARAYEVA, N. I. 1991. Patterns of development in primitive angiosperm pollen. Pp. 257–268 *in* S. Blackmore & S. H. Barnes (editors), Pollen and Spores: Patterns of Diversification. Clarendon Press, Oxford.

GIBBS, R. D. 1957. The Mäule reaction, lignin, and the relationships between woody plants. Pp. 269–312 *in* K. V. Thimann (editor), The Physiology of Forest Trees. Ronald Press, New York.

GIFFORD, E. M. & A. S. FOSTER. 1989. Morphology and Evolution of Vascular Plants. W. H. Freeman, New York.

GILLESPIE, L. J. & J. W. NOWICKE. 1992. Systematic implications of pollen morphology in *Gnetum*. 8th Int. Palynol. Congr. (Aix-en-Provence) Abstr. 53.

GOLENBERG, E. M., D. E. GIANNASI, M. T. CLEGG, C. J. SMILEY, M. DURBIN, D. HENDERSON & G. ZURAWSKI. 1990. Chloroplast DNA sequence from a Miocene *Magnolia* species. Nature 344: 656–658.

HAMBY, R. K. & E. A. ZIMMER. 1992. Ribosomal RNA as a phylogenetic tool in plant systematics. Pp. 50–91 *in* P. S. Soltis, D. E. Soltis & J. J. Doyle (editors), Molecular Systematics of Plants. Chapman & Hall, New York.

HART, J. A. 1987. A cladistic analysis of conifers: Preliminary results. J. Arnold Arbor. 68: 269–307.

HASEBE, M., K. KOFUJI, M. ITO, M. KATO, K. IWATSUKI & K. UEDA. 1992. Phylogeny of gymnosperms inferred from *rbc*L gene sequences. Bot. Mag. (Tokyo) 105: 673–679.

HILLIS, D. M. 1987. Molecular versus morphological approaches to systematics. Annual Rev. Ecol. Syst. 18: 23–42.

——— & J. J. BULL. 1993. An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. Syst. Biol. 42: 182–192.

HUGHES, N. F. & A. B. MCDOUGALL. 1990. Barremian-Aptian angiospermid pollen records from southern England. Rev. Palaeobot. Palynol. 65: 145–151.

ITO, M. 1987. Phylogenetic systematics of the Nymphaeales. Bot. Mag. (Tokyo) 100: 17–35.

IWABE, N., K. KUMA, M. HASEGAWA, S. OSAWA & T. MIYATA. 1989. Evolutionary relationship of archaebacteria, eubacteria, and eukaryotes inferred from phylogenetic trees of duplicated genes. Proc. Natl. Acad. Sci. U.S.A. 86: 9355–9359.

KELLOGG, E. A. & C. S. CAMPBELL. 1987. Phylogenetic analyses of the Gramineae. Pp. 310–322 *in* T. R. Soderstrom, K. W. Hilu, C. S. Campbell & M. E. Barkworth (editors), Grass Systematics and Evolution. Smithsonian Institution, Washington, D.C.

KLUGE, A. J. 1989. A concern for evidence and a phylogenetic hypothesis of relationships among *Epicrates* (Boidae, Serpentes). Syst. Zool. 38: 7–25.

KRASSILOV, V. A. 1991. The origin of angiosperms: New and old problems. Trends Ecol. Evol. 6: 215–220.

LES, D. H., D. K. GARVIN & C. F. WIMPEE. 1991. Molecular evolutionary history of ancient aquatic angiosperms. Proc. Natl. Acad. Sci. U.S.A. 88: 10119–10123.

LE THOMAS, A. 1980–1981. Ultrastructural characters of the pollen grains of African Annonaceae and their significance for the phylogeny of primitive angiosperms. Pollen & Spores 22: 267–342, 23: 5–36.

LOCONTE, H. & D. W. STEVENSON. 1990. Cladistics of the Spermatophyta. Brittonia 42: 197–211.

——— & ———. 1991. Cladistics of the Magnoliidae. Cladistics 7: 267–296.

MADDISON, D. R. 1991. The discovery and importance of multiple islands of most-parsimonious trees. Syst. Zool. 40: 315–328.

———, M. RUVOLO & D. L. SWOFFORD. 1992. Geographic origins of human mitochondrial DNA: Phylogenetic evidence from control region sequences. Syst. Biol. 41: 111–124.

MADDISON, W. P. & D. R. MADDISON. 1992. MacClade: Analysis of Phylogeny and Character Evolution, Version 3. Sinauer Associates, Sunderland, Massachusetts.

METCALFE, C. R. 1987. Anatomy of the Dicotyledons, 2nd ed., vol. III. Magnoliales, Illiciales, and Laurales (*Sensu* Armen Takhtajan). Clarendon Press, Oxford.

MEYEN, S. V. 1984. Basic features of gymnosperm systematics and phylogeny as evidenced by the fossil record. Bot. Rev. 50: 1–112.

MISHLER, B. D. 1994. The cladistic analysis of molecular and morphological data. Amer. J. Phys. Anthropol. 94: 143–156.

MONEY, L. L., I. W. BAILEY & B. G. L. SWAMY. 1950. The morphology and relationships of the Monimiaceae. J. Arnold Arbor. 31: 372–404.

NIXON, K. C. & J. I. DAVIS. 1991. Polymorphic taxa, missing values and cladistic analysis. Cladistics 7: 233–241.

———, W. L. CREPET, D. STEVENSON & E. M. FRIIS. 1994. A reevaluation of seed plant phylogeny. Ann. Missouri Bot. Gard. 81: 484–533.

OSBORN, J. M., T. N. TAYLOR & E. L. SCHNEIDER. 1991. Pollen morphology and ultrastructure of the Cabombaceae: Correlations with pollination biology. Amer. J. Bot. 78: 1367–1378.

QIU, Y.-L., M. W. CHASE, D. H. LES & C. R. PARKS. 1993. Molecular phylogenetics of the Magnoliidae: Cladistic analyses of nucleotide sequences of the plastid gene *rbc*L. Ann. Missouri Bot. Gard. 80: 587–606.

QUEIROZ, A. DE. 1993. For consensus (sometimes). Syst. Biol. 42: 368–372.

RAUBESON, L. A. & R. K. JANSEN. 1992. Chloroplast DNA evidence on the ancient evolutionary split in vascular land plants. Science 255: 1697–1699.

ROBINSON-BEERS, K., R. E. PRUITT & C. S. GASSER. 1992. Ovule development in wild type Arabidopsis and two female-sterile mutants. Pl. Cell 4: 1237–1249.

ROTHWELL, G. W. 1982. New interpretations of the earliest conifers. Rev. Palaeobot. Palynol. 37: 7–28.

SANDERSON, M. J. & M. J. DONOGHUE. 1989. Patterns of variation in levels of homoplasy. Evolution 43: 1781–1795.

SERBET, R. & G. W. ROTHWELL. 1992. Characterizing the most primitive seed ferns: I. A reconstruction of *Elkinsia polymorpha*. Int. J. Pl. Sci. 153: 602–621.

SEWARD, A. C. 1904. Catalogue of Mesozoic Plants in the British Museum. The Jurassic Flora, Part 1. British Museum (Natural History), London.

STEVENSON, D. W. 1990. Morphology and systematics of the Cycadales. Mem. New York Bot. Gard. 57: 8–55.

SWOFFORD, D. L. 1991. PAUP: Phylogenetic Analysis

Volume 81, Number 3
1994

Doyle et al.
Integration of Morphological and rRNA Data

445

Using Parsimony, Version 3.0. Illinois Natural History Survey, Champaign, Illinois.

———— & W. P. MADDISON. 1987. Reconstructing ancestral states under Wagner parsimony. Math. Biosci. 87: 199–229.

TAKHTAJAN, A. L. 1969. Flowering Plants: Origin and Dispersal. Smithsonian Institution, Washington, D.C.

TAYLOR, D. W. & L. J. HICKEY. 1990. An Aptian plant with attached leaves and flowers: Implications for angiosperm origin. Science 247: 702–704.

———— & ————. 1992. Phylogenetic evidence for the herbaceous origin of angiosperms. Pl. Syst. Evol. 180: 137–156.

TROITSKY, A. V., YU. F. MELEKHOVETS, G. M. RAKHI- MOVA, V. K. BOBROVA, K. M. VALIEJO-ROMAN & A. S. ANTONOV. 1991. Angiosperm origin and early stages of seed plant evolution deduced from rRNA sequence comparisons. J. Molec. Evol. 32: 253–261.

VAN CAMPO, M. & B. LUGARDON. 1973. Structure grenue infratectale de l'ectexine des pollens de quelques Gymnospermes et Angiospermes. Pollen & Spores 15: 171–187.

VAN KONIJNENBURG-VAN CITTERT, J. H. A. 1992. An enigmatic Liassic microsporophyll, yielding Ephedripites pollen. Rev. Palaeobot. Palynol. 71: 239–254.

VINK, W. 1988. Taxonomy in Winteraceae. Taxon 37: 691–698.

VOGEL, E. F. DE. 1980. Seedlings of Dicotyledons. Centre for Agricultural Publishing and Documentation, Wageningen.

WAKE, D. B. 1991. Homoplasy: The result of natural selection, or evidence of design limitations? Amer. Naturalist 138: 543–567.

WALKER, J. W. 1976. Evolutionary significance of the exine in the pollen of primitive angiosperms. Pp. 1112–1137 in I. K. Ferguson & J. Muller (editors), The Evolutionary Significance of the Exine. Academic Press, London.

WHEELER, W. C. 1990. Nucleic acid sequence phylogeny and random outgroups. Cladistics 6: 363–367.

ZAVADA, M. 1984. Angiosperm origins and evolution based on dispersed fossil pollen ultrastructure. Ann. Missouri Bot. Gard. 71: 444–463.

ZIMMER, E. A., R. K. HAMBY, M. L. ARNOLD, D. A. LEBLANC & E. C. THERIOT. 1989. Ribosomal RNA phylogenies and flowering plant evolution. Pp. 205–214 in B. Fernholm, K. Bremer & H. Jörnvall (editors), The Hierarchy of Life. Elsevier, Amsterdam.

APPENDIX.

MORPHOLOGICAL CHARACTERS. Character definitions and justification follow Doyle & Donoghue (1992) for characters 1–50 (except 48) and Donoghue & Doyle (1989a) for characters 57–82, unless otherwise indicated. Where the nine-angiosperm (fossil and extant) and 12-angiosperm (extant) data sets differ in definition of characters, definitions for the former analysis are given first.

1. Branching (0) apical, (1) axillary.
2. Axillary buds (0) single, (1) multiple.
3. Phyllotaxy (0) spiral, (1) opposite or whorled, (2) distichous.
4. Leaves (0) all dichotomous, (1) linear or dichotomous plus cataphylls, (2) simple pinnate plus cataphylls,

(3) pinnately compound plus cataphylls, (4) palmately veined (actino- or acrodromous) plus cataphylls; extant analysis: (0) simple pinnate, (1) linear or dichotomous, (2) palmately veined. States 0 and 1 of Donoghue & Doyle (1989a) (elliptical or obovate, secondary veins at constant angle or lower angle at base, vs. ovate, basal secondaries crowded, at higher angle) are combined under simple pinnate; the only taxa with Donoghue & Doyle's state 1 are Austrobaileya and Calycanthaceae, only one of which appears in each of the present data sets.
5. Rachis (0) bifurcate, (1) simple.
6. Laminar venation (0) open, (1) reticulate.
7. Laminar vein orders (0) one, (1) two or more.
8. Guard cell poles (0) raised, (1) level with aperture.
9. Stomates (0) anomocytic, (1) mostly paracytic, (2) laterocytic or variable, (3) tetracytic. Core Laurales were scored as unknown in Donoghue & Doyle (1989a), but the basal state with the ingroup relationships assumed here (Fig. 8) is paracytic.
10. Apical meristem (0) without tunica, (1) with tunica.
11. Stele (0) protostele, (1) eustele with external secondary xylem only, (2) eustele with internal secondary xylem.
12. Primary xylem (0) mesarch, (1) endarch.
13. Nodes (0) unilacunar, one-trace, (1) multilacunar (more than three traces from separate primary xylem bundles, arcuate in petiole), (2) unilacunar, two-trace, (3) trilacunar. The medullosan condition (many traces derived from one solid mass or several arcs of primary xylem, scattered in petiole) was treated as a separate state in Doyle & Donoghue (1992), but because it is uninformative we have eliminated the state and rescored medullosans as unknown; this change should have no effect on the results. Trochodendrales are scored as trilacunar because Trochodendron is polymorphic but Tetracentron is trilacunar (Cronquist, 1981). Core Laurales have various numbers of traces, but these are usually formed by the splitting of two traces (Money et al., 1950; Beck et al., 1982), so we interpret the group as basically two-trace.
14. Primary xylem (0) with scalariform pitting in the metaxylem, (1) with no scalariform pitting (coniferopsid type).
15. Secondary xylem (0) with circular bordered pitting or perforations only, (1) with at least some scalariform pitting or perforations. Scoring of angiosperms based on Metcalfe (1987).
16. Vessels (0) absent, (1) present. Donoghue & Doyle (1989a) treated vessels in the roots only as a third state, but in the present data sets this occurs only in monocots. To preserve the unordered nature of the previous character, monocots could be scored as unknown, but this would obscure the similarity between monocots and other groups in ability to produce the vessel cell type. To preserve this information, we have redefined the character to express this ability and scored monocots as 1.
17. Rays (0) uniseriate or biseriate, (1) at least some multiseriate.
18. Cortical secretory structures (0) absent, (1) cavities, (2) canals.
19. Lignin with (0) no Mäule reaction, (0) Mäule reaction (Gibbs, 1957).
20. (0) Micro- and megasporophylls pinnately organized, (1) microsporophylls pinnately organized, megasporophylls simple, (2) micro- and megasporophylls simple; extant analysis: sporophylls (0) pinnately organized, (2) simple. Chloranthaceae, core Laurales, and Piperaceae

TABLE 1. Nine-angiosperm matrix, extant and fossil taxa. PROG = "progymnosperms" (*Aneurophyton*, *Archaeopteris*); ELKI = *Elkinsia* of Serbet & Rothwell (1992) = Devonian "seed fern" of Doyle & Donoghue (1992); MEDU = Medullosaceae; CALL = *Callistophyton*; CORD = Cordaitales; CONI = Coniferales; GINK = Ginkgoales; CORY = Corystospermaceae; PELT = *Peltaspermum*; CYCA = Cycadales; GLOS = Glossopteridales; CAYT = *Caytonia*; BENN = Bennettitales; PENT = *Pentoxylon*; EPHE = *Ephedra*; WELW = *Welwitschia*; GNET = *Gnetum*; MAGN = "core" Magnoliales (Magnoliaceae, *Degeneria*, Myristicaceae, Annonaceae); AUST = *Austrobaileya*; CHLO = Chloranthaceae; WINT = Winteraceae; EUDI = eudicots (Ranunculidae, *Nelumbo*, Trochodendrales, Hamamelidales); ARIS = Aristolochiaceae; PIPE = Piperales (Piperaceae, Saururaceae); NYMP = Nymphaeales (Nymphaeaceae, Cabombaceae); MONO = monocots. ? = character state unknown; − = character not included; A = 0/1; B = 0/2; C = 1/2; D = 1/3; E = 2/3.

```
                1         2         3         4         5         6         7         8
      1234567890123456789012345678901234567890123456789012345678901234567890123456789012
PROG  0?00?00???00000000???00?00000?00??000?0000?00????00000??00??-00?????10?00???????
ELKI  ??0300000?00000010?0000?010010000?000??000?000????0000??00??-00?????10?0?????????
MEDU  100300000?00?00012?0010?010011101?000?0000?0?0???00000?000??-00????00?00????????
CALL  1003100?0?10000011?00100110021100?111010?00??0????00000??00??-00?????A0000???????
CORD  1001?0000?11000001?2010?000121 10??001010?000?0????00000?000??-00????A0?00????????
CONI  1001?000001101000202010?100?201000011010100000000100000000000-00????A0?00??????1
GINK  1001?000001121000102010??000211000110?101000000000000000000-00???010000???????1
CORY  ??0300000?20100001?0020110002?1??0111110?1?????????000???00??-00?????A0000???????
PELT  ???3100?0??????????0010010002?1??0110??0?0?????????000??????-00????01000?????????
CYCA  ??02?0000011100012000102?1000?11010110?1010000000000000000000-00???010000???????
GLOS  1002?10?0?1??0000??00?00000020100111171?0??0?????000??700??-00????A0000?????????
CAYT  1003?1010???????????00200?100201001111110?1??????????000??????-00????10000???????
BENN  1002?0001?11001012?103010110?0110?110?110?2011???0????0000?000??-00????010000???????
PENT  1002?0000?2?201010?103010??0201001110?20?1???????????000???00??-00????010000???????
EPHE  1111?0000111210110120111?01112011000210?211101000010100000000000-00???010?00??????0
WELW  1111?11010111210111212111?0111201100110?21?11111111100000000100-00????010000????????
GNET  1112?110111111?11012111?0?1110110002?0?2??1111111110000000?00-00????120?10?????????1
MAGN  1022?1111111D01110100201?100?0?001110?20122100110111111100011-00010000000100A0000?
AUST  1012?111211120111010 0201?100?0?001110?30122100110111111100001-?0000011110200010000?
CHLO  1012?1112111201110101?0301?1?1?0?001110?30122100110?1111110?001-?12??0111112011?0000
WINT  1002?111111130101010 0201?100?0?0011?0?30122100110111111100001-0000001110020011 1001
EUDI  1004?111B111301?10100201?10A?0?0013?0?30122100110111111100000-010A011111020001A000
ARIS  1024?1110111301110100201?100?0?001110?30122100110?111111??121-?0111111100?10010000
PIPE  1024?11131111011010100301?101?0?001110?301221A011011111110100 1-102?111201A200B10111
NYMP  1004?111011130?0?0?00201?100?0?001110?E01221001100111111111000-?0A??100000200D11011
MONO  10???111?11110?1?0100201?100?0?001110?30122100110111111111020-?0111101?00100???00?
```

are scored as unknown to allow equivalence of their uniovulate carpels with the condition in Gnetales; however, Piperales as a taxon in the nine-angiosperm data set are scored 0, because Saururaceae have multiovulate carpels.

21. Ovule position (0) appendicular, (1) terminal.

22. Cupule (0) radial, lobed, (1) absent, (2) anatropous, (3) orthotropous unlobed; extant analysis: (0) absent, (1) anatropous, (2) orthotropous.

23. Outer integument derived from two appendages (0) absent, (1) present.

24. Ovules per anatropous cupule or potential homolog (0) several, (1) one.

25. Microsporangia (0) terminal, marginal, or adaxial, (1) abaxial. Microsporangia vary from abaxial to adaxial in angiosperms, but we have scored them as unknown because of the highly modified nature of angiosperm stamens. It is problematical whether the different positions of the pollen sacs of angiosperms, which are unique in being fused lengthwise to the sporophyll, can be equated with conditions recognized in other seed plants.

26. Microsporangia (0) free, (1) fused at least basally.

27. Microsporophylls (0) spiral or in more than one whorl, (1) in a single whorl. Chloranthaceae are scored as unknown to allow equivalence of the three-lobed androecium of *Chloranthus* with the whorled microsporophylls of Bennettitales and Gnetales.

28. Strobili (0) lacking or simple, (1) compound. Of states in the inflorescence character of Donoghue & Doyle (1989a), spikes and racemes are scored as potentially homologous with compound strobili, solitary flowers with simple strobili; cymes are not represented, except in association with solitary flowers.

29. Seeds (0) absent, (1) radiospermic, (2) platyspermic. Omitted in the extant data set (*Gnetum* is the only unequivocally radiospermic taxon).

30. Integument (0) simple, (1) with sclerotesta and sarcotesta.

31. Megasporangium with (0) lagenostome, (1) simple pollen chamber.

32. Micropyle (0) normal, (1) tubular. Angiosperms were scored as unknown in Doyle & Donoghue (1992), because their ovules are so reduced, but we have rescored

Volume 81, Number 3
1994

Doyle et al.
Integration of Morphological and rRNA Data

447

TABLE 2. Morphological matrix, extant taxa. CYCA = Cycadales; GINK = *Ginkgo*; CONI = Coniferales; EPHE = *Ephedra*; WELW = *Welwitschia*; GNET = *Gnetum*; MAGN = "core" Magnoliales (Magnoliaceae, *Degeneria*, Myristicaceae, Annonaceae); WINT = Winteraceae; CHLO = Chloranthaceae; CALY = Calycanthaceae; LAUR = "core" Laurales (*Hortonia*, Monimiaceae, Atherospermataceae, Siparunaceae, *Gomortega*, Hernandiaceae, Lauraceae); SAUR = Saururaceae; PIPE = Piperaceae; ARIS = Aristolochiaceae; NYMP = Nymphaeales (Nymphaeaceae, Cabombaceae); RANU = Ranunculidae; TROC = Trochodendrales; MONO = monocots. ? = character state unknown; − = character not included; A = 0/1; B = 0/2; C = 1/2; D = 1/3; E = 2/3.

```
              1         2         3         4         5         6         7         8
     1234567890123456789012345678901234567890123456789012345678901234567890123456789012
CYCA -?00-0-000--10001201000-1000-1-0-00---00-00000000000000000000000????010000????????
GINK -001-0-000--21000100000-1000-1-0-00---00-00000000000000000000000????010000???????1
CONI -001-0-000--01000200000-100?-0-0-00---00-00100000100000000000000?????A0000???????1
EPHE -111-0-001--21011010101-0111-0-1-01---11-10100010100000000000000????010?00???????0
WELW -111-1-010--21011210101-0111-0-1-00---11-11111111100000000100000????010000????????
GNET -110-1-011--11?11010101-0?11-0-1-01---1?-11111111100000000?00000????120?10???????1
MAGN -020-1-111--D0111011010-?100-0-0-10---10-22100110111111100011?00010000000100A0000?
WINT -000-1-111--30101011010-?100-0-0-10---20-22100110111111100001?00000011100200111001
CHLO -010-1-121--2011101?020-?1?1-0-0-10---20-22100110?1111110?001??12??0111112011?0000
CALY -010-1-111--20111011010-?100-0-0-1?---20-221001?0?11111?001110000000100?02103000??
LAUR -01B-1-111--2011101?010-?100-0-0-11---10-22100110?1111110001110?0001110?12?12?000?
SAUR -022-1-131--10?1?011020-?101-0-0-10---20-22100110?11111010010102?1112010200010111
PIPE -0?2-1-131--1011101?020-?101-0-0-10---?0-2211011011111?10??01?102?1112011?0?210111
ARIS -022-1-101--30111011010-?100-0-0-10---20-22100110?111111??121??0111111100?10010000
NYMP -002-1-101--30?0?0?1010-?100-0-0-10---C0-22100110011111110000?0A??100000200D11011
RANU -002-1-101--D0111011010-?100-0-0-12---20-22100110111111110?00010101??11?1?200??1000
TROC -002-1-121--30101011010-?101-0-0-12---20-22100110?111110000000?1?00111110200010?0?
MONO -0??-1-1?1--10?1?011010-?10?-0-0-10---20-22100110?111111110200?0111101?00100???00?
```

them as normal, since they certainly have nothing closely comparable to the gnetalian–bennettitalian state.

33. Nucellus (0) not vascularized, (1) vascularized.

34. Nucellar cuticle (0) thin, (1) thick; extant analysis: redefined as (inner) integument (0) fused most of the way to the nucellus and (1) free nearly to the base, which seems to have the same distribution but is better documented in extant taxa.

35. Pollen with (0) tetrad scar, (1) sulcus, (2) no aperture, (3) three colpi; extant analysis: (0) sulcus, (1) no aperture, (2) three colpi. Conifers are scored as having a tetrad scar in the nine-angiosperm data set, based on primitive fossil representatives, but as sulcate in the extant data set. The disulculate pollen of Calycanthaceae is scored as unknown to allow derivation from either sulcate or inaperturate; they were scored as state 1 in Donoghue & Doyle (1989a), but this state included sulcate, inaperturate, and sulculate.

36. Pollen symmetry (0) radial, (1) bilateral. As in Doyle & Donoghue (1992), we score *Gnetum* as unknown, because its pollen has global rather than radial or bilateral symmetry. We also score angiosperms with radial pollen as unknown, not only because their symmetry is unlikely to be homologous with radial symmetry in the sporelike pollen of primitive seed plants, but also because it is correlated with other characters included in the matrix: globose shape in Winteraceae, three colpi in eudicots. Uninformative in the extant data set.

37. Pollen (0) nonsaccate or subsaccate, (1) saccate.

38. Saccus structure (0) eusaccate (alveolae detached from nexine), (1) protosaccate (alveolae continuous from tectum to nexine).

39. Exine structure (0) spongy alveolar, (1) honeycomb alveolar, (2) granular, (3) columellar; extant analysis: (0) alveolar, (1) granular, (2) columellar. Unordered for reasons discussed in text. Nymphaeales are scored as uncertain because Osborn et al. (1991) reported that Cabombaceae are columellar.

40. Exine striations (0) absent, (1) present. Doyle & Donoghue (1992) scored *Gnetum* as unknown, on the grounds that its tectum is so reduced that any striations would have been lost; this is now supported by ultrastructural observations that the tectal spines of *Gnetum* resemble striations of *Ephedra* and *Welwitschia* (Gillespie & Nowicke, 1992).

41. Megaspore tetrad (0) tetrahedral, (1) linear.

42. Megaspore wall (0) thick, (1) thin, (2) absent (ordered). Ordering of this character is in keeping with the treatment of thin and absent as one state in Doyle & Donoghue (1986, 1992).

43. Microgametophyte with (0) more than four nuclei, (1) four nuclei, (2) three nuclei. In Doyle & Donoghue (1986, 1992), the three-nucleate state of angiosperms was omitted and angiosperms were scored as unknown (to allow derivation from either state), but this state is a potential synapomorphy of angiosperms in the present data sets.

44. Sperm transfer (0) zooidogamous, (1) siphonogamous. Conifers are scored as zooidogamous in the nine-angiosperm data set, based on lack of a sulcus in primitive fossil representatives (cf. Doyle & Donoghue, 1992), but siphonogamous in the extant data set.

45. Megagametophyte (0) monosporic, (1) tetrasporic.

46. Egg (0) cellular, (1) free-nuclear.

47. Early embryogenesis (0) free-nuclear, (1) cellular.

48. Fertilization (0) single, (1) double. See discussion in text. Calycanthaceae lack double fertilization (Loconte & Stevenson, 1991), but because this is associated with apomixis (Davis, 1966) we score them as unknown.

49. Embryo (0) without feeder, (1) with feeder.

50. Seed germination (0) hypogeal, (1) epigeal. Angiosperm data from de Vogel (1980) and Endress (1983); contrary to Loconte & Stevenson (1991), Endress reports that germination of *Austrobaileya* is epigeal.

51. Companion cells in phloem (0) absent, (1) present.

52. Microsporangia (0) various, (1) in two pairs.

53. Endothecium (0) absent, (1) present.

54. Pollen germination (0) in pollen chamber, (1) on stigma.

55. Megagametophyte (0) large, (1) eight-nucleate. The larger tetrasporic embryo sacs of Piperaceae are conservatively scored as unknown, although they show similarities with the eight-nucleate type.

56. Endosperm (0) absent, (1) present.

57. Radicle (0) persistent, (1) replaced by adventitious roots.

58. Habit (0) woody, (1) herbaceous. Groups with interfascicular cambium not producing normal secondary xylem are scored as unknown.

59. End-wall pits or vessel perforations (0) multiple, (1) simple.

60. Sieve-tube plastids (0) starch, (1) PI type, (2) PII type. See text for addition of PII type, scoring of Magnoliales (Behnke, 1988). Core Laurales were scored as unknown in Donoghue & Doyle (1989a), but the basic state with the ingroup arrangement assumed here is PI.

61. Oil cells (0) absent, (1) present.

62. Benzylisoquinoline alkaloids (0) absent, (1) present. Uninformative in the nine-angiosperm data set.

63. Stipules (0) absent, (1) adnate-axillary. Other stipule types are scored as unknown.

64. Chloranthoid teeth on leaf margins (0) absent, (1) present.

65. Perianth (0) more than two whorls (or spiral or chaotic), (1) two whorls, (2) absent. Donoghue & Doyle (1989a) scored Trochodendrales as having one whorl, but because this state is autapomorphic in the present data set we have rescored the group as unknown.

66. Perianth symmetry (0) various, (1) at least calyx trimerous.

67. Stamen number (0) various, (1) multiples of three.

68. Stamens (0) laminar, (1) with well-differentiated filament.

69. Pollen (0) boat-shaped, (1) globose. Groups with saccate and sporelike pollen scored as unknown because it is unclear whether their shape conditions can be compared with those in nonsaccate, basically monosulcate groups.

70. Pollen size (0) large (> 50 μm), (1) medium, (2) small (< 20 μm) (ordered). Previously this character was used only in the angiosperm study (Donoghue & Doyle, 1989a). We have scored taxa with saccate pollen (except *Caytonia*) as uncertain (0/1), for two reasons. First, saccate pollen tends to be conspicuously larger and more massive than pollen of nonsaccate groups, suggesting that there is a functional correlation between presence or absence of sacs and size. If so, scoring saccate groups on the angiosperm-based size scale of Donoghue & Doyle might excessively weight transitions between saccate and nonsaccate and produce spurious groupings among non-angiospermous taxa. Second, it is possible that the more appropriate comparison is between size of nonsaccate pollen grains and size of the central body (nexine) in saccate pollen, which is often much smaller than total grain size.

71. Tectum (0) continuous or finely perforate, (1) foveolate-reticulate. We have modified the definition of state 1 from semitectate-reticulate in Donoghue & Doyle (1989a) and rescored *Austrobaileya* as having this state, since its relatively large foveolae and rounded muri seem more comparable to the sculpture of Chloranthaceae and other reticulate groups than the much more finely perforate sculpture of groups such as Magnoliales.

72. Aperture membrane (0) (nearly) smooth, (1) sculptured. Conifers are scored as unknown in the nine-angiosperm data set because a sulcus is absent in primitive forms. Scoring of Aristolochiaceae is based on SEM photos of *Saruma* pollen kindly provided by Long Huo (Guangzhou); scoring of Piperales is based on Bornstein (1989) and Doyle & Hotton (1991).

73. Supratectal spinules (0) absent, (1) present. Scoring of Piperales is based on Bornstein (1989) and Doyle & Hotton (1991).

74. Endexine (0) thick, laminated, (1) absent, (2) thin, nonlaminated, except under apertures. Unordered for reasons discussed in the text. Following Chlonova & Surova (1988), we have rescored Chloranthaceae as having endexine, not unknown as in Donoghue & Doyle (1989a).

75. Hypanthium (0) absent, (1) present.

76. Ovules per carpel (0) several, (1) one apical. A third state in Donoghue & Doyle (1989a), one basal, is represented only in Piperaceae, which are therefore scored as unknown.

77. Fruit (0) dehiscent, (1) berry, (2) drupe (with endocarp), (3) dry indehiscent. As noted by Loconte & Stevenson (1991), the spongy, indehiscent or irregularly dehiscent fruits of Nymphaeaceae (Cronquist, 1981), scored as dehiscent in Donoghue & Doyle (1989a), are better characterized as berries. Winteraceae were scored as unknown in Donoghue & Doyle (1989a), for either dehiscent or berries, but because the dehiscent condition is restricted to *Takhtajania*, which appears to be nested within the family (Vink, 1988), we have rescored them as having berries.

78. Testa (0) multiplicative, (1) non-multiplicative.

79. Exotesta (0) normal, (1) palisade.

80. Tegmen (0) normal, (1) sclerotic.

81. Nutritive tissue (0) endosperm only, (1) endosperm plus perisperm. Megagametophyte tissue might be considered a third state, but this would be redundant with endosperm formation (character 56).

82. Chromosome number (0) $n = 6-8$, (1) $n = 12-16$. Donoghue & Doyle (1989a) defined state 1 as $n = 12-19$, but because numbers above $n = 12$ are of uncertain and potentially complex origin, we have redefined the states and scored taxa with $n > 16$ as unknown. A third state in Donoghue & Doyle (1989a), $n = 10-11$, is represented only by Calycanthaceae, which are therefore scored as unknown. Data on non-angiospermous groups from Ehrendorfer (1976).

RIBOSOMAL RNA CHARACTERS. 1–86: Characters from the 18S subunit, in terms of positions in soy.

1: 93; 2: 120; 3: 131; 4: nucleotide not present in soy; 5: 132; 6: 134; 7: 176; 8: 181; 9: 189; 10: 190;

11: nucleotide not present in soy (between 191–192); 12: 200; 13: 220; 14: 222; 15: 236; 16: 239; 17: 240; 18: 241; 19: 242; 20: 244;

TABLE 3. Ribosomal RNA matrix. Order of taxa as in morphological matrix. R = A/G; Y = C/T; M = A/C;
K = G/T; S = C/G; W = A/T; H = A/C/T; B = C/G/T; V = A/C/G; D = A/G/T; N = A/C/G/T.

```
         1         2         3         4         5         6         7         8
1234567890123456789012345678901234567890123456789012345678901234567890123456789012345 67
ACCTCGCTTTTSCGCGTBDTCGWAYTGCTMCASGATAGCYGACACCATATTGCCCRGTTTTTGCCTGTGCGTAATACATTAGCTTAG
ACCCTACTTC?CCGCCTGCTCGTACCGCTGCACGATA??TGATACCGT?TTGCTCGTTCGTCGCCT??????AATAAATTAGCTTA?
GYCYWRCTCTTTCGCCTWCCCGWACCRYTTTASGATAGCCGACACCCTGGTGCCCGTTCCTCGCTT??GCCGAATACRTTAGYTTAG
GCCCTAGTTC?TCGCCTTGTCGAATCGTCACWYKRCWGTCGATTCCCCGTCG?CTCATTCCGGATTGCGTGGTACGGGCCAACTTGG
GCCCTWGTTC?TCATCGTGTTGAGCCGTCACACTGCA?AGGGTTTCCCGGTGTCTCGCAGTTGATT??????TACGCGCCAATTTGG
?CCCTAGTCT?TCGTTTTTTCGAGG?ACTTCAGTGCA?TCGACTTCCCGGTGTCTGACGCTCGACT??????A?GGGTCAATTTGG
AGACYGACTCGTYGCTCGCCTATATCGCTCYWKGACTGCCGATTCCATATCGCYYGTGACCCACTCTC?GGTAGTGCGTTGGCCCAT
AGT?CAACTC?TTACTTGCCTAAACCGCTC?WYKRCWGACAATTCTATATCGYTTGTCACCCACTCTCTKRTAGTGCGTTGGCCCAY
AGT?CGACTC?TTGCT?GCCTATACCGCTTCTCGACT?TCGATTCCATATCGCTTGT?ACCCACTC?????TAGTGCGTTGGCCTA?
?GTCT?CCTC?TTGCTCGCCTATACCGTCTCTTGACTGACGATTCCATATCGCTTGCGACCCACTC??T?GTAGTGCGTTGGCCCAT
AGT?CTACTC?TTGCTCGCCCATAYYGCTCCGTGACTTACGATTCCATATCGCTTGTGAYCCACTS???TGTAGTGCGTTGGCCCA?
AGC?TGCCTC?TTACTTGATT?TAC?GTCT?TCGACATACGGTTCCATATCTCCCGATCCCCACTCTCGTTTAGTGCGTTGGCCCAT
AGT?TRCCTCGTTAYTYKMMYATATCGTTTTKCKACW?ACGATTCCATATCTCCTGYCWCCCACTY?TGTTTAGTGCGTTGGCYYAC
AGC?TGMCTC?YTACTTGYYTATACTGYTTCGCRACT?ACGATTCCATAKCGCTTGTGACCCACTCTCYGATAGTGCGTTGGCCCAT
AGC?TRCCTC?CCRCATGTTTGTACCGYTYCWCAACTGATGAYTCCATATCGCCTGWMCCCRCTC??GGGAAGTGMGTTGGCTTAK
AGT?AGACTC?TTATTTGCCTATA???????TTGACTGCCGATTCCMTATCGGTTGTACCCACTCTCTGATAGTGAGTTGACCTAT
AGT?CGACTC?TTGCTTGCCCATACCAC??TATGACT?CCAATTCTATATCGCCTGTAACCTACTCTCTGATAGTGCGTTGGCCCAT
AKCKSGMCTC?YYGCNTGMYBAYAYHGYTCCGCGACTGACGAYWCCATAYCGCTCGYRWCCCGCTCTCKGGTARTRCGTTGGYCCAG


         9         0         1         2         3         4         5         6         7
890123456789012345678901234567890123456789012345678901234567890123456789012345678901234
TCGCTCTATATTAG?CCCCCRATCGCCTATTWACCTCGCGYCTATGYCTGGCRYTCTWYRTWCGYSSGGGTTAGGGTGATGGGAGAGR
??????TATACCAG?CCGCAATCGCCTATTAACCCCGTGTTTATGCCCGGCGCTCTACACCCGTCGCGGTTAGGGTGATAGG?G?AG
TCGTYCCATACYARCCCGCAATCGCCTRTTVACYYCGTGTCTATGCCTGGYGCYCYHCATCCGYTCGGGTTAGKGTGATAGGGGARG
?TCTTCCATGCCAACCCGCGACTGCTTATTAAC?TCGCGCCTATRCTCGCCTGCTAACGCCCGCCTCGGTTAGAGCGTCTGGGGAGA
ATCCAGTGTATTCACCCGCAACCACTTATAAATCACGTGCCTATGCCCACTTGCCTCCGTCAGSTTGGG?TTGAACGTCTGA?AGGG
ATGTAATGTACTCA??TGCGGCCACTTATCAACT?CGTGC????????????????????????GGGGGATTAAAACAGCTGG?AGAA
TCGTAATGCGCCAGTCCTCAGYYCTCGGACGGCGCTACGC???????????????????CGCGCGCCTCGGCCGGTGCGATAAGAGAGG
TCGAATTGCGCCAATTCTTGGTCGTCGGACGGTGCTGCGMACGCA?YCGTCATATACTGCGTATTTCTACCGATGYARTAAGRGAGG
?CTAATTGCGCC?A??????GGTCGTCGGACG?CGCTGCA?ACGCA?TCGCCGTATTCTGCGCGCCTCGGCCGGTGTGATAAA?GAGG
TCGAATTGCGCCAG??????????TAGGACGGCGCTGCGC????????????????????????????GGCCGGTGTGATAGG???GG
???RWYTGCGCMVG??CTCGGTCGYCGGACGGCGCTGCGC???????????????????CGCGCGCCTCGGCCGGTGTGATAAGAGAGG
TTGTAATGCGCC?ATCCTCAGTCGTCGGACKTCGCTGCGC?????????????????????????????????????????AGGGAGG
TTGTATTACGCCAATCCTCAGCTGTCGGWCGGMGCTGCGCTCGCA?CCGCCGCATTCCGCGCGCMTCSGCCGRTGCRATAAGGGAGG
TTGAATTACGCMAA??TTCGGTCGTCGGACGGCGCTGCGC?????CCGCCGTATT?CGCGCGTYYCGGCCGGTGTRATAAGGGAGG
TCGTAGTGCGTCAR?TYTCGGTTGTCGGACGRCSYYAYAYWCGCATTCGCCGYATTYCRCGCAYTYVTSCCGATGCRATAAGGGAGG
TCGATTTACGCCAATCTTCGGTCGTCGGACGGCGCTGCGCACGCA?TCATTGCATTCCGTGTGCTTCGGCCGGTGTGATAAGGGAGG
TCGAAATGC?CAAGGCTTTGGGTGTCGAACGACGCTGCGC???????????????????GCGCGTT?CGGCCGGTGTGATAAGGGAGG
TYGHWGTGCRCCWAYCCTCGGYCGYCGGACGGCGCYGCGCHCGSG?CCGCYGYATTYYGCGCGTYKCGGCCGGTGYGATARGRRARG
```