1990

Ribosomal RNA and the Early Evolution of Flowering Plants.

Robert Keith Hamby
Louisiana State University and Agricultural & Mechanical College

Recommended Citation

Hamby, Robert Keith, "Ribosomal RNA and the Early Evolution of Flowering Plants." (1990). *LSU Historical Dissertations and Theses*. 5053.
https://repository.lsu.edu/gradschool_disstheses/5053

U·M·I

University Microfilms International
A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
313/761-4700 800/521-0600

Order Number 9123196

Ribosomal RNA and the early evolution of flowering plants

Hamby, Robert Keith, Ph.D.

The Louisiana State University and Agricultural and Mechanical Col., 1990

U·M·I
300 N. Zeeb Rd.
Ann Arbor, MI 48106

RIBOSOMAL RNA AND THE EARLY EVOLUTION OF FLOWERING PLANTS

A Dissertation

Submitted to the Graduate Faculty of the
Louisiana State University and
Agricultural and Mechanical College
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

in

The Department of Biochemistry

by
Robert Keith Hamby
B.S.Ch.E., Auburn University, 1978
M.S., Virginia Polytechnic Institute and State University, 1984
December 1990

ACKNOWLEDGEMENTS

There are many people to whom I am indebted for their help, encouragement, support and friendship through the many years of graduate school which are now ended. As this was the absolute last thing I added to this dissertation, I almost certainly omitted the names of some people. I ask everyone's indulgence.

I would like to thank Dr. Elizabeth Zimmer, my major professor, for the direction she provided in the completion of this project and her friendship and help in the transition from the world of engineering to the world of molecular biology. Dr. Zimmer has also provided generous support for research and travel over the past five years, and for that I am also grateful.

I want to thank the people who have collaborated and conspired with me, especially the past and present members of the Zimmer lab including Joey Spatafora, Eldon Jupe, Lynne Sims, Mike Arnold, Julie Dowd, Kelly Mullen, Gretchen Stein, Laurie Issel, Vishal Sachdev, Bryan Peavy, Jay DeSalvo, Dianne Dennis, Mary Bowen, Monique LeBlanc and all the other people who have worked in our lab. I wish to thank Russ Chapman and the people from his lab, including Debra Waters, Rick Zechman, Tom Kantz and Mark Buchheim, for their encouragement and comraderie.

I am indebted to Sue Bartlett and the members of her lab for their generous help with my cloning experiments and to Sue in general for her friendship and support. I want to express my gratitude to the other members

of my committee, all of whom have been very helpful: Kathy Morden, Martin Hjortso, Andy Deutsch, Simon Chang and Russ Chapman.

I have many friends outside of LSU to whom I am also grateful for moral support and encouragement and understanding when I had to "go to the lab." This list includes, but is by no means limited to, Rene and Smith Jackson, Wayne Lowther, Charlotte Adcock, Jim Anding, Gary Daniels, David Motes and James and Peggy Denny.

I would like to dedicate this dissertation to my parents, Bob and Ann Hamby who gave me life, and to David Walker who taught me about living life.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Since the end of the last century, the predominant theories of the early radiation of the angiosperms have been that the earliest flowering plants were most similar to the present-day Magnoliidae (*sensu* Takhtajan, 1969). This position has been adopted by many, though there are some who suggest that the base of angiosperm radiation lies within the monocots (Burger, 1981) or a combination of monocots and dicots (Burger, 1977; Donoghue and Doyle, 1989a, 1989b). Many different ancestors to the angiosperms themselves have been proposed including, at one time or another, most of the extant gymnosperms, extinct gymnosperms and the extinct seed ferns. Morphologically-, cytologically- and phytochemically-based classifications have not provided unequivocal phylogenies of the angiosperm lineages, although recent cladistic treatments of morphological characters by Crane (1985) and Donoghue and Doyle (1989a) provide a logical framework for testing molecular genealogies. The most fundamental comparison between homologous molecules of different species is a comparison of the primary nucleotide structure. In this dissertation, I report on comparisons of the primary structure of the nuclear-encoded cytoplasmic ribosomal RNAs (rRNAs) to produce phylogenetic hypotheses for the extant angiosperms and other seed plant lineages. Computer-assisted phylogenetic analyses based on the comparisons of 1700 nucleotides from five regions of the nuclear-encoded cytoplasmic 18S rRNA and three regions of the nuclear-encoded cytoplasmic 26S rRNA from 46 angiosperm taxa, 12 gymnosperm taxa and two seedless vascular plants (as outgroups), suggest that: (1) The seed plants (gymnosperms and angiosperms) are a natural (monophyletic) group; (2) The angiosperms arose from within the gymnosperms and are a natural group; (3) The Gnetales are a coherent group with tenuous support as the sister group of the angiosperms; (4) The earliest angiosperm divergences involve the *paleoherbs* of Donoghue and Doyle (1989a, 1989b), i.e., the Piperales (Piperaceae and Saururaceae), the Nymphaeales (Nymphaeaceae, Cabombaceae, Barclayaceae, but not Ceratophyllaceae or Nelumboaceae) and the monocots; (5) Both the monocots and dicots are paraphyletic groups.

INTRODUCTION

The flowering plants (angiosperms) are the most diverse flora on the earth today with almost 300,000 species (Cronquist, 1968). Since their first appearance at least 120 million years ago, the flowering plants have become the predominant form of vegetation in the world. They exist and thrive in habitats as diverse as tropical rain forests, deserts and the Arctic tundra; some are even marine. The angiosperms have been subdivided into two (putatively) natural classes, the monocotyledons and dicotyledons, so named for the number of primordial leaves on the emerging seedling axis. The closest living relatives of the angiosperms are the gymnosperms, the other and older group of seed bearing plants. Since the late nineteenth century, the origin and the earliest radiation of the flowering plants have been studied by many investigators who have tried to identify the group from which the first angiosperms were derived and to determine the characteristics of the most primitive flowering plants. Most of these investigators have at one time or another invoked Darwin's evaluation of the situation as "an abominable mystery."

Comparisons of flower, pollen and stem morphology, cytology and phytochemistry have been used to develop classifications of the extant angiosperm taxa. Looking at a series of closely related species can suggest the direction of evolution of certain characteristics as can contrasting angiosperm traits with those of non-angiosperm plant species. This information can be used to try to assign an evolutionary progression among different groups of similar plants.

Paleopalynology, the study of fossil pollen and spores, has contributed significantly to the present-day perceptions of the direction of evolution in angiosperms. Similarly, fossil leaves and to a lesser extent fossil wood, fructifications and flowers have added more data. However, no single angiospermous ancestor has been identified in the fossil record, and in fact, the earliest unequivocal angiosperm pollen was already diversified into three or four groups including representatives of both monocots and dicots (Walker and Walker, 1984).

Despite much work in this field, the evolutionary history of the angiosperms is today still largely unsolved. In turn, each of the extant gymnosperm lineages, many of the extinct gymnosperms and the extinct seed ferns have been proposed as having given rise to the flowering plants. The dicotyledonous Magnoliales and their close allies which make up the superorder Magnoliidae (the list of allies varies with author), plants with large, showy flowers consisting of many stamens and carpels, are almost a consensus choice as the most primitive angiosperms (Cronquist, 1968; Takhtajan, 1969; Thorne, 1974). There are, however, others who suggest that perhaps the earliest angiosperms were monocots (Burger, 1981) or a group composed of monocots and certain dicots (Burger, 1977; Donoghue and Doyle, 1989a).

The identification of DNA as the means by which phenotypic traits are inherited and passed from ancestor to descendant has resulted in an explosion of available techniques for the study of evolution and, hence, systematics. The central dogma of molecular biology holds that information passes from generation to generation in the form of DNA, and that DNA can pass information on to proteins, but not vice versa (Smith, 1989). Taken in the framework of cladistic analysis (Hennig, 1965), which holds that systematic classifications should reflect the true evolutionary history of the taxa in question as opposed to grouping the taxa based on perceived similarity, biochemical comparisons of DNA and proteins offer very powerful tools for the inference of phylogenetic relationships.

Early systematic applications of biochemical methods at the molecular level employed immunological and electrophoretic techniques to measure similarities between homologous proteins from different species. Another common systematic application of biochemical methodology is to compare the primary sequences of homologous proteins of representatives of different taxa. Zuckerkandl and Pauling (1962) proposed the concept of the molecular clock based on such comparisons of the amino acid sequences of hemoglobin molecules. The clock hypothesis holds that the amino acid and nucleotide sequences of homologous proteins in different species are evolving at similar rates. The clock may be different for different genes, and is understood not to tick like a metronome, but to have periods of rapid and slow change. Sarich and Wilson (1967), using the immunological cross reactivity of albumins among higher primates, asserted a much closer relationship among human, chimp and gorilla lineages than had previously been proposed. The results were highly controversial, but led to reevaluation of fossil evidence and, with the accumulation of additional biochemical evidence, their time scale has gained acceptance. Still the clock concept is not universally accepted and may not always be a safe *a priori* assumption.

At the DNA level, digestion with restriction endonucleases followed by electrophoresis and probe hybridization identifies mutations at sites recognized by the highly specific enzymes. Differences in the patterns of digestion are convenient markers for discerning relationships among related taxa. This technique, while rapid and simple, only samples the parts of the DNA molecule which are recognized by the restriction enzymes. DNA and RNA sequencing protocols, on the other hand, allow for the elucidation of the primary structure of individual genes. Thus, the most fundamental comparison possible between homologous genes of different species is at the level of the primary nucleotide sequence.

In this dissertation, I report on comparisons of the primary sequence of nucleotides from homologous regions of the nuclear-encoded ribosomal RNA (rRNA) from many different extant plant species and the inferred early history of the angiosperms. Ribosomal RNA was chosen as the molecular "yardstick" with which all species were compared because all living organisms possess rRNA, an essential component of cellular protein synthesis. Ribosomal RNA was present in the earliest forms of life and is in fact older than the plant kingdom itself. The ubiquity of rRNA throughout living organisms and the development of techniques for the rapid determination of the primary nucleotide sequence of rRNA molecules make rRNA a useful source of characters for inferring evolutionary relationships.

Comparisons based on molecular sequence have several potential advantages over morphological comparisons. One is the ability to minimize subjective interpretations; determination of molecular sequences is objective and can even be automated. Another advantage of DNA characters is that they can be understood at a primary genetic level when assigning homology. In evolutionary studies, *homology* means more than just similarity; homology implies descent from a common ancestor. For example, in trying to determine the progenitor of angiosperms, much work has focused on trying to find within the gymnosperms and seed ferns the homologs to the bitegmic ovule and to the enclosed carpel found in all angiosperms. Different interpretations made in the absence of knowledge of the genetic contributions to such characters can and do lead to different conclusions.

In the chapters that follow, I first describe the function and structure of ribosomal RNA and the evolution of the rDNA locus. I then discuss the cladistic method of inferring evolutionary relationships using specific examples and briefly compare cladistics to the phenetic method. In the Materials and Methods section, I present the experimental methodology and the basics of the data management and analysis. In the Results section, I first report on the use of rRNA sequences to determine intrafamilial relationships within the grass family (Poaceae) and then on the trees inferred from comparing rRNA sequences from 60 extant plant species including 46 angiosperms, 12 gymnosperms and two seedless vascular plants.

LITERATURE REVIEW

RIBOSOMAL RNA

Introduction. All living organisms have within their genome DNA sequences which code for ribosomal RNAs (rRNAs), essential components of cellular protein synthesis. In plants, ribosomal DNA (rDNA) is found in nuclear, mitochondrial and chloroplast genomes. The ubiquity of rRNA throughout nature and the development of techniques for the rapid determination of the primary nucleotide sequence of rRNA molecules make rRNA a good tool for inferring evolutionary relationships. Not all regions of the rDNA are evolving at the same rate, so while some regions are useful for comparisons at or below the genus level, other regions are only useful at the family level or above.

Until recently, the greatest use of rRNA sequences had been in the investigations of bacterial evolution. Woese (1987) used a parsimony analysis (see below) of complete 16S rRNA sequences to propose three main lines of descent in nature: eubacteria, archaebacteria and eukaryotes. Other analyses of the same data support the archaebacteria tree (Gouy and Li, 1989a). Lake (1988) disputes this interpretation of the rRNA sequence data suggesting that the archaebacteria are paraphyletic.

Aside from work in this laboratory (Hamby and Zimmer, 1988, 1991; Zimmer *et al.*, 1989; Knaak *et al.*, 1990) and that of our collaborators

(Chapman and Avery, 1989; Buchheim *et al.*, 1990; Kantz *et al.*, 1990; Zechman *et al.*, 1990), there has been little use of comparative rRNA sequences in plant evolutionary studies. Nickrent and Franchina (1989) are using nuclear 18S rRNA sequences to define the relationships within the parasitic flowering plants of the order Santalales. Wolfe and coworkers (1989) have compared published 18S and 26S sequences to calibrate the divergence of monocot and dicot lineages.

Table 1 contains a partial list of investigators who have used chloroplast or nuclear rRNA sequences to study taxonomic or phylogenetic relationships.

Ribosomal RNA function. The main function of the rRNAs is in protein synthesis. It was previously thought that the rRNAs served primarily as a scaffolding for the ribosomal proteins, but recent evidence suggests that rRNA molecules are the basic functional element of the ribosome and that the proteins serve to mediate interactions between messenger RNA (mRNA), transfer RNA (tRNA) and rRNA (reviewed by Gerbi, 1985 and Dahlberg, 1989). Most detailed studies of ribosome action are based on ribosomes of the bacterium *Escherichia coli*, but the results are generally valid for higher taxa as well. The 70S *E. coli* ribosome consists of a 30S subunit and a 50S subunit which come together in the presence of mRNA and other cofactors. The 16S rRNA (analogous to the plant cytoplasmic 18S rRNA) is part of the

Table 1. A partial list of investigators who have used ribosomal DNA or RNA for systematics studies.

Investigators	Groups	Subunit	Comments
Kumazaki *et al.*, 1983	Protists	Nuclear 5S	Green algae share common ancestor with vascular plants.
McCarroll *et al.*, 1983	Eukaryotes	Nuclear 18S	*Dictyostelium* represents earliest divergence of eukaryotes.
Hori *et al.*, 1985	Plants	Nuclear 5S	*Cycas* is a gymnosperm. Land plants are most closely related to charophyte algae.
Woese, 1987	Bacteria	16S	There are three primary lines of descent: archaebacteria, eubacteria and eukaryotes.
Hori and Osawa, 1987	Prokaryotes and Eukaryotes	5S and Nuclear 5S	Red algae most primitive eukaryotes. Archaebacteria and eukaryotes split off after eubacteria.
Vossbrinck *et al.*, 1987	Eukaryotes	Nuclear 18S	Microsporidia are very early divergence of eukaryotic evolution.
Lake, 1988; 1989	Bacteria	16S	Evolutionary parsimony analysis says archaebacteria are paraphyletic.
Edman *et al.*, 1988; Stringer *et al.*, 1989	Protozoa and Fungi	Nuclear 16S	*Pneumocystis carinii* is a fungus.
Field *et al.*, 1988; Raff *et al.*, 1989	Animals	Nuclear 18S	Cniderians are separate from other animal lineages. Coelomates are monophyletic.
Nairn and Ferl, 1988	Eukaryotes	Nuclear 18S	Angiosperms are monophyletic.
Gouy and Li, 1989a	Bacteria	16S and 23S	Neighbor joining and maximum parsimony analysis support Woese above.
Gouy and Li, 1989b	Eukaryotes	Nuclear 18S and 26S	Fungi diverged first from the common ancestor of plants and animals.

Table 1 (con'd).

Investigators	Groups	Subunit	Comments
Perasso *et al.*, 1989	Algae	Nuclear 26S	Rhodophytes, chromophytes and chlorophytes are each monophyletic groups. Plants are closest to chlorophytes.
Wolfe *et al.*, 1989	Angiosperms	Nuclear 18S and 26S. Chloroplast 16S.	Monocots and dicots diverged from one another 200 million years ago.
Turner *et al.*, 1989	Prokaryotes	16S	Prochlorophytes are holophyletic with cyanobacteria and chloroplasts, but not progenitors of chloroplasts.
Watanabe *et al.*, 1989	Protozoa and Fungi	Nuclear 5S	*Pneumocystis carinii* is closer to Zygomycota fungi than to ascomycota or basidiomycota.
Scheifer and Ludwig, 1989	Bacteria	23S	23S rRNA trees support the 16S rRNA trees as well as thought based on EF Tu and subunit of ATPase.
Hillis and Dixon, 1989	Vertebrates	Nuclear 28S	Coelacanths belong among the tetrapods. Weak support for a bird-mammal relationship.
Sogin *et al.*, 1989	Eukaryotes	Nuclear 18S	Earliest eukaryotes are microsporidia and diplomonads. Fungi, plants and animals diverged relatively recently.

30S subunit; the 5S and 23S (analogous to the plant cytoplasmic 26S rRNA) combine with various proteins to make up the 50S subunit. In plants and other eukaryotes, the large subunit of the ribosome also contains a 5.8S rRNA molecule.

During translation initiation, sequences near the 3' end of the 16S rRNA molecule base pair with the Shine-Delgarno sequence upstream of the initiation codon in bacterial mRNA. Interference with this base pairing interaction by mutation in the 16S rRNA molecule leads to significant reductions in the level of protein synthesis (Jacob *et al.*, 1987; Hui *et al.*, 1988). Ribosome activity can be restored by a compensatory mutation in the Shine-Delgarno sequence of the target mRNA (Hui *et al.*, 1988). Base pairing between mRNA and the same region of the 16S rRNA molecule also may be responsible for maintaining the correct reading frame during elongation (Trifonov, 1987; Weiss *et al.*, 1987; 1988). In addition, translation termination at the stop codons appears to rely upon specific RNA-RNA interactions between the 16S rRNA and mRNA (Murgola *et al.*, 1988). It should be noted that eukaryotic mRNAs do not possess a Shine-Delgarno sequence, and protein synthesis is proposed to be initiated by other means.

The proper association of the small and large subunits also is dependent to some degree on sequences within the 16S rRNA molecule though no particular sequence dependence has been identified within the 23S rRNA molecule (Dahlberg, 1989). Methylation of two consecutive adenine residues near the 3' end of 16S rRNA is required for correct association of the subunits. The stem structure immediately upstream of the stem-loop containing the methylated adenines is also important in the formation of an active ribosome as is the sequence around position 790 (of 1542 bases in the 16S molecule).

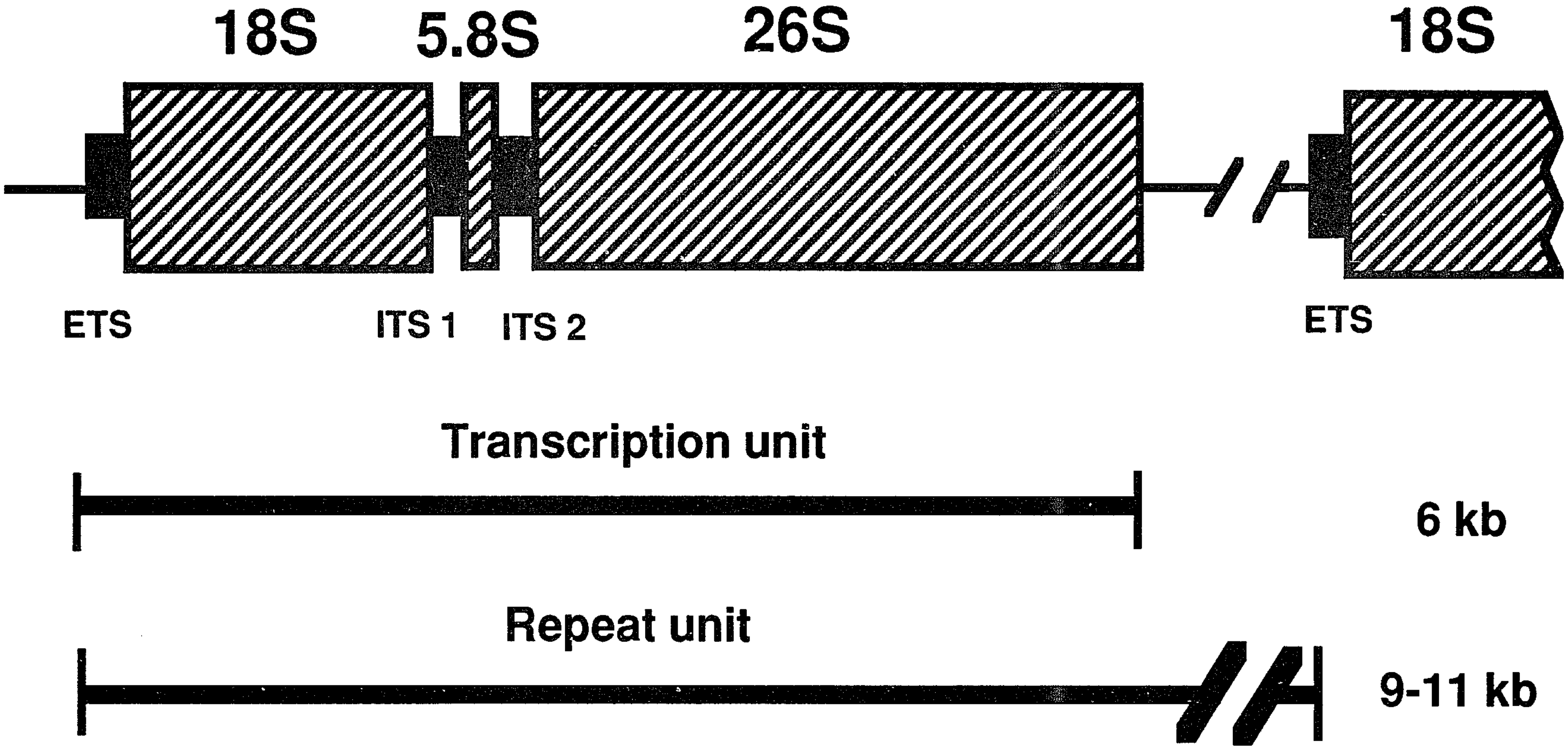
The activities within the ribosome decoding site which consists of the aminoacyl (A) site and the peptidyl (P) site are dependent on the tertiary structure of the 16S rRNA. Several different regions of the 16S rRNA secondary structure are brought together by three-dimensional folding to line the cleft of the 30S subunit which has been shown to be only a few angstroms from the codon-anticodon site. Transfer RNA protection experiments indicate that the tRNAs interact with specific 16S rRNA nucleotides in this cleft region (Noller *et al.*, 1987). Footprinting experiments have implicated specific nucleotides within the 16S rRNA as sites of action for antibiotic agents known to cause miscoding; resistance to the antibiotic is associated with modifications of the rRNA sequence (Moazed and Noller, 1987). Recently, Moazed and Noller (1989) have identified sequences within the 23S rRNA that make up parts of the A and P sites on the 50S subunit. They have also described the E site, the site where the deacylated tRNA resides before it dissociates from the ribosome completely, and have shown that the CCA conserved nucleotides at the end of all tRNA molecules interact with the 23S rRNA at the A, P and E sites.

The peptidyl transferase activity of the ribosome catalyzes the formation of the peptide bond between the growing protein and the new amino acid (Dahlberg, 1989). This activity can be significantly disrupted by base modifications in domain V of the 23S rRNA. The action of antibiotics known to inhibit transferase activity also map to this domain. Finally, specific nucleotides in the 23S rRNA have been shown to be involved with translocation of the peptidyl tRNA from the A site to the P site (Dahlberg, 1989).

Nuclear ribosomal gene organization. The nuclear genes which code for rRNA (rDNA) are reiterated thousands of times within the typical plant genome (Appels and Honeycutt, 1986). In fact, they can comprise as much as 10% of the total plant DNA (Hemleben *et al.*, 1988). Ribosomal DNA is arranged in tandem repeats in one or a few chromosomal loci. Only among closely related species are the chromosomal locations homologous.

Each repeat unit consists of a transcribed region separated from the next repeat unit by an intergenic spacer (IGS). Figure 1 shows that, beginning from the 5' end, the transcribed region consists of an external transcribed spacer (ETS), the 18S gene, an internal transcribed spacer (ITS1), the 5.8S gene, a second ITS (ITS2) and the 26S gene. Transcription by RNA polymerase I (which only transcribes rDNA) may end immediately after the 26S gene, although in some animal systems, transcription can

Figure 1. A typical plant rDNA repeat unit. The coding regions are marked by hatched boxes. The other transcribed regions are denoted by thick black lines, and the nontranscribed regions are denoted by thin black lines.



continue on through most of the IGS and end just before transcription of the next repeat unit begins (DeWinter and Moss, 1986; Labhart and Reeder, 1986). In wheat, most transcripts end at or near the 3' end of the 26S gene, but some transcription proceeds through the intergenic regions as in *Xenopus* and *Mus* (Vincentz and Flavell, 1989). Presumably, the 3' trailer is rapidly discarded to yield the precursor rRNA molecule. This 45S precursor is enzymatically cleaved and trimmed to produce the three mature rRNA molecules.

There is another cytoplasmic rRNA molecule, the 5S rRNA, which is transcribed by RNA polymerase III. In prokaryotes and some lower eukaryotes, the 5S gene is linked to the other rDNA, but in higher eukaryotes the 5S genes lie in independent unlinked arrays (Appels and Honeycutt, 1986). In maize, for example, rDNA arrays are on the short arm of chromosome 6 (McClintock, 1934; Givens and Phillips, 1976; Phillips, 1978), while the 5S rDNA repeats have been localized to the long arm of chromosome 2 (Steffensen and Patterson, 1979; Mascia *et al.*, 1981).

Evolution of the rDNA locus. The most remarkable feature of ribosomal DNA is the overall sequence homogeneity among members of the gene family. If all parts of the genome are evolving independently, comparisons of nucleotide sequences between members of the same gene family within a species would show about the same level of similarity as comparisons of the same gene between two closely related species. This is true so long as the duplication events creating the gene family preceded the divergence of the two species. Studies consistently show that this is not the case for rDNA (Arnheim, 1983). Brown and coworkers (1972) first demonstrated by hybridization experiments that, within the species *Xenopus laevis*, the several hundred rDNA repeats were essentially identical at both the coding and the intergenic regions. In contrast, when the rDNAs of *X. laevis* were hybridized to those of *X. borealis* (misidentified as *X. mulleri* in the original reference), a much lower level of overall similarity was found. While the coding regions were still highly conserved, the IGSs were found to be sharply divergent, although within each species the IGS was conserved. This motif of conserved coding regions and nonconserved intergenic spacers with species-specific mutations has been identified in the rDNA of all species studied (Dover and Flavell, 1984). The phenomenon in which this pattern of intraspecific homogeneity and interspecific heterogeneity is maintained has been called horizontal evolution (Brown *et al.*, 1972) and coincidental evolution (Hood *et al.*, 1975), but now is usually termed concerted evolution (Zimmer *et al.*, 1980).

Concerted evolution initially was proposed to operate via either a sudden correction model or a gradual correction model (Brown and Sugimoto, 1974). All models of concerted evolution require that the rate of mutation be lower than the rate of fixation (Arnheim, 1983). In the sudden correction model, the possible mechanisms included saltatory replication (Britten and Kohne, 1968) and master-slave correction (Callan, 1967). In saltatory replication, on the order of every 10 to 50 generations, one or a few of the repeat units are laterally amplified to replace all the other copies within the gene family. This process is a means to obtain homogeneity, but not to maintain it, since after the saltatory event, each member of the gene family would be able to accumulate mutations reducing intraspecific similarity (Li *et al.*, 1985). Master-slave correction is a process by which one member of the gene family is used as the template for replication of the entire gene family each generation. This cannot be the case for rDNA since some species exhibit variation in the length of the repeat unit within the same chromosomal locus (Li *et al.*, 1985).

The mechanisms of gradual correction are the ones now accepted as the preferred means of concerted evolution. Primarily these are unequal crossingover or unequal exchange, and gene conversion (Dover 1982; Arnheim, 1983). In order to achieve overall homogeneity, one or both of these processes (and possibly others) must take place within each individual locus, between rDNA loci on homologous chromosomes and between rDNA loci on non-homologous chromosomes.

Unequal crossingover (Tartof, 1975; Smith, 1976) has been examined within the rDNA families of yeast (Petes, 1980) and *Drosophila* (Coen *et al.*, 1982). In an unequal exchange, a recombination event will lead to a sequence duplication in one chromatid or chromosome and a deletion in the other chromatid or chromosome. If there are six consecutive repeats with the same mutation at one locus and the sister chromatids align so that crossing over occurs between the second mutant repeat of one chromatid and the fourth of the other, one chromatid might end up with four copies of the mutant gene and the other would have eight at the completion of the exchange. The effect of the process is to make each daughter chromatid more homogeneous for the wild type or mutant type than either parental chromatid (Li *et al.*, 1985). The copy number of the family also will vary due to unequal exchange and one variant of the gene will eventually become fixed within the population. Computer modeling studies and analytical treatments have shown that unequal exchange can eventually lead to the fixation of a mutant gene within a population even with only one or a few original copies of the mutant (Smith, 1974, 1976; Ohta, 1983).

Gene conversion is another mechanism which produces or maintains sequence homogeneity within a gene family. One strand from each of two different genes forms a duplex and if there is a mismatch due to a mutation in one of the genes, cellular DNA repair enzymes will correct the mismatch. In yeast, there is evidence for gene conversion occurring between genes on the same chromosome (Klein and Petes, 1981), on homologous chromosomes (Fogel *et al.*, 1978) and on nonhomologous chromosomes (Scherer and Davis, 1980). Using the above example with six mutant genes on two sister chromatids, heteroduplex formation between a mutant and a wild type gene might convert a mutant back to wild type leaving one chromatid more homogeneous for the wild type and the other unchanged. This nonreciprocal process always will leave one chromatid (or chromosome) more homogeneous for one variant and the other unchanged. Appels and Dvorak (1982b) have proposed that perhaps rRNA itself mediates gene conversion events by forming a heteroduplex with rDNA. Theoretical studies have shown that gene conversion, like unequal exchange, can lead to fixation of a variant within a population even beginning with a single copy of the mutant gene (Birky and Skavaril, 1976; Ohta, 1984). Gene conversion also can contribute to the variation in copy number within a single family locus (Li *et al.*, 1985).

Experimentally, the rate of concerted evolution within a population is dependent upon a number of variables, including the size of the gene family, the architecture of the arrays (i.e., tandem or interspersed) and the chromosomal location of the repeat units. The number of unequal crossover events required to achieve fixation increases roughly with the number of repeats in the family (Smith, 1974). Unequal exchange can have deleterious effects if the genes are interspersed instead of tandem, making it an inefficient mechanism for homogenization. Interchromosomal exchange can be significantly facilitated if the rDNA clusters are located at the ends of the chromosome as they are for humans (Arnheim, 1983). In meiotic human cells, with rDNA located on five different chromosome pairs, rDNA from more than one locus will sometimes combine to form an active nucleolus, site of rRNA synthesis. The proximity of the rDNA regions to one another could facilitate gene conversion or unequal exchange events between nonhomologous chromosomes (Arheim, 1983). In mice, the overall level of homogeneity between loci on different chromosomes is not as high as that in humans, perhaps due to the fact that the rDNA arrays are in the middle of the chromosome making interchromosomal strand exchange difficult or impossible. There is no evidence for more than one locus being able to contribute to a nucleolus in mice (Arheim, 1983).

Theoretically a gene conversion can proceed in either direction when a heteroduplex is recognized, that is, the mutant may be converted to wild type or vice versa. However, if there is even a small bias in one direction or the other, the rate of concerted evolution can be significantly increased (Nagylaki and Petes, 1982). Dover (1982) has called the phenomenon of gene family homogenization and fixation due to unequal crossingover and biased gene conversion molecular drive. Transposition may also play an important role in molecular drive, but it has not yet been demonstrated as a mechanism in the concerted evolution of rDNA families. Experimental studies on the relative importance of various mechanisms that can produce concerted evolution remain to be done in plant systems. It is clear, however, both from restriction mapping and nucleic acid sequencing studies (Appels and Honeycutt, 1986; Zimmer *et al.*, 1988) that plant rDNA arrays exhibit standard patterns of concerted evolution.

Nuclear rDNA copy number variation. The copy number of rRNA repeat units is highly variable in plants (Appels and Honeycutt, 1986) as well as animals (Long and Dawid, 1980). In plants, the variation exists at the interspecific and intraspecific levels as well as between individuals of the same population (Rogers and Bendich, 1987). Within a species, rDNA copy number can have a four-fold level of variation (Jorgensen and Cluster, 1988). Among inbred lines of maize, rDNA copy number has been shown to have a 10-fold range (Rivin *et al.*, 1986). Within a population of wild barley, a six-fold range in the copy number was detected between different individuals, and within a large population of broad bean, the copy number ranged from 500 to 44,000 per individual and the copy number was found to vary in different tissues (Rogers and Bendich, 1987). Experiments in *Drosophila* have shown that there is a minimum level of rDNA required and possession of genes in excess of those required has no discernible effect on phenotype (Shermoen and Kiefer, 1975; Tartof, 1975). An overabundance of rDNA is one way for the cell to insure that at critical times during development or in cases of stress there is sufficient cellular machinery for protein synthesis.

There is evidence that there is a large excess of rDNA within the plant nuclear genome; structural studies in maize (Phillips, 1978) and DNAse digestion experiments in wheat have shown that a large amount of rDNA lies within the heterochromatic, that is, the nontranscribed, region of the chromosome (Flavell, 1986). Those rRNA genes that are transcribed lie in the nucleolar organizer region (NOR) of the chromosome. The genes within the NOR are methylated to a lesser degree than those in the heterochromatin; the loss of methyl groups from cytosine residues in animal genes has been associated with gene activation (Razin and Riggs, 1980). In wheat the relative size of the NOR at a chromosomal locus, and hence the activity of that NOR, is proportional to the fraction of the rRNA genes without methylated cytosines (Flavell *et al.*, 1983). Deletion of the NOR with the high activity results in a decrease in the methylation at the other NORs and a concomitant increase in rDNA expression at the other NORs (Flavell, 1986). Similar inactivity of hypermethylated rDNA genes recently has been demonstrated in maize (Jupe and Zimmer, 1990).

Unequal crossingover between ribosomal arrays on sister chromatids or homologous chromosomes coupled with deletions is probably responsible for the high variation in rDNA copy number seen in plants and other organisms (Flavell, 1986). The process of gene conversion can also increase or decrease the number of repeats in an array (Dvorak, 1989).

Nuclear rDNA length variation. Restriction site analysis shows that there is no measurable variation in the lengths of the coding regions of the rDNA repeat units of plants (Jorgensen and Cluster, 1988). Sequencing of the soybean, maize and rice 18S genes has shown these cistrons to be 1807, 1809 and 1812 bp in length, respectively (Eckenrode *et al.*, 1985; Messing *et al.*, 1984; Takaiwa *et al.*, 1984). Among plants, only the 26S gene of rice has been completely sequenced and it is 3376 bp in length (Takaiwa *et al.*, 1985a). The 26S genes of two species of *Saccharomyces* are 3550 and 3549 bp, while mammalian 26S genes range from 4869 bp in mouse to 5184 in human (reviewed in Gutell and Fox, 1988). The lengths of the 5.8S genes of rice and broad bean are 163 bp (Takaiwa *et al.*, 1985b; Tanaka *et al.*, 1980). No plant ribosomal genes are known to have intervening sequences (IVS) within the coding regions so the lengths of the mature RNAs are the same as those of the coding regions. Some species of insect and protozoa do have an IVS within a subset of their 25S genes (Appels and Honeycutt, 1986) and recently an IVS was identified within the 18S gene of *Pneumocystis carinii* (Edman *et al.*, 1988). In *Drosophila*, the genes with the intervening sequences are not expressed (Long and Dawid, 1980), but in *Tetrahymena* the precursor rRNA acts as a catalyst for splicing out the IVS to form the mature rRNA (Cech, 1983).

In the rDNA of rice and cucumber, ITS1 is 194 and 229 bp and ITS2 is 233 and 245 bp, respectively (Hemleben *et al.*, 1988). No ITS length variation was detected within species of broad bean and species of pea, but comparisons between different legume genera showed some slight length variation (Jorgensen and Cluster, 1988).

The length of the intergenic spacer ranges from 1 to 8 kbp in plants thus far examined (Jorgensen and Cluster, 1988). The IGS heterogeneity accounts for the interspecific range of 8 to 15 kbp in repeat unit length (Hemleben *et al.*, 1988). The IGS may also show considerable length variation within populations of one species, within individuals of a population and even within individual chromosomal loci (Schaal and Learn, 1988).

Intraspecific variation in IGS length is caused by the presence of varying numbers of subrepeats in the middle region of the IGS. In most plant species, the subrepeats range from 100-200 bp. In species of wheat, barley and broad bean, the subrepeats are 130 bp, 115 bp and 325 bp (consisting of two copies of a 155-bp repeat and an unrelated 14-bp fragment), respectively (Appels and Dvorak, 1982a; Saghai-Maroof *et al.*, 1984; Yakura *et al.*, 1984). In corn, the 10 subrepeats are not constant in size, but range from 165 to 234 bp in length (McMullen *et al.*, 1986). Samples of wheat have shown heterogeneity for IGS length between individuals of a population, each variant differing from the others by a multiple of 130 bp (Appels and Dvorak, 1982a). In broad bean, individual plants can exhibit as many as 20 different size classes of IGS each differing by a multiple of 325 bp. The broad bean has only one chromosomal locus for rDNA, so the heterogeneity must occur among neighboring repeat units (Rogers *et al.*, 1986). Not all species show length heterogeneity, however. Soybean and *Lisianthius skinneri* have shown no variation within their rDNA for repeat unit size (Doyle and Beachy, 1985; Sytsma and Schaal, 1985). The mechanism for the variation in IGS length presumably is unequal crossingover within an individual repeat unit.

Nuclear rDNA sequence variation. Within the coding regions of the small (18S-like) and large (26S-like) rRNAs are stretches of nucleotides conserved across all species examined, including bacteria, yeast, plants and animals (Gerbi, 1985). Other regions of the small and large rRNA primary sequence are conserved only between more closely related phyla or classes, while a certain fraction of the rRNA is not conserved to any significant extent. In some of the areas where the primary sequence is divergent, computer modeling and chemical probing have suggested that the secondary structures of the rRNA molecules are conserved. Both the small and large rRNA molecules have areas of base-paired nucleotides which form stems; at the ends of these stems lie single-stranded loops. It is believed that this core secondary structure is maintained through selection by the stringent requirements of protein synthesis (Gerbi, 1985). In the double-stranded stems, there may be compensatory mutations which restore base pairing after one nucleotide of the pair changes (Wheeler and Honeycutt, 1988).

Comparisons between the rRNA molecules of bacteria and various eukaryotes have revealed that the insertion of so-called expansion segments within the bacterial sequences can account for the differences in length (e.g., 2500 for the *E. coli* 23S and 3300 for rice 26S) (Clark *et al.*, 1984). These expansion segments are proposed to be located such that major secondary structure elements are conserved in prokaryotes and eukaryotes (Gerbi, 1985). The expansion segments are usually found in the same location in the rRNA of different eukaryotes, but their lengths and sequences are not conserved.

The 5.8S sequences are conserved at the same level as the 18S and 26S sequences: sequencing has shown that there is only 1 bp difference between pea and broad bean and 2 bp different between pea and lupine (Jorgensen and Cluster, 1988). The sequences of the internal transcribed spacers are much more divergent. Comparisons of ITS1 of pea and broad bean showed one region of 16-18% difference and the remainder at 55% difference. The second ITS was constructed similarly, with two regions of different levels of conservation (Jorgensen and Cluster, 1988). The two levels of conservation could reflect the presence of processing signals within the ITS regions, perhaps for the post-transcriptional modifications.

The intergenic spacer is by far the most divergent part of the rDNA gene, making it useful for microevolutionary phylogenetic comparisons. The sequences of the subrepeats within the IGS are substantially conserved within a species, though not necessarily identical. Sequencing the broad bean subrepeats indicated that only five or fewer of the 325 nucleotides were not conserved through all copies of the subrepeat (Yakura *et al.*, 1984). Interspecifically there is generally little conservation of subrepeat structure, although some similarity has been detected between wheat and maize subrepeats (reviewed in Schaal and Learn, 1988). It is possible that in the genome the subrepeats function as hotspots for recombination or possibly as enhancers of transcription (Rogers and Bendich, 1987). In *Xenopus* the subrepeats within the IGS have been shown to possess enhancer activity: they increase the level of transcription from downstream coding regions irrespective of their orientation (Reeder, 1984).

The region downstream of the subrepeats which contains the ribosomal gene promoter shows little interspecific conservation; only short stretches are similar among closely related species. Sequence comparisons from different taxa have shown that there does not seem to be a consensus sequence analogous to the TATA box of genes transcribed by RNA polymerase II. In animal systems, it has been shown that RNA polymerase I of one species generally is incapable of transcribing the rDNA from another species (Grummt *et al.*, 1982). This stands in stark contrast to RNA polymerase II transcription, in which yeast can faithfully transcribe mammalian genes. The lack of sequence conservation and the species-specific nature of Polymerase I transcription indicate that the promoter region of the rDNA IGS has been evolving rapidly and that the RNA polymerase I must be co-evolving at a similar rate (Flavell, 1986).

THE ORIGIN AND EARLY RADIATION OF THE ANGIOSPERMS

The angiosperms are the most recently evolved of the major groups of plants and presently make up the largest and most diverse group of "flora" in the world. Estimates of the number of extant angiosperm species range from 240,000 to 300,000; this is more than the combined number of species of algae, bryophytes (liverworts, hornworts and mosses), pteridophytes (ferns), and gymnosperms (Friis *et al.*, 1986). Only insects among higher eukaryotes have more extant species than the flowering plants.

There are a number of morphological and developmental features that unite the angiosperms: Presence of flowers, bitegmic ovules, enclosed carpels, reduced size of the gametophytes, double fertilization, endosperm formation, tectate pollen and vessels in the xylem (Taylor, 1981). While certain of these features are absent in some angiosperms, or are occasionally found in groups other than the angiosperms, the formation of endosperm and double fertilization are uniquely derived conditions (synapomorphies) of angiosperms, although recent evidence points to a variation on angiospermous double fertilization in the Gnetalean genus *Ephedra* (Friedman, 1990). Most researchers believe that the large suite of characters which unites the angiosperms indicates a monophyletic origin of the angiosperms, that is, all angiosperms share a single common ancestor (Beck, 1974; Donoghue, 1989). It is hard to conceive of so complicated a process as double fertilization arising more than once during evolution, although there are some, like Meeuse (1967) who suggest that modern angiosperms arose independently from several different lineages, i.e., a polyphyletic origin.

Among living plants, angiosperms are most closely related to the other seed plants, the gymnosperms. The name gymnosperm means "naked seed" and represents one of the primary characteristics which separates the gymnosperms from the angiosperms. Angiosperms also have reduced male and female gametophytes compared to those of gymnosperms and a more sophisticated vascular system than gymnosperms. Results of studies of pollen and leaf fossils suggest that the gymnosperms first appeared during the late Devonian period about 360 million years ago (mya) (Friis *et al.*, 1986). There are four divisions of extant gymnosperms: Coniferophyta (conifers), Cycadophyta (cycads), Ginkgophyta (ginkgo) and Gnetophyta (Gnetales). Also important to a discussion of seed plant phylogeny are key fossil lineages. Cordaites are extinct gymnosperms related to the conifers; Bennettitales are an extinct order of Cycadophyta. The seed ferns, which include Caytoniales, Glossopteridales, Callistophyton, Corystospermaceae and Medullosa, represent an extinct division of gymnosperms, generally thought to have been the antecedents of the Cycadophyta (Cronquist, 1968).

Within the angiosperms, all species can be classified as monocotyledons or dicotyledons, based on the number (one or two) of primordial leaves (cotyledons) on the axis of an emerging seedling. Both groups are diverse. There are about 200,000 species of dicots including most trees and shrubs (except for the gymnospermous conifers, ginkgo and cycads), as well as many herbaceous plants like the composites, and 60,000 species of monocots, including the cereals, palms and orchids. There are other features which serve to separate the monocots from the dicots: among them are leaf venation patterns, number of floral parts, pollen type, vascular arrangement and presence of secondary xylem (wood). In dicots, the venation pattern is net-like with primary, secondary, tertiary and quaternary ranks of veins; in monocots, the veins usually lie in parallel arrangements of approximately equal rank after the primary vein. Dicot flower parts typically come in fours or fives, while monocot floral parts usually come in threes. Dicot pollen is mostly triaperturate while monocot pollen is normally uniaperturate. The vascular bundles of dicots are arranged in a ring, while those of monocots are more dispersed. There are exceptions to all of these generalities except that true secondary xylem is absent in all monocots (Raven *et al.*, 1986, p. 354).

Since the later part of the nineteenth century there has been much discussion of the origin and early evolution of the angiosperms. Almost all theories about the evolution of angiosperms and most classification schemes for angiosperms have been based on morphological, cytological, developmental and, to a lesser extent, phytochemical comparisons between species. Shared characteristics, especially leaf and floral morphology, have been used to place taxa into different groups. Before a truly phylogenetic classification can be proposed and the evolutionary relationships between the different groups can be assigned however, the polarity of character evolution should be inferred according to cladistic principles. That is, the primitive and derived conditions of each character should be determined. This will be discussed below.

The fossil record has been used to polarize some characters, most notably, those of pollen and leaves in progressively younger sediments near the Potomac basin of Virginia and Maryland (Hickey and Doyle, 1977). Beginning at about the Barremian age of the Early Cretaceous (about 118 mya), the oldest unequivocal angiosperm pollen grains had one germinal furrow (i.e., they were monosulcate) and a columellar exine structure in the pollen wall, similar to that of extant monocots and some members of the Magnoliidae. Moving up through the younger sediments, the triaperturate pollen types were found: tricolpate, then tricolporate and later triporate. These observations, along with the fact that most gymnosperms have monosulcate and never triaperturate pollen, strongly suggest that in angiosperms uniaperturate pollen is primitive and triaperturate is advanced. The columellar exine which facilitates adhesion of pollen grains to insects is also indicative of primitive entomophily (insect pollination) in the angiosperms (Hickey and Doyle, 1977). The oldest angiosperm leaves from the same sediments were mostly small, simple (i.e., not compound) and pinnately veined with several orders of reticulate venation. Most of the leaves had entire margins, though a few had irregular teeth in the leaf margins. In progressively younger strata, leaf diversity increased significantly; palmate venation appeared and both pinnate- and palmate-lobed leaves were first seen. Later still, compound leaves were found for the first time (Hickey and Doyle, 1977).

The wood, fruit and flower fossil records of angiosperms are not very complete and they cannot be used to determine the polarity of many characters (Hughes, 1976). In many cases it is assumed that evolution proceeded in such a manner that individual organs fused to form fewer, but more complex organs. For example, in a flower the condition of apocarpy, more than one carpel (female reproductive organ) each separate from the other, is considered primitive compared to the condition of syncarpy when the carpels are fused together. For the same reasons, compound leaves were considered a derived condition relative to simple leaves before there was solid fossil evidence to support the hypothesis. In general, any trend toward reduction and simplification is considered to be an evolutionary advance by many botanists.

In other cases, the primitive state of a character is determined by comparisons to outgroups like the gymnosperms. For example, in vascular plants other than the angiosperms, the predominant leaf arrangement is spiral, that is, only one leaf emerges from each node and in successive nodes, the leaves wind into a spiral (Cronquist, 1968). Consequently, within the angiosperms the condition of spiral phyllotaxy is considered primitive compared to opposite (two leaves at one node arranged opposite one another) and whorled (three or more leaves at the node). Similarly, the herbaceous habit is unknown in gymnosperms, so that within angiosperms it is assumed that being woody is primitive and herbaceous advanced (Cronquist, 1968). Further illustration is provided by the xylem of most gymnosperms which is made up solely of tracheids. In angiosperms, the xylem has both tracheids and vessels which are more efficient at water delivery and which give angiosperms a competitive advantage over gymnosperms. A few angiosperm groups have genera with vesselless xylem and the groups containing these genera have for this reason been presumed to be more primitive. Still, not all characteristics can be polarized and, as can be expected, the proposed phylogenetic relationships within angiosperms can be significantly affected by the presumed polarity of any character or suite of characters.

Another problem encountered in comparative morphological studies is in the assignment of homology. In evolutionary terms, if two organs are truly homologous, they are descended with modifications from a common ancestor, but not necessarily descended directly one from the other. A telling example is the effort to define a homologous structure to the enclosed angiospermous carpel within the potentially ancestral gymnosperms (Friis *et al.*, 1986). The carpel is the female reproductive unit and consists of a stigma lying atop a style which extends downward into the ovary where one or more ovules lie. The ovules of angiosperms are enclosed and protected from exposure and predators. In Gnetales and Bennettitales, the ovules are borne on the ends of stalks without associated leaves or anything else that would seem able to lead to carpel formation; however in several seed fern lineages, fossil evidence shows that the ovules were borne on leaf-like appendages which are easier to homologize to the enclosed carpel (Friis *et al.*, 1986).

Convergent evolution and reversals of characters also lead to difficulties in determining phylogenetic relationships. When a character or character state is shared by two otherwise distantly-related taxa it can serve to erroneously indicate a more recent common ancestry. An often-cited example of this convergent evolution is the common appearance of wings in birds and in mammalian bats. Reversals occur when a derived or advanced condition reverts back to the primitive state; failure to recognize a reversal (which usually is accomplished by considering the relative advancement of other, unrelated characters) also can lead to incorrect phylogenies. It appears that most, or perhaps all, the major trends recognized in plant evolution are reversible (Thorne, 1976; Endress, 1987). In part, this may occur because immobile plants must be more "plastic" in order to adapt to changing environments from which they cannot flee.

The groups of greatest interest relative to the results to be presented here are the dicot orders Magnoliales, Piperales and Nymphaeales and the monocots as a whole. The placement of these four groups and the extant gymnosperm orders will be emphasized in the discussion of theories of angiosperm evolution to follow.

At one time or another, most extinct and extant gymnosperms have been proposed as the group from which the angiosperms were derived. In the more recent classical taxonomic treatments, the authors have rejected the extant gymnosperms and presented angiosperms as derived from one of the groups of extinct seed ferns (Cronquist, 1968; Thorne, 1976; Rothwell, 1982; Meyen, 1984). Beck (1981) proposed that the gymnosperms and angiosperms were derived from two different lineages of Devonian progymnosperms, *Archeopteris* and *Aneurophytes*, with the former giving rise to extant conifers and ginkgo and the later giving rise to cycads, seed ferns and angiosperms.

A theory proposed by Bessey (1897) was that the monocots and dicots diverged early in angiosperm history, neither giving rise to the other, and that the most primitive dicots were the Ranales. The Ranales included the families of Magnoliales and Nymphaeales as well as others; he considered the Piperales to be very advanced. This Ranalian hypothesis of early dicots was supported by the work of Arber and Parkin (1907), authors of the Strobilus or Euanthial Theory. They proposed that the earliest angiosperms were woody and had flowers that were derived from unbranched strobili with many spirally-arranged male and female reproductive organs, similar to the strobili of the extinct gymnosperm group Bennettitales. They believed that the angiosperms were related to the Gnetales and Bennettitales. The flower of the Magnoliaceae and others of the Ranalian complex were considered to be the most similar to the earliest angiosperm. THis kind of flower is bisexual, beetle-pollinated, apocarpous with each carpel containing several ovules, and has many floral parts (sepals, petals, stamens and carpels) spirally arranged on a long axis. An opposing idea was the Pseudanthial Theory of Wettstein (1907) who proposed that the earliest dicots were derived from the Gnetales and were similar to the extant Piperales and the Amentiferae, a group including walnut and pecan, which consists of several families with very simple, anemophilous (wind-pollinated) flowers on a catkin (inflorescence). The carpels of these groups typically have only one unitegmic ovule. He suggested that the more complex flowers were derived through condensation of several smaller flowers. The Amentiferae are part of the subclass Hamamelidae, a group that proponents of the strobilus theory thought were derived from the Ranales.

In various forms the Ranalian theory continues to enjoy much support among plant systematists. Cronquist (1968), Takhtajan (1969) and Thorne (1976), in exhaustive classifications of the angiosperms, all have placed the subclass Magnoliidae (or its equivalent) at the base of the early radiation of the angiosperms because it is this group which contains more of the character states regarded as primitive. Stebbins (1974) agrees with their placement of Magnoliidae as the most similar to the primitive angiosperms. They all suggest that the other subclasses of angiosperms were derived from within the Magnoliidae. Cronquist (1968) and Takhtajan (1969) believe that the monocots arose from dicots related to the Nymphaeales. Within the Magnoliidae, Cronquist has placed six orders including the Magnoliales, Piperales and Nymphaeales. Takhtajan's Magnoliidae also include the Magnoliales, Piperales and Nymphaeales. Thorne's basal group is the superorder Annoniflorae which he has divided into three orders the Annonales (equivalent to Cronquist's Magnoliales), Berberidales (Cronquist's Ranunculales) and Nymphaeales. The Piperales are demoted to suborder status (Piperinae) by Thorne and placed within the order Annonales.

Cronquist has placed the families Piperaceae, Saururaceae and Chloranthaceae within the Piperales. Takhtajan moved the Chloranthaceae to another order within Magnoliidae. Thorne's suborder Piperinae contains only the families Piperaceae and Saururaceae. Cronquist's order Nymphaeales is composed of the families Nymphaeaceae, Nelumboaceae and Ceratophyllaceae. Takhtajan assigned the Nelumboaceae to a separate order within the Ranunculidae subclass, otherwise his Nymphaeales has the same composition as Cronquist's. Thorne's Nymphaeales are essentially identical to Cronquist's, except that he recognizes two different families Nymphaeaceae and Cabombaceae from within Cronquist's Nymphaeaceae. The different classification systems are summarized in Table 2.

Table 2. Summary of the groupings of taxa key to this study by various authors.

	Cronquist	Takhtajan	Thorne
Basal angiosperm group	Magnoliidae (subclass)	Magnoliidae (subclass)	Annoniflorae (superorder)
Members of basal group	Magnoliales Piperales Nymphaeales Aristolochiales Ranunculales Papaverales	Magnoliales Piperales Nymphaeales Aristolochiales Laurales Rafflesiales	Annonales Berberidales Nymphaeales
Families of Nymphaeales	Nymphaeaceae Ceratophyllaceae Nelumboaceae	Nymphaeaceae Ceratophyllaceae	Nymphaeaceae Cabombaceae Ceratophyllaceae Nelumboaceae
Families of Piperales (Thorne's Piperinae)	Piperaceae Saururaceae Chloranthaceae	Piperaceae Saururaceae	Piperaceae Saururaceae

Using a specific hypothesis for the basal angiosperm associations, each author has proposed a prototype "early flowering plant." The first angiosperm, according to Cronquist (1968), was an evergreen tree or large shrub of moist tropical habitat. Its leaves were small and spirally arranged on the axis, and they had entire margins and pinnate venation. The large, bisexual flower was at the end of a leafy branch and had a well-developed perianth and numerous free stamens and carpels. The flower was pollinated by beetles. Takhtajan (1969) and Thorne (1976) described the first angiosperm similarly, except that they did not expect the perianth to be differentiated into petals and sepals. Stebbins (1974) did not believe it necessary to assume spiral phyllotaxis and emphasized that while he thought the original angiosperm had a flower with spirally arranged parts, he did not believe it was derived from the strobili of Bennettitales or conifers.

Though most investigators believe that the dicots, specifically those from the subclass Magnoliidae, lie at the base of angiosperm radiation, the opinion is not unanimous. Burger (1977) challenged many of the traditional interpretations of the direction of floral evolution including the concept of the complex flower as primitive and the reduced flower as derived. He proposed that the primitive flower had one perianth part, two stamens and one pistil and that more complicated flowers evolved from this one by condensation. He concluded that the Piperales and monocots were very closely related and at the base of the angiosperm radiation. Burger (1981) expanded on his thesis later arguing that the leaf-like stamens of Degeneria (a member of Magnoliales), often cited as evidence for the primitive nature of the genus, were actually advanced features and challenging the presumed polarity of a number of other flowering plant features. He said that the most primitive angiosperms were small stemless herbaceous monocotyledonous plants and that woody stems evolved later during the diversification of dicots in the mid-Cretaceous. Burger's theories suggesting the monocots were basal and gave rise to the dicots through Piperales, Nymphaeales or Ranunculales are not inconsistent with the fossil pollen record.

There have only been a few robust cladistic studies of the origin and evolution of seed plants including angiosperms based on morphological features. Most notable are those of Crane (1985), Doyle and Donoghue (1986) and Donoghue and Doyle (1989a, 1989b). These analyses were based on morphological comparisons between extant and extinct gymnosperms and angiosperms. Crane (1985) suggested that, based on parsimony analyses of morphological characters, including floral structure, leaf node anatomy, vascular structure and many others that: (1) The seed plants were all descended from one common ancestor, i.e., they are monophyletic; (2) The Gnetales are a united monophyletic group, set apart from all other gymnosperms; (3) The Gnetales and the angiosperms are sister groups; (4) Along with the extinct Cordaites, Ginkgo and the extant conifers constituted a monophyletic group. One of Crane's two consensus trees is shown in Figure 2; the two consensus trees have some differences based on different assignments of certain homologies, but the results discussed here were not affected by the differing interpretations. Within the Gnetales he found that *Welwitschia* and *Gnetum* were more closely related to one another than either was to *Ephedra* and that *Ephedra* was the most primitive of the genera. As for the group(s) from which the angiosperms are descended, Crane's analyses indicate that the Gnetales and angiosperms together are derived from the same stock that gave rise to the extinct gymnosperms Bennettitales and *Pentoxylon*, in concordance with the hypothesis of Arber and Parkin (1907).

Doyle and Donoghue (1986) and Donoghue and Doyle (1989a, 1989b) expanded the analysis of Crane by the addition of more taxa and characters and by recoding the data set to minimize dependence on questionable polarity assignments. The addition of a second progymnosperm allowed them to test the hypothesis of Beck (1981) that the seed plants arose twice from two different progymnosperms. Their results based on a parsimony analysis are shown in Figure 3 and were very similar to those of Crane in that: (1) The seed plants were found to be monophyletic, having arisen from within the progymnosperms; (2) The angiosperms, Bennettitales, *Pentoxylon*, and Gnetales shared a common ancestor, although the Gnetales are not the sister group to the angiosperms, but rather to Bennettitales and *Pentoxylon*; (3) Extant conifers were found to be more closely related to ginkgo than to

Figure 2. One of Crane's (1985) trees for seed plants based on cladistic analyses of morphological data. Fossil taxa are in italics * denotes seed ferns

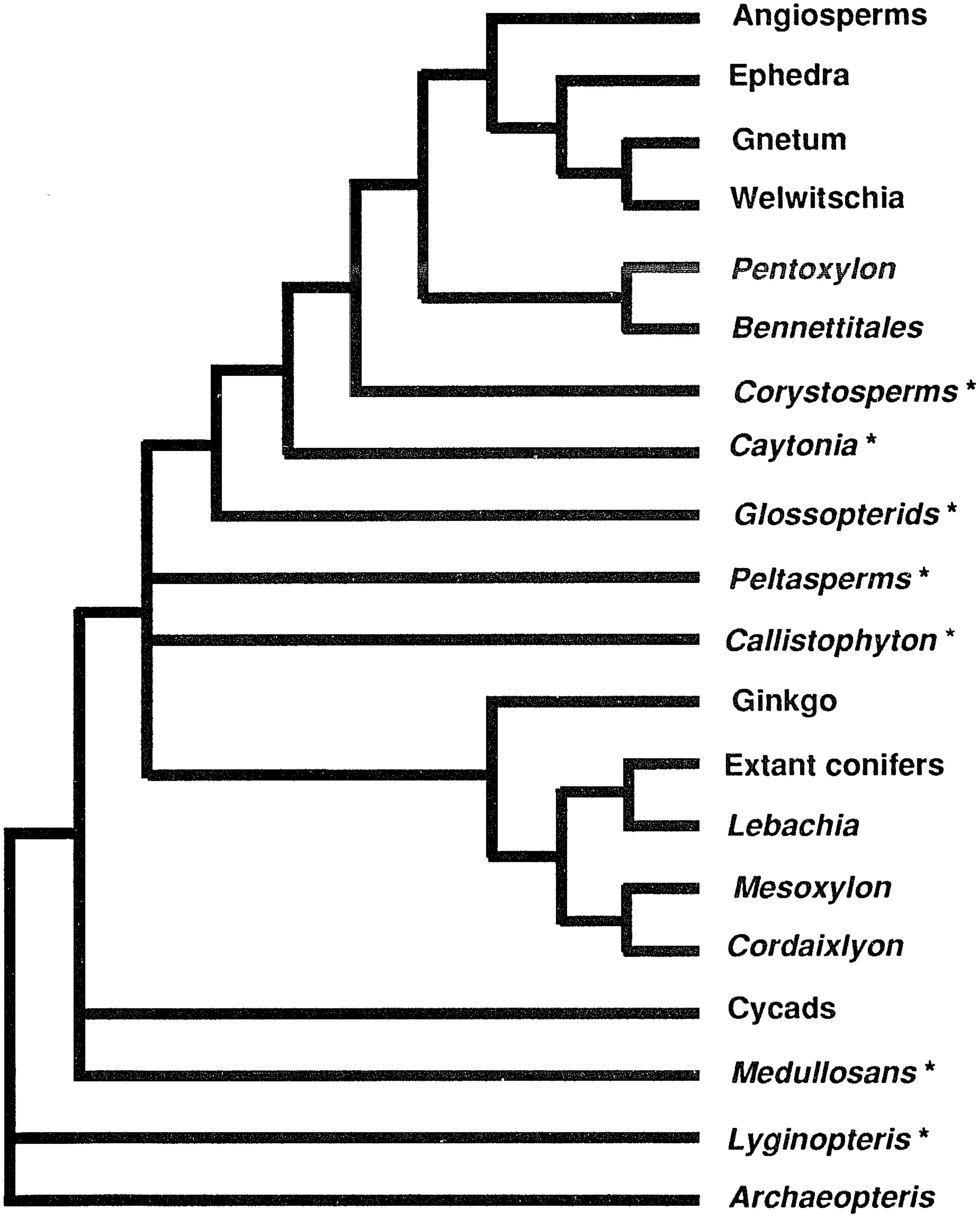
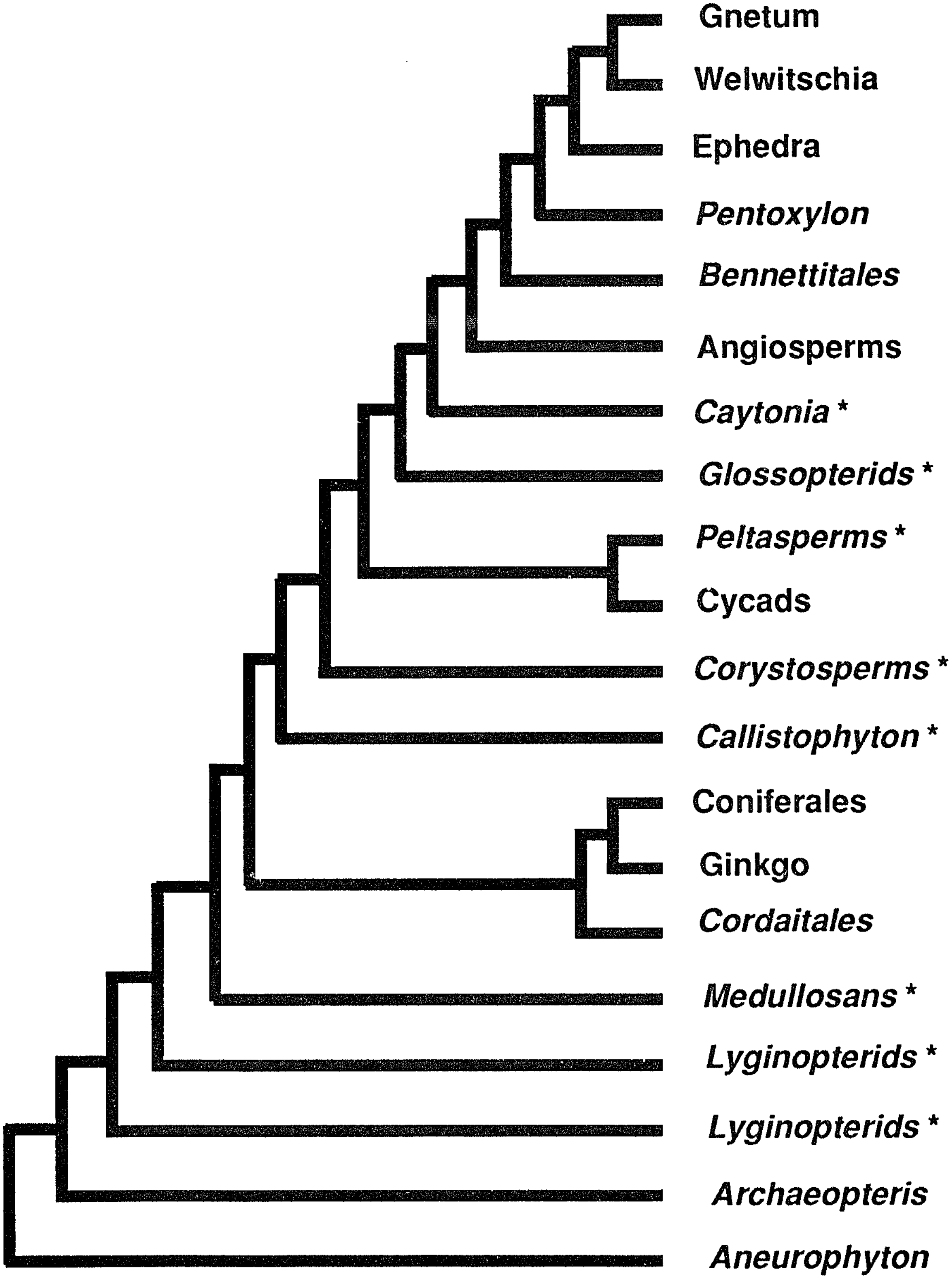


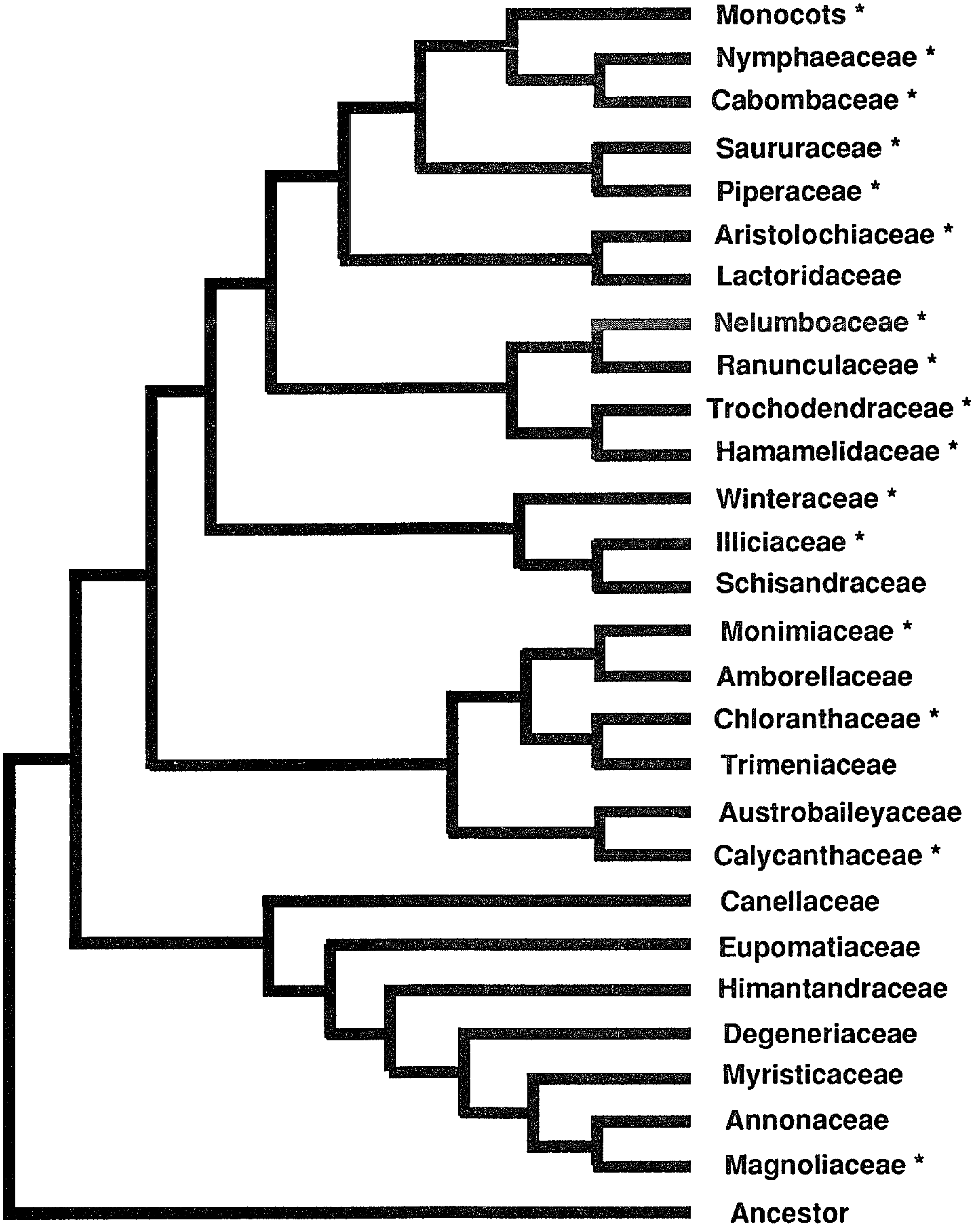
Figure 3. One of Donoghue and Doyle's (1989a) most parsimonious trees for seed plants, from a cladistic analysis of morphological data. Fossil taxa are in italics. * marks the seed ferns



extant cycads and the coniferopsid group (extant conifers, Cordaites and ginkgo) were monophyletic. Beck's hypothesis that the seed plants arose independently from Archaeopteris and Aneurophyton was not supported by the most parsimonious tree. It could, however, be supported on a tree only slightly less parsimonious (one step longer), but the authors point out that this is a result of the conservative nature of their data set and the omission of other characters which would provide additional support to the monophyly of seed plants. The trees of Crane (1985) and Donoghue and Doyle (1989a, 1989b) indicate that the seed ferns are not a natural group and that all extant seed plants arose from within them, although they differ as to which groups of seed ferns are most closely linked to the angiosperms.

Donoghue and Doyle (1989a, 1989b) performed a second parsimony analysis based on morphological features of 26 dicot families and the monocots, with the monocots treated as a single terminal taxon. Their most parsimonious trees place the families of the order Magnoliales (*sensu* Cronquist, 1968) at the root of the angiosperm tree (Figure 4). Their trees suggest that the Magnoliidae are not a natural group, but Donoghue and Doyle do recognize another natural group, one they named "Paleoherbs," which consists of Piperaceae, Saururaceae, Nymphaeaceae, Cabombaceae, Aristolochiaceae, Lactoridaceae and the monocots. Within the paleoherbs, Nymphaeaceae and Cabombaceae always make up a sister group to the monocots. Chloranthaceae were always excluded from the paleoherbs; in

Figure 4. One of Donoghue and Doyle's (1989a) most parsimonious trees for angiosperms, based on cladistic analyses of morphological data. * marks taxa included in this study



some of the most parsimonious trees, Nelumboaceae was part of the paleoherb clade, but not united with Nymphaeaceae and Cabombaceae. Ceratophyllaceae were not tested. Donoghue and Doyle (1989a, 1989b) also recognized a larger natural group consisting of the paleoherbs and the triaperturate dicots, which they called the "Palmates". Although the most parsimonious trees placed the ancestors of the Magnoliales as the most primitive angiosperms, the parsimony penalty to re-root the trees so that the paleoherbs were basal was only one or two steps (relative to a shortest tree of 178 steps). The shortest of these alternative trees was 179 steps long and placed the Nymphaeales at the base, followed by the monocots linked to the Piperales, an arrangement quite similar to some of Burger's (1977, 1981) ideas. The recent identification of a fossil leaf from the lower Cretaceous with low rank venation and other similarities to members of the paleoherbs also supports this alternative rooting of the flowering plants (Taylor and Hickey, 1990).

Martin and Dowd (1989) have used a parsimony analysis on the partial amino acid sequence of the small subunit of ribulose bisphosphate carboxylase/oxidase (rubisco) to study the evolution of flowering plants. They find that the most basal angiosperms are of the family Schisandraceae (a member of Magnoiliales [Cronquist, 1968]) and the next most basal is a group which includes the Nymphaeaceae and Cabombaceae. According to their trees, the Piperales are closely related to the monocots (Saururaceae were not represented in their trees) and the lineage leading to them diverged before the lineage leading to the Magnoliales. Their analysis suffers from dividing the tree up into branches which were each individually optimized and then re-combined. Combining these separate branches into one large tree does not guarantee that the globally most parsimonious tree has been found. Their results also are based only on comparisons of a small number of nucleotides; those inferred from the first 40 amino acid residues of the rubisco protein. When Archie (1989c) analyzed these inferred sequences in combination with DNA sequences inferred from two other proteins for a subset of these plant taxa, he found that the data were not any more informative than random sequences (see below).

In another recent study Troitsky *et al.* (1990) have analyzed 263 nucleotides from the nuclear-encoded 18S rRNA to propose the evolutionary relationships within the seed plants. In their 18S tree, the gymnosperms and angiosperms are sister groups, the Gnetales are split among the other gymnosperms, the monocots are a paraphyletic group at the base of angiosperm radiation, and the dicots are derived from the monocots. They only have one representative from the dicotyledonous paleoherb groups, *Peperomia*, and it is not near the base of the flowering plant radiation. Their parsimony analyses suffer from the same problem as those of Martin and Dowd (1989); they broke the data sets down into subsets and a locally most parsimonious tree was found for each subset and then an overall tree was constructed from the subtrees. The fact that the data set is relatively small, that the angiosperms do not arise from within the gymnosperms, that the Gnetales are not a coherent group and that the dicots are underrepresented, make the rest of their results also seem questionable.

At the start of the work described in this dissertation, it was clear that to properly address the relationships between gymnosperm and angiosperm groups and the early radiation of the angiosperms, rRNA sequences would be required minimally from *Ginkgo* and representatives of cycads, conifers, Gnetales, Piperales, Nymphaeales, Magnoliales, monocots and the more advanced dicots. Samples of all three genera of the Gnetales were acquired to test the naturalness of this order. As the study progressed, the choice of additional taxa was guided by the new work of Donoghue and Doyle (1989a, 1989b), resulting in a wide range of representatives of paleoherbs, of putatively primitive Magnoliidae and of their assumed close dicot relatives.

PHYLOGENETIC SYSTEMATICS

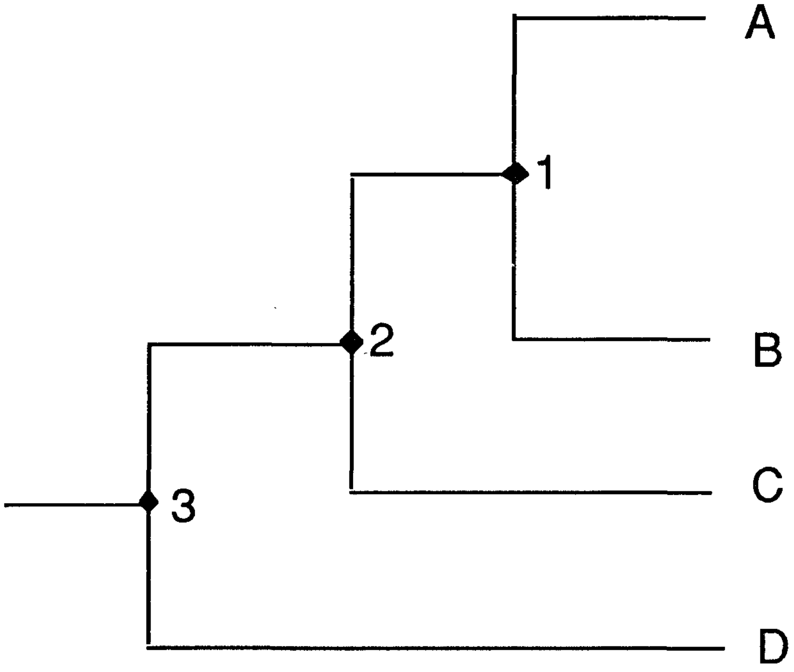
The concepts of phylogenetic systematics were first put forth by Willi Hennig (1950, 1965, 1966), a German entomologist. Hennig's method for the formulation of classification systems is based on several principles. Most important among these is that the only true hierarchical classification system of any group of organisms is one which reflects the evolutionary history of that group. All extinct and living organisms are phylogenetically related to one another at some level, because they are all descended from the first life form on earth. Therefore, saying two species are phylogenetically related is redundant, since all species are phylogenetically related. What is important, then, is the *relative* phylogenetic relationship among the species of interest. That is, asking the question "Are species A and species B more closely related to one another than either is to species C?" If species A and species B *are* more closely related to one another than either is to C, it implies that during the course of evolution, species A and species B shared a common ancestor more recently than species A, B and C shared a common ancestor. If true, it also means that species A, species B and their common ancestor (call it species AB) form a natural or monophyletic group, that is, a group which includes an ancestor and all its descendants. Monophyletic, or natural, groups are sometimes called clades. The concept of monophyly is also a relative one; species A, species B and species AB form a monophyletic group with respect to species C. One monophyletic group can be a subset of another, for example, if one goes back far enough on the evolutionary tree of life, some point will eventually be reached at which species A, species B, and species C form a monophyletic group. The natural group of A, B and AB is a subset of this group. It is not possible, however, for two monophyletic groups to partially overlap. Groups that contain a common ancestor, but not all the descendant species, are paraphyletic groups.

The characters that are used to unite species into natural groups must be characters whose present state arose during the common evolution of the members of that group (Hennig, 1965). This is another of Hennig's principles, the principle of cladistics, that species are united in monophyletic groups based on shared derived characters (or character states). In cladistic analyses, species are not placed into natural groups based on shared characters, or shared character states, that are considered to be primitive relative to the group under study. For example, if the common ancestor to all beetles were thought to have red eyes, then the possession of blue eyes among three beetle species would be a shared derived character state (synapomorphy) useful for uniting these three beetle species into a natural group within the larger natural group of beetles. However, the retention of red eyes is not a valid character state for grouping the remaining beetle lineages into another monophyletic group, because that would constitute uniting the species based on a shared primitive character state (symplesiomorphy). Usually the primitive state of any character is determined by comparison to an outgroup species, a closely-related species that is not a member of the group of interest (the ingroup). When a character state is present in the outgroup and some members of the ingroup, then it is considered to be the primitive state, and retention of that state is not sufficient grounds for uniting taxa within the ingroup into a natural group.

A derived condition that is unique to one of the ingroup species does not provide any information for cladistic analyses, either. This character state, called an autapomorphy, serves only to indicate that the species that possesses the autapomorphy is different from the other species, but this is already known. Using the beetle example again, the condition of green eyes unique to a fourth species would be an autapomorphy which would contribute nothing toward inferring a new natural group within the beetles. If later another species is identified with green eyes, the autapomorphy would become a shared derived character which would serve to join the two green-eyed species into a natural group.

In a phylogenetic tree inferred by a cladistic analysis, each node represents the common ancestor of the taxa at the tips of the branches that emerge from that node. In the phylogenetic tree shown in Figure 5, node 1 represents the species that was the common ancestor of species A and species B. Node 2 is the common ancestor of species A, B and C, and node 3 represents the common ancestor of species A, B, C and D. Node 1, species A and species B form a monophyletic group, as do nodes 1 and 2 along with species A, B and C. Those character state changes which occurred on the branch connecting node 3 to node 2 are changes that unite species A, B and C and the species represented by nodes 1 and 2 into a natural group. Similarly, the changes that occurred in the branch connecting node 2 to node 1 are the shared derived characters which unite species A and B into a monophyletic group. From the phylogenetic tree, it is possible to infer the character states of each ancestral taxon on the tree.

Figure 5. A sample phylogenetic tree. Nodes are marked by black dots and numbered. Terminal taxa are represented by letters.

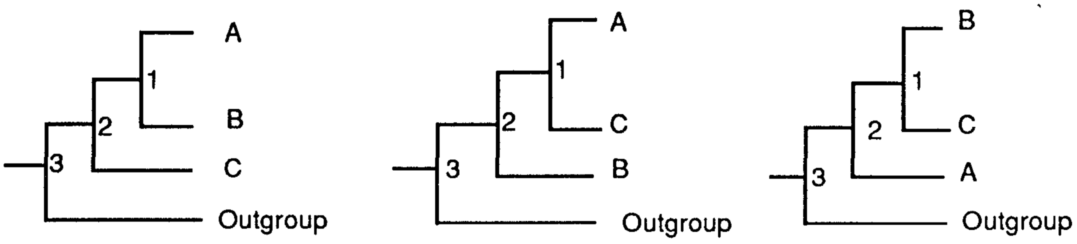


There are several different cladistic techniques available to infer evolutionary trees; the one used in this study is maximum parsimony. Maximum parsimony infers a tree that minimizes the total number of changes necessary to account for the distribution of the character states among the taxa of interest. Maximum parsimony can be analyzed under the constraints of the Wagner (Farris, 1970), Dollo (Farris, 1977), Camin-Sokal (1965) or Fitch (1971) algorithms. In Wagner parsimony, the character states are ordered, that is, they cannot change, for example, from red eyes directly to green eyes, without having been blue eyes in between. Wagner parsimony allows reversions at the same rate as forward changes, i.e., under Wagner parsimony, it is permitted for the descendants of one lineage to revert back to red eyes from blue, or from green eyes to blue. Dollo and Camin-Sokal parsimony also assume ordered characters, but both have restrictions on the number of times certain events are allowed. Under Dollo parsimony, a forward change is allowed to occur only once, but any number of reversions is allowed. Under Camin-Sokal parsimony, any number of forward changes is allowed, but no reversions are allowed. Fitch parsimony is used for unordered characters. Unordered characters can change from one state to any other without passing through intermediate states. Fitch parsimony does not penalize multiple occurrences or reversals.

In a real data set, resulting from real processes of evolution, not all characters are going to be distributed in such a manner that there will be one and only one phylogenetic tree whose topology perfectly accounts for each character. A sample data set is presented in Figure 6 with four characters and four taxa, one of which is the outgroup. If the tree is to be rooted by the outgroup, then there are only three possible arrangments of the ingroup taxa, A, B and C. These three possible arrangements also are shown in Figure 6. In topology I, parsimony would suggest that character 1 changed from the primitive to the derived state on the branch connecting node 3 to node 2. This would account for the distribution of character 1 among species A, B and C by one change. It is also possible that character 1 changed three times, once on each branch connected to a terminal taxon, or it could have changed twice, once on the branch connecting node 2 to node 1 and once on the branch connecting node 2 to species C, but these explanations require three and two changes, respectively, and therefore are less parsimonious than the one-change hypothesis. The change from primitive to derived for the second and third characters would be assigned most parsimoniously to the node connecting node 2 to node 1, requiring one change each. To fit the third character to topology I requires a change from primitive to derived between node 2 and species C, and another change between node 1 and species A. An equally parsimonious solution for character 4 would propose a change from primitive to derived on the branch connecting node 3 to node 2 and a reversal from derived to primitive; each solution proposed for character 4 requires two changes. A total of five

Figure 6. An example data set and alternative topologies for four taxa rooted by an outgroup.

	1 Carpel Condition	2 Leaf Phyllotaxis	3 Perianth	4 Leaf Margins
A	Syncarpous	Whorled	Present	Serrate
B	Syncarpous	Whorled	Present	Entire
C	Syncarpous	Spiral	Absent	Serrate
Outgroup	Apocarpous	Spiral	Absent	Entire



Topology I

Topology II

Topology III

changes, or steps, one each for characters 1, 2 and 3 and two for character 4, are necessary to account for the distribution of the characters according to topology I.

Topology II requires one change to explain the distribution of character 1, on the branch connecting node 3 to node 2. Characters 2 and 3 each require two changes, one on the branch connecting node 2 to species B and one on the branch connecting node 1 to species A. Character 4 may be accounted for by one change on the branch connecting node 2 to node 1. There are a total of six changes necessary to explain the distribution of the characters with topology II. Topology III requires one change to explain character 1, and two changes to explain characters 2, 3 and 4 for a total of seven changes over the entire data set.

In this example, then, maximum parsimony would choose topology I over topology II and topology III to best explain the distribution of characters among the taxa of interest because it is the shortest tree - the one requiring the fewest number of changes or steps. Character 4 is a homoplaseous character according to topology I, that is, it requires more than the minimum number of changes possible to account for its distribution. The minimum number of changes required to account for each character is one less than the number of different states present in the data set at that character. There are two states for character 4 (serrate margins and entire margins), so the minimum number of steps required to account for its distribution is one. With the beetle example above, there were three different character states for eye color: red (primitive), green and blue. If the data are unordered so that eye color can change from red to green or to blue, the minimum number of changes required for this character are two, one for a change from red to blue and a second to change from blue to green. Homoplaseous characters always require at least two gains (change from primitive to derived state) or at least one gain and one reversal.

The principle of phenetics, as opposed to cladistics, clusters species together based on overall similarity, that is, it treats shared derived and shared primitive characters as equally valid characters for construction of natural groups. Phenetics also treats uniquely derived characters as informative ones for grouping species together; it groups those species together that do not possess the autapomorphy, thus, phenetics groups taxa based on symplesiomorphies. As opposed to cladistic analyses, phenetic anayses do not necessarily have an evolutionary (or phylogenetic) connotation, though they are sometimes interpreted in this manner (Wiley, 1981). In phenetic analyses, the raw data, which may include melting point temperatures, allele frequencies, protein or nucleotide sequences, are converted to distances by various formulae, and then the taxa are clustered together based on minimizing distances between taxa. The nodes of a tree inferred from a phenetic analysis (a phenogram) do not represent any ancestral taxon and no character information may be inferred at the nodes.

While phenetic analyses violate Hennig's principles and there is evidence that phenetic analyses are not robust with the addition of greater amounts of data (Felsenstein, 1982), some types of data, like DNA-DNA hybridization and immunological data, can only be analyzed phenetically. Another drawback of phenetic techniques is that most assume an overall constant rate of change throughout the species being analyzed; parsimony is not as dependent on a constant rate of change (Wiley, 1981).

In this study, the characters of the data set are nucleotide sequences. The state of each character is G or A or T or C or absent. There are 58 ingroup taxa, various representatives of seed plants, and two outgroup taxa, seedless plants. The data were analyzed by maximum parsimony using the method of Fitch (1971) which allows the characters to change from one state to another without being required to pass through any intermediate state. Biologically this means that any nucleotide was allowed change to any other nucleotide, that is, a G could change to a C without having to first be an A. The other assumption of the parsimony analysis was that reversals were possible; a species could change from G to C and back to a G again at a particular nucleotide position. Phenetic analyses were performed for purposes of comparison, and the phenetic technique used, neighbor-joining (Saitou and Nei, 1987), was chosen because it is not dependent on a constant rate of change.

MATERIALS AND METHODS

PLANT MATERIALS

Table 3 contains the list of taxa used in this study, including the subclass, order, family, genus and species designations. The table also lists the source of each plant material.

RNA ISOLATION

Introduction. RNA sequencing with reverse transcriptase, synthetic oligonucleotide primers and dideoxynucleotides is an attractive choice for comparing ribosomal RNAs (rRNAs). The highly-conserved nature of rRNAs allows identical oligonucleotide primers to be used successfully with templates from all lineages of eukaryotes (Zimmer and Sims, 1985; Jupe *et al.*, 1988; Hamby and Zimmer, 1988). Similarly, "universal" primers can be synthesized for prokaryotic rRNAs (Lane *et al.*, 1985). Direct sequencing methods for RNA offer the advantages of bypassing labor-intensive cloning steps and, in the case of multigene families, of providing sequence information on those genes which are actually transcribed. These methods are most applicable to systems in which a large percentage of the total RNA preparation is a specific, homogeneous product (e.g., ribosomal RNAs [Zimmer and Sims, 1985; Lane *et al.*, 1985], abundant mRNAs [Martin *et al.*, 1981; Tolan *et al.*, 1984] and viral RNAs [Pace *et al.*, 1986]).

Table 3. Taxa sequenced in this study. The higher classifications are those of Takhtajan (1969). The affiliation of the source for each plant material is listed in the footnotes. If no affiliation is listed, then the source is from LSU. The v listed after a source name indicates that a voucher was prepared for the plant material and is on record at LSU.

Dicots

Subclass	Order	Family	Genus	Species	Common	Source
Magnoliidae	Magnoliales	Winteraceae	Drimys	winteri	drimys	J.Affolter[11]
		Magnoliaceae	Magnolia	grandiflora	magnolia	L.Sims
			Liriodendron	tulipfera	tulip tree	C.Knaak v
		Annonaceae	Asimina	triloba	pawpaw	M.Bowen v
	Laurales	Calycanthaceae	Calycanthus	occidentalis	carolina allspice	C.Knaak v
		Chloranthaceae	Chloranthus	spicatus	chloranthus	J.Doyle[8]
		Monimiaceae	Hedycarya	sp.		L.Thien[4]
	Piperales	Piperaceae	Piper	nigrum	black pepper	J.Wendell[2]
			Peperomia	sp.	peperomia	D.Nickrent[5]
		Saururaceae	Saururus	cernuus	lizard tail	R.Chapman
	Aristolochiales	Aristolochiaceae	Aristolochia	gigantea	Dutchman's pipe	J.Wendell[2]
			Saruma	henryi	saruma	J.Kress[3]
	Nymphaeales	Nymphaeaceae	Nymphaea	odorata	white waterlily	F.Givens v
			Nuphar	luteum	spatterdock	P.Raven[6]
		Cabombaceae	Cabomba	caroliniana	fanwort	E.Schneider[7]
		Barclayaceae	Barclaya	longifolia		D. Bryne[1] v
		Ceratophyllaceae	Ceratophyllum	sp.	coontail	F.Givens
Ranunculidae	Ranunculales	Ranunculaceae	Ranunculus	acris	buttercup	L.Sims v
	Nelumboales	Nelumboaceae	Nelumbo	nucifera	lotus	M.LeBlanc v
	Illicales	Illiciaceae	Illicium	floridanum	starbush	C.Knaak v
Hamamelidae	Trochodendrales	Trochodendraceae	Trochodendron	aralioides	trochodendron	S.Chaw[12]
	Hamamelidales	Hamamelideceae	Liquidambar	styraciflua	sweetgum	L.Sims v
		Platanaceae	Platanus	occidentalis	sycamore	L.Sims v
Rosidae	Fabales	Fabaceae	Glycine	max	soybean	S.Bartlett
			Pisum	sativa	pea	S.Bartlett
	Apiales	Apiaceae	Petroselinum	crispum	parsley	L.Sims
	Rosales	Rosaceae	Duchesnea	indica	indian strawberry	L.Sims v
Caryophyllidae	Caryophyllales	Caryophyllaceae	Stellaria	media	chickweed	L.Sims v
		Chenopodiaceae	Spinacia	oleracea	spinach	L.Sims

Table 3 (con'd)
Monocots

Subclass	Order	Family	Genus	Species	Common	Source
Alismatidae	Alismatales	Alismataceae	Echinodorus	cordefolius	echinodorus	F.Givens
			Sagittaria	lancifolia	arrowhead	E.Jupe v
	Najadales	Najadaceae	Najas	guadaliensis	pond weed	C.Knaak v
		Potamogetonaceae	Potamogeton	sp.	potamogeton	P.Hoch[6]
Arecidae	Arales	Araceae	Colocasia	antiquorum	elephant's ear	L.Sims
			Pistia	stratoides	water lettuce	P.Hoch[6]
	Arecales	Arecaceae	Sabal	minor	palmetto	J.Drost v
Commelinidae	Poales	Poaceae	Zea	mays	maize	L.Sims
			Tripsacum	dactyloides	tripsacum	K.Newton[9]
			Sorghum	bicolor	sorghum	K.Newton[9]
			Saccharum	officinarum	sugarcane	L.Sims
			Oryza	sativa	rice	L.Sims
			Hordeum	vulgare	barley	L.Sims
			Avena	sativa	oats	L.Sims
			Triticum	aestivum	wheat	L.Sims
			Arundinaria	gigantea	bamboo	K.Hamby v
Liliidae	Liliales	Liliaceae	Hosta	japonica	plantain lily	L.Sims v

Table 3 (con'd)
Gymnosperms

Subclass	Order	Family	Genus	Species	Common	Source
	Gnetales	Welwitschiaceae	Welwitschia	mirabilis	welwitschia	J.Folsom[13]
		Gnetaceae	Gnetum	ula	gnetum	J.Doyle[8]
		Ephedraceae	Ephedra	distachya		J.Doyle[8]
			Ephedra	tweediana	mormon tea	J.Doyle[8]
	Coniferales	Pinaceae	Pinus	taeda	pine	O.Stubbs[10]
		Taxodiaceae	Cryptomeria	japonica	cryptomeria	J.Drost
		Cupressaceae	Juniperus	ashei	rock cedar	J.Drost v
	Cycadales	Cycadaceae	Cycas	revoluta	cycad	L.Sims
		Zamiaceae	Zamia	ottonis	zamia	D.Nickrent[5]
			Zamia	floridana	zamia	D.Nickrent[5]
			Encephalartos	ferox	encephalartos	D.Nickrent[5]
	Ginkgoales	Ginkgoaceae	Ginkgo	biloba	maidenhair	L.Sims
Outgroups						
			Equisetum	hyemale	horsetail	R.Chapman
			Psilotum	nudum	psilotum	G.Learn[14]

Footnotes:
1. Suwanee Laboratories
Lake City, FL
2. Iowa State University
Ames, IA
3. Smithsonian National Arboretum
Washington, D.C.
4. Tulane University
New Orleans, LA
5. University of Illinois Herbarium
Champaign, IL
6. Missouri Botanical Garden
St. Louis, MO
7. Southwest Texas State University
San Marcos, TX
8. U.C. Davis Arboretum
Davis, CA
9. Univeristy of Missouri
Columbia, MO
10. Louisiana Dept. of Wildlife and Fisheries
Baton Rouge, LA
11. Berkeley Botanical Gdn.
Berkeley, CA
12. Academia Sinica
Taipei, Taiwan
13. Huntington Botanical Gdn.
San Marino, CA
14. Washington University
St. Louis, MO

Below, two techniques for total RNA isolation are presented and the advantages of each are briefly discussed. These relatively straight-forward total RNA isolation procedures allow collection of nuclear 18S and 26S rRNA as well as the 16S and 23S chloroplast rRNA. Consequently, proper design of the oligonucleotide primer allows the selective sequencing of any of the four molecules. I also discuss primer preparation, the sequencing method itself and some of the variables tested in order to optimize success in sequencing rRNA from a broad range of species.

RNA Isolation Protocols All glassware and spatulas were baked (200°C for 3 hours) to minimize RNase contamination. All plastic tubes, the Polytron (Brinkman Instruments) probe and the Miracloth (CalBiochem) were autoclaved. All solutions were made with DEPC-treated water (prepared as follows: water was brought to a final concentration of 0.1% DEPC, allowed to stand 12 hours, and then autoclaved). All plant tissue was collected in advance, quick-frozen with liquid nitrogen and stored at -80°C in airtight plastic bags until extraction.

There were two useful RNA extraction procedures. The first used a hot borate buffer and was a modification of the procedure of Hall *et al.* (1978). The second procedure used a guanidinium isothiocyanate extraction buffer and was a modification of the procedures of Glisin *et al.* (1974) and Chirgwin *et al.* (1979). The step-by-step protocols for both are published in Hamby *et al.* (1988).

The best yield and best RNA quality were obtained with young tissue and a high buffer-to-tissue ratio, i.e., between 5 and 10 ml of buffer per gram of tissue. For some taxa, e.g., conifers, cycads and ferns, it was best to use fresh material, freezing the sample only immediately before extraction. Overall yields were species dependent and were typically 40 to 400 μg of total RNA per gram of tissue.

The phenol:chloroform extraction was the most critical step of the hot borate method. After a successful phenol:chloroform extraction, the pellet changed from a slimy green mass to a clean white solid. Up to that step, the pellet did not stick tightly to the bottom of the tube. After the extraction it did. While the guanidinium method does not explicitly call for a phenol extraction, it may increase yield and help to deproteinize the RNA by adding a phenol:chloroform extraction after step 9 (Hamby *et al.*, 1988).

Initially, in a survey of about 40 taxa, the hot borate method was successful with 75-85% of the taxa. The guanidinium method was successful with only about 50% of the taxa. The hot borate method is simpler and quicker and does not require an ultracentrifuge. However, with certain species this method was not successful. For instance, it was only possible to isolate RNA from the ephedras, *Ephedra tweediana* and *E. distachya*, with the guanidinium method. There was no absolute pattern in success with either technique - the hot borate preparation was successful on tissue from *Welwitschia* and *Gnetum*, the two other genera (along with *Ephedra*) of the

Gnetales. Therefore the first attempt at RNA isolation from a new taxon was with the hot borate method; material was retained for the alternate method it case it were necessary.

OLIGONUCLEOTIDE PRIMER DESIGN AND PURIFICATION

Design. As mentioned above, it is possible to selectively collect sequence information from any of four different molecules: the nuclear-encoded 18S and 26S ribosomal RNAs and the chloroplast-encoded 16S and 23S ribosomal RNAs. The selective step is the design of the oligonucleotide primer.

The primer was designed to anneal to an invariant region of the target molecule which was identified by comparison of primary sequence data from several different known rRNA sequences. For example, in design of nuclear 18S primers, sequences of *Glycine max* (Eckenrode *et al.*, 1985), *Zea mays* (Messing *et al.*, 1984), *Rattus sp.* (Torczynski *et al.*, 1983), *Xenopus laevis* (Salim *et al.*, 1981), *Saccharomyces cerevisiae* (Rubstov *et al.*, 1980) and *Oryza sativa* (Takaiwa *et al.*, 1984) rRNAs were used. The first primers were up to 30 bases long, but subsequent experiments showed that high levels of specificity were obtained with 18-mers, the current design length.

The GAP program of the University of Wisconsin Genetics Computer Group package (Devereux *et al.*, 1984) was used to define regions on the other molecules to which the primer could possibly anneal. All primers had more than four mismatches with other potential target sites to minimize the chances of cross-hybridizations. A mismatch of at least three consecutive bases, or four out of the 18 nucleotides of the primer is usually sufficient to ensure selectivity.

Synthesis. The primers were synthesized on an automated DNA synthesizer (Applied Biosystems, model 380A) using phosphoramadite chemistry (Beaucage and Caruthers, 1981; Matteucci and Caruthers, 1981). In this automated process, the oligo is synthesized on a column in the 3' to 5' direction (the column is chosen according to the nucleotide at the 3' end of the oligo). Before incorporation into the oligo, the individual nucleotides are bound to phosphoramidite at the 3' end and to a dimethoxy trityl group at the 5' end. In addition, the amine sites of each base are protected by bulky groups and the reactive oxygens of the phosphate backbone are protected by methyl groups. Nucleotides are added one at a time in a cycle consisting of four steps: (1) The trityl group on the nucleotide at the 5' end (the growing end) of the oligo is removed by addition of trichloroacetic acid or $ZnBr_2$. (2) The next phosphoramidite nucleoside is added along with tetrazole to initiate the linkage reaction. (3) Some of the 5' ends that were deprotected in step (1) will not bind with the next nucleotide in step (2), so these 5' ends must be capped to prevent synthesis of N-1mers. This is accomplished by acetylating the 5' ends with acetic anhydride. (4) The phosphate backbone is oxidized by reaction with I_2-H_2O-lutidine-THF.

At the end of the synthesis, thiophenol was added to remove the methyl groups from the oxygens of the phosphate backbone. The oligo was separated from the support by addition of ammonium hydroxide. The bases were deprotected by heating to 55°C for 12 hours (the oligo was still in an ammonium hydroxide solution).

Purification. The purification was completed by separating the failed capped sequences from the complete sequences on a Poly-Pak reverse-phase chromatography column. The protocol below is faster and easier than the one published in Hamby *et al.* (1988).

(1) Add 1 ml of dH_2O to the oligo.

(2) Wash the column with 2 ml acetonitrile.

(3) Wash the column with 5 ml 2M TEAAc.

(4) Load the diluted oligo solution onto the column. Save the eluted volume and reapply it to the column. Save the final eluted volume because it may still contain some oligo if the cartridge is saturated. Only tritylated oligos should bind to the column.

(5) Flush the cartridge three times with 5 ml of dilute ammonium hydroxide (a 1:10 dilution of 30% ammonium hydroxide).

(6) Flush the column two times with 5 ml of dH_2O. Steps 6 and 7 remove impurities and untritylated sequences.

(7) Using a new syringe, wash the column two times with 5 ml of 2% TFA

to detritylate the bound oligonucleotide.

(8) Flush the column two times with 5 ml dH_2O.

(9) With a new syringe, elute the detritylated full-length oligo by flushing the column three times with 0.5 - 1 ml of 20% acetonitrile.

(10) Dry the eluate in the speed-vac and resuspend in 250 μl of TE. Dilute 1:100 and determine the concentration on the spectrophotometer. For an oligo, 1 OD_{260} is equal to about 35 μg/ml. OD_{260}/OD_{280} should be around 1.8.

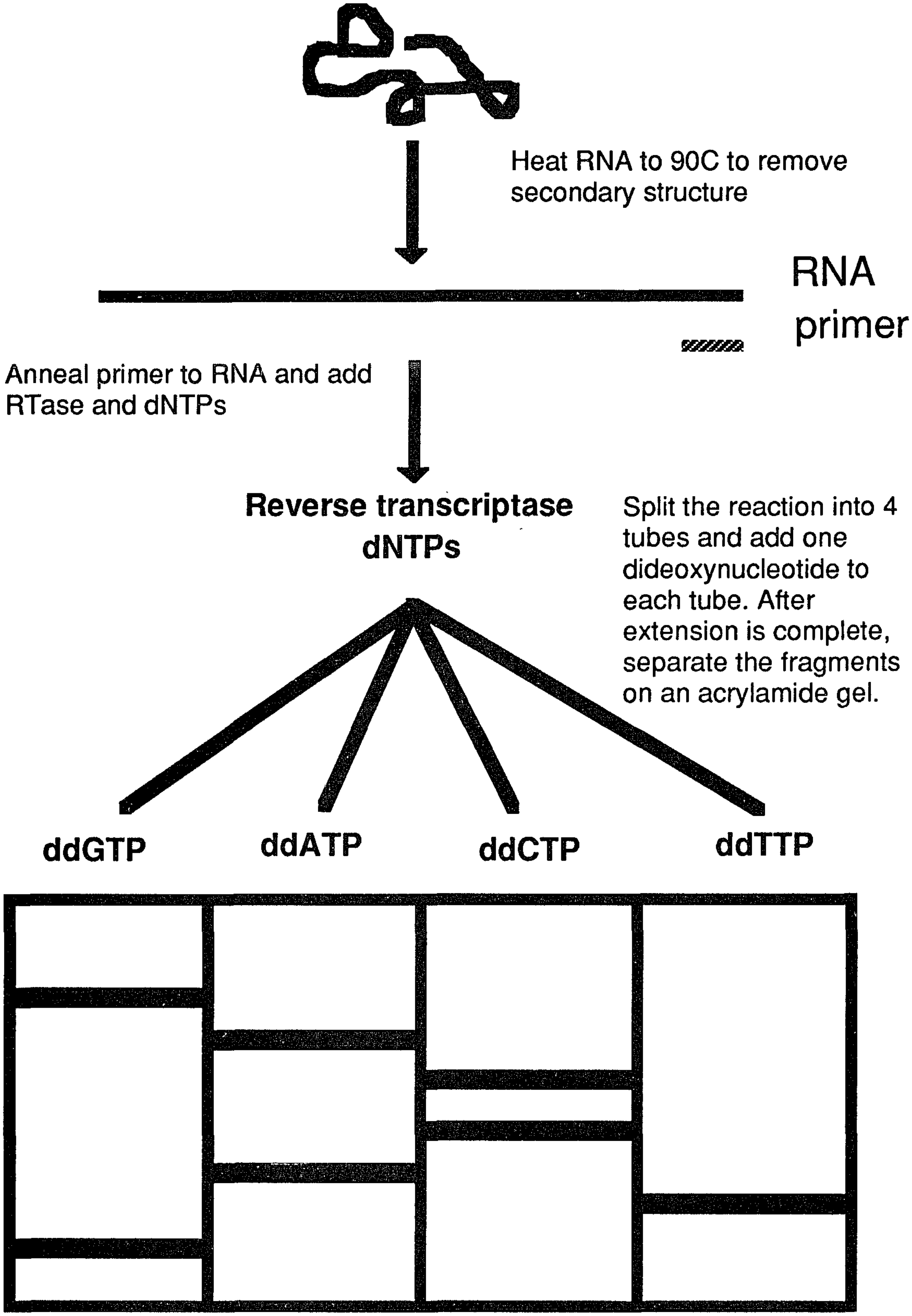
2M TEAAc is made by dropwise addition of 2 moles of triethylamine into an aqueous solution (500 ml) containing 2 moles of acetic acid in an ice bath. Adjust the pH to 7.0 and dilute to 1 liter with dH_2O.

RNA SEQUENCING REACTIONS AND GELS

Introduction. This procedure for reverse transcriptase sequencing with oligonucleotide primers and dideoxynucleotides is a modification of the techniques first described by Youvan and Hearst (1981) and Qu *et al.* (1983).

The procedure, shown schematically in Figure 7, was to first uncoil and linearize the RNA by heating it to 95°C for five minutes. Then the oligonucleotide primer was added and allowed to anneal to the RNA as it is cooled to 42°C. After the primer had annealed to the RNA, the mixture was divided into four tubes. A solution containing reverse transcriptase, all four

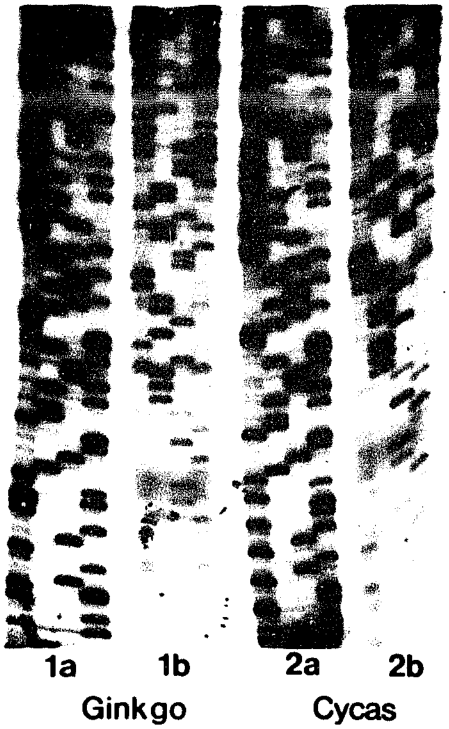
Figure 7. A schematic of the procedure for direct rRNA sequencing.


Heat RNA to 90C to remove secondary structure
RNA
primer
Anneal primer to RNA and add RTase and dNTPs
Reverse transcriptase
dNTPs
Split the reaction into 4 tubes and add one dideoxynucleotide to each tube. After extension is complete, separate the fragments on an acrylamide gel.
ddGTP
ddATP
ddCTP
ddTTP

deoxynucleotide triphosphates (one of which was radioactively labelled) and one of the four dideoxynucleotide triphosphates was added to each of the four tubes of RNA and primer. A different dideoxynucleotide triphosphate was added to each tube. Reverse transcriptase then directed extension from the 3' end of the primer to make a DNA strand complementary to the RNA to which the primer was annealed. Each time a dideoxynucleotide was incorporated into the growing strand of DNA, the strand was terminated because a dideoxynucleotide does not possess a hydroxy group at the 3' site of the nucleotide. For example, in the tube which contained dideoxyadenosine triphosphate, there were some species of DNA which terminated at the first adenosine in the growing chain, some which terminated at the second adenosine, some at the third and so on. After the reactions proceeded for *ca.* 20 minutes, a chase mixture of all four deoxynucleotide triphosphates was added to ensure that there were no chains terminated simply because the reverse transcriptase ran out of appropriate deoxynucleotide. The contents of each of the four tubes were then separated electrophoretically on a polyacrylamide gel. The gel was then dried and exposed to a piece of X-ray film and developed. A typical autoradiogram is shown in Figure 8. The sequence of the complementary DNA and, by inference, the sequence of the template RNA was read from the autoradiogram.

The exact details of the protocol are published in Hamby *et al.* (1988).

Figure 8. A typical autoradiogram.



Discussion There are at least two other ways to label the sequencing reactions for autoradiography: with ^{35}S-labelled deoxynucleotide, and with ^{32}P-end labelled, or "kinased", primer. There are few significant differences in the protocols for each. These modifications are detailed in Hamby *et al.* (1988).

In direct comparisons of gels using ^{35}S labelling and ^{32}P labelling, the results were nearly identical, but occasionally there were more stops in the reaction mixtures using ^{35}S. This occurrence of additional stops in the gels, along with the added inconvenience of fixing the gels and the fact that lab members routinely used ^{32}P in nick translation made ^{32}P labelling the method of choice.

Labelling with ^{32}P can be accomplished by either using a labelled deoxynucleotide in the extension reactions or by using a labelled primer. Results were satisfactory with the kinased primer, but the kinasing procedure must be repeated every 2-3 weeks, and any unused primer is lost. Consequently, α-^{32}P labelling in the extension reactions was chosen over kinasing the primers.

Periodically the concentrations of dideoxynucleotides must be fine-tuned by trial and error. When the concentration of dideoxynucleotide is too high, there is too much chain termination early in the extension step. This results in a gel in which the lane with too high a dideoxynucleotide concentration has very dark bands at the bottom of the gel, while at the top of the gel, no bands can be observed. Conversely, if the concentration of dideoxynucleotide is too low, there will not be any significant level of chain termination and all incorporation will be in long cDNAs which are not resolved on the sequencing gel.

A common problem with sequencing gels is the occurrence of compressions among the bands, especially in GC-rich areas, making it difficult to read through certain sections of the sequence. Several different approaches to alleviate this problem were tried: replacemenat of dGTP with 7-deaza-dGTP, addition of formamide to the gel mix, and substitution of inosine triphosphate for dGTP. None of these approaches offered any significant improvement to our rRNA sequencing. Some sequence ambiguities were resolved with terminal deoxynucleotidyl transferase (DeBorde *et al.*, 1986; Jupe, 1988).

DATA HANDLING

After the autoradiograms were developed, the RNA sequence was read and recorded. The sequences were then compared to a published sequence, usually soybean or rice, and any differences were confirmed by rechecking the autoradiogram. The final corrected sequence was then entered into the program SEQED of the University of Wisconsin Genetics Computer Group (UWGCG) package of programs (Devereux *et al.*, 1984) which runs on the College of Basic Science's VAX cluster. SEQED is an interactive program for data entry and editing.

Once the nucleotide sequences were collected, they were aligned using the UWGCG program GAP. First each new sequence was GAPed against a common sequence (usually soybean or rice); this program makes optimal pairwise alignments with the Needleman-Wunsch algorithm, inserting gaps into either sequence as necessary. GAPing against a common taxon also oriented each new taxon similarly to the ones already aligned. The resulting sequence (including any gaps) was then imported into the LINEUP program, an interactive editor for aligned sequences, and the alignments were fine-tuned by visual inspection. Any apparent anomalies revealed by the alignments were confirmed by another check of the autoradiogram. LINEUP may display up to 31 sequences simultaneously.

For archival purposes, a separate file was maintained for each species-primer combination (60 taxa times 8 primers equals 480 separate files) on the College of Basic Sciences' VAX computer. The files were organized into separate subdirectories, one for each primer. Each file was named in the same manner: six or fewer letters to describe the species name followed by a three-character extension which named the primer. For example, there was a file called soy.18J in a subdirectory named 18J, soy.18L in a subdirectory named 18L etc. The consistent use of this naming protocol simplified file retrieval. After each sequence was GAPed, the results were written to a new file named with up to nine characters to describe the species and primer followed by the extension .GAP, e.g., soy18g.gap. This was the file that was imported into LINEUP. At the end of an editing session, LINEUP renames the individual sequences by replacing the previous extension with .FRG so it is important to have all descriptive information before the extension. Otherwise, the next time one calls up LINEUP, the program may not retrieve the correct files.

DATA ANALYSIS The complete aligned sequences, including the invariant positions, were then transferred to a Macintosh computer via a modem and the file was edited so that it was in the proper format for Swofford's PAUP 3.0 (1989) which calculates phylogenetic relationships based on the principle of parsimony. PAUP has three different algorithms to calculate the most parsimonious solutions, the exhaustive search in which all possible topologies are considered, a branch-and-bound search procedure (Hendy and Penny, 1982) and a heuristic procedure. Only the exhaustive search, which evaluates every possible topology, is guaranteed to find the most parsimonious solution, but it is limited to about 10 or 11 taxa because the number of possible trees increases very rapidly with the addition of each new taxon. The number of possible trees can be calculated by the formula $\#trees=[2n-5]!!$ (Felsenstein, 1982) where n=number of taxa. The double factorial notation means multiplication by every other number beginning with $(2n-5)$ and continuing down to 1. The branch-and-bound algorithm is a modification of the exhaustive search procedure in which an initial upper bound of the tree is estimated, and if in the process of trying rearrangements on a particular branch it becomes clear that a certain arrangement will lead to trees that exceed the upper bound, the search along that particular branch is terminated and searching commences on the next branch. In practice the branch-and-bound algorithm is limited to 16 or 17 taxa, though Hendy and Penny have developed a new algorithm which will handle more taxa (Penny *et al.*, 1990). The heuristic search procedure takes certain shortcuts and approximations to try and find the shortest tree in a reasonable period of time. Basically, a first estimate of the best tree is constructed and then various branch swapping options are invoked to try to find shorter arrangements.

The program Hennig86 (Farris, 1986), another parsimony program which runs on the IBM PC, was used to compare to PAUP. PAUP has a utility to convert NEXUS data sets (the PAUP and MacClade format) into the proper format for Hennig86.

The MacClade program package of Maddison and Maddison (1990) was also useful in data analysis. It has a data editing window and a tree editing window. The data editor can be used to manually align sequences from more than 100 different taxa. Therefore, for new entries, it is possible to skip the LINEUP step on the VAX computer and enter GAPed sequences directly into the MacClade data editor. The tree editor permits interactive rearrangement of phylogenetic trees and recalculates tree parameters according to the new arrangement. Because this is a test version of MacClade, all of the results were separately confirmed by PAUP. (This task was not difficult because the had Maddisons consulted with Swofford to create a data format common to both PAUP and MacClade.)

PAUP can also carry out the bootstrap procedure of Felsenstein (1985). In bootstrapping, certain characters of the data set are randomly selected and eliminated. They are then replaced with other characters of the data set also chosen at random. This means that for example, in a data set with 100 characters, characters 2-20, 35 and 99 may be eliminated and replaced with characters 45-56, 59-66 and 77. This results in the replacement characters being counted twice, once in their original positions and again as replacements for the eliminated characters. After the data set is modified, a search is conducted for the shortest tree by one of the three available algorithms. After the shortest tree is found, the data set is modified again and another parsimony search undertaken. This is a test to determine which nodes of the tree are statistically supported. Felsenstein says that in order to be statistically significant, a particular arrangement of taxa must appear identically in 95 out of 100 bootstrap replications.

Archie's (1989a) randomization program, which runs on an IBM PC, requires a data set in the format of PAUP 2.4. PAUP 3.0 data sets were converted into PAUP 2.4 format by exporting the PAUP 3.0 files as Hennig86 files and editing them in a word processing program. This program creates random data sets from the original data set by randomly permuting the character state assignments at each individual site while maintaining the same character state distribution (see below).

The distance analyses were performed by the neighbor joining program of Saitou and Nei (1987). The PAUP data matrix was rewritten by MacClade so that the data were not interleaved, and then the data were input into a computer program that I wrote (see Appendix). This program, which runs on the VAX, calculates pairwise distances for each pair of taxa (1770 comparisons for 60 taxa) by three different formulae: a total dissimilarity equal to the number of differences divided by the number of bases compared; the Jukes-Cantor (1969) distance which compensates for multiple changes at one position; and the Kimura two-parameter distance (1980) which gives more weight to the less frequent transversion events. The computer program calculates the distances and then creates three different data sets for entry into the Neighbor-Joining program which runs on an IBM PC.

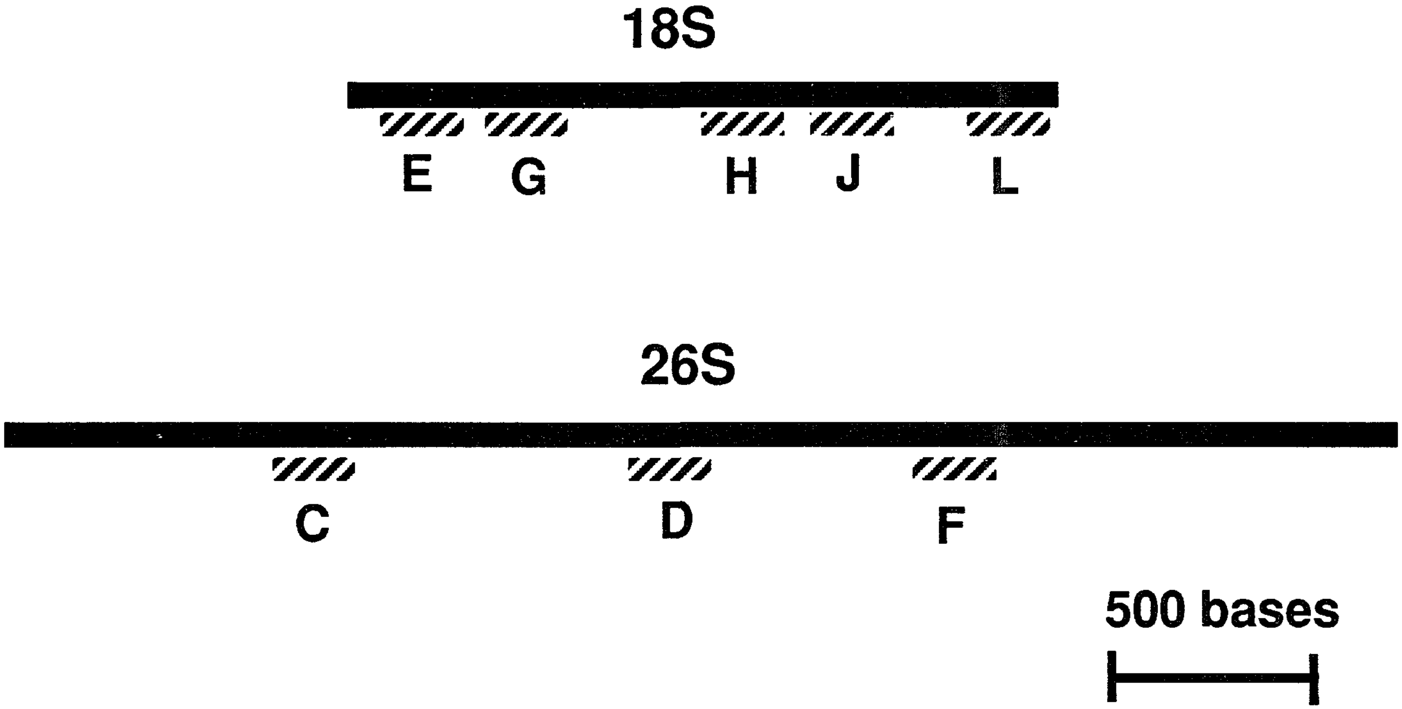
RESULTS

The primary nucleotide sequence was determined for five different regions (representing *ca.* 60%) of the 18S rRNA molecule and three regions (representing *ca.* 15%) of the 26S rRNA molecule for 60 plant taxa. The primers used were 18E, 18G, 18H, 18J and 18L in the small ribosomal subunit and 26C, 26D and 26F in the large subunit. Table 4 gives the sequences of these oligonucleotide primers and the regions of the reference rRNA molecules (from soybean or rice) to which they anneal. The relative positions of the primers are indicated in Figure 9. The number of nucleotide positions determined with each primer ranged from 191 to 250 and a total of 1097 nucleotides from the 18S molecule and 604 from the 26S molecule were compared. One short stretch of nucleotides (about 20 positions) in the region sequenced with the 18E primer was almost universally unreadable for all 60 taxa and was consequently eliminated from the alignments. Similarly two regions (totalling about 45 positions) within those sequenced with the 26F primer, and one region of about 20 nucleotides within the 18L region were also unreadable and these also were eliminated from the alignments. Various attempts were made to sequence through these problem regions, including substituting inosine for guanosine and chasing the reaction with terminal deoxynucleotidyl transferase (DeBorde *et al.*, 1987). None of the modifications succeeded, although the terminal transferase has worked in

Table 4. A list of primers used in direct rRNA sequencing, their sequence and the positions to which the primers anneal. The 18S positions are numbered relative to soybean (Eckenrode *et al.*, 1985). The 26S positions are numbered relative to rice (Takaiwa *et al.*, 1986).

NAME	LENGTH	PRIMER SEQUENCE	ANNEALS TO
18E	25	TACCATCGAAAGTTGATAGGGCAGA	SOY 308-332
18G	18	TGGCACCAGACTTGCCCT	SOY 554-571
18H	30	GCCCTTCCGTCAATTCCTTTAAGTTTCAGC	SOY 1131-1160
18J	27	TCTAAGGGCATCACAGACCTGTTATTG	SOY 1424-1450
18L	26	CACCTACGGAAACCTTGTTACGACTT	SOY 1762-1787
28C	22	GCTATCCTGAGGGAAACTTCGG	RICE 948-969
28D	18	CTTGGAGACCTGCTGCGG	RICE 1836-1853
28F	22	CAGAGCACTGGGCAGAAATCAC	RICE 2172-2193

Figure 9. Location of the regions sequenced by each primer. The length of the hatched bars corresponds to the length of the sequenced region.



other rRNA sequencing experiments at a different poblematic location (Jupe, 1988). The sequences of the eight separate regions for each taxon were concatenated in the computer to make one long file of 1701 nucleotides for each species. Gaps in the aligned sequences due to deletion or insertion events were coded separately and appended to the end of the alignments. There were only 13 sites within the eight sequenced regions where gaps were inferred to create exact alignments. Thirteen characters representing either the absence or presence of a gap at each of these sites were appended to the end of the data set for a grand total of 1714 positions. Table 5 shows the location of each of the gap sites within the eight sequenced regions and its corresponding position within the alignments (i.e., position 1702-1714).

Ribosomal RNA sequences and evolution in Poaceae. An initial investigation was undertaken by comparing 18S and 26S rRNA sequences of members of the grass family, Poaceae. This preliminary analysis was done to permit development of data handling procedures and to provide experience with the data analysis techniques, some of which are not available for large data sets. It also provided a test to determine if these rRNA sequences were able to resolve relationships below the family level. Poaceae were chosen because of our studies of grass ribosomal gene genetics and because in other analyses with more taxa, the members of this family were

Table 5. Location of gap sites within the 1701 aligned nucleotide sequences, and the corresponding position of this gap score in the last 13 positions of the 1714 entries in the data set. For example, the first gap site in the aligned sequences is at position 42 (of 1701 aligned) and the absence or presence of this gap is scored at position 1702 of the 1714 total sites in the data set.

Gap	Site	Position
1	42	1702
2	352	1703
3	394	1704
4	813-815	1705*
5	819	1706
6	823	1707
7	853-854	1708*
8	898	1709
9	965-966	1710*
10	1129	1711
11	1444-1448	1712*
12	1454-1456	1713*
13	1469-1470	1714*

* (Score gap as present if there is a gap at one or more of these positions)

consistently found to form a natural group. The genera represented are *Zea* (maize), *Tripsacum*, *Sorghum*, *Saccharum* (sugarcane), *Oryza* (rice), *Hordeum* (barley), *Avena* (oats), *Triticum* (wheat) and *Arundinaria* (bamboo); *Colocasia* (elephant's ear), another monocot of the family Araceae, was used as an outgroup.

The data are summarized in Table 6. Of the 1714 positions aligned, only 143 (i.e., 8.3%) were variable (i.e., 1571 sites were invariant). Of the 143 variable positions, 88 were variable only because of an autapomorphy, that is, a change inferred as unique to a particular terminal taxon. As stated previously, autapomorphies do not provide phylogenetic information because they serve only to separate that one taxon possessing the unique change from the other nine taxa. The 88 autapomorphies were divided such that 54 were specific to the outgroup, *Colocasia*, and 34 were autapomorphies within the grasses. The remaining 55 positions were variable and phylogenetically informative. Fifty-four of the informative sites changed via base substitution; only one of the six variable gap positions was informative. At 31 of the variable and informative sites, the changes were restricted to transitions, while only transversions had occurred at 14 sites. At the other nine sites, both transition and transversion events had to be postulated during the differentiation of these grass genomes.

Although almost twice as many positions were sequenced within the 18S rRNA molecule, the 18S molecule had only slightly more variable sites

Table 6. Summary of rRNA data over Poacea and *Colocasia*. Tn=transition, Tv=transversion, MH=multiply hit.

Primer	Region	Sites	Variable	Tn	Tv	MH	Informative	Tn	Tv	MH
18E	90-308	191	28	16	6	6	6	4	1	1
18G	300-554	250	8	2	4	2	1	0	0	1
18H	910-1134	215	13	8	4	1	8	5	2	1
18J	1210-1429	214	13	8	4	1	4	2	2	0
18L	1535-1766	227	12	7	5	0	1	0	1	0
18S total		1097	74	41	23	10	20	11	6	3
26C	740-949	202	13	6	3	4	4	0	2	2
26D	1625-1836	202	25	16	7	2	11	6	3	2
26F	1960-2172	200	25	19	4	2	19	14	3	2
26S total		604	63	41	14	8	34	20	8	6
18S+26S		1701	137	82	37	18	54	31	14	9
Gaps		13	6	-	-	-	1	-	-	-
GRAND TOTAL		1714	143	82	37	18	55	31	14	9

(74 to 63) and fewer informative sites (22 to 34) in comparison to the 26S rRNA molecule. The most variable regions were those sequenced with 18E, 26D and 26F. The region sequenced with 18G was the most conserved with only eight variable sites among the 250 sequenced.

The aligned sequences were read into PAUP 3.0g and the heuristic search process found the tree shown in Figure 10 to be the most parsimonious arrangement. Only one most parsimonious tree was found, with a length of 187 steps. Both a branch-and-bound search (Hendy and Penny, 1982) and an exhaustive search in which all possible topologies are tested, guaranteeing the most parsimonious solution, found the same shortest tree shown in Figure 10. With nine ingroup taxa and one outgroup, there are 2,027,025 possible arrangements. The exhaustive search tried each of these possible arrangements and found the lengths to be distributed as shown in Figure 11. On a MacIntosh IIcx, the heuristic and branch-and-bound algorithms executed completely in a matter of two or three seconds. The exhaustive search took a few minutes.

In the most parsimonious tree, *Arundinaria* branches first off the tree, leaving the other eight taxa as a natural group. This group is split into two smaller monophyletic groups: one contains *Triticum*, *Avena* and *Hordeum*; the other consists of *Oryza*, *Zea*, *Tripsacum*, *Sorghum* and *Saccharum*. *Zea* and *Tripsacum* form a natural group as do *Sorghum* and *Saccharum*. These four genera also form another monophyletic group. There is one tree of 188

Figure 10. The most parsimonious tree for Poaceae inferred from rRNA sequences.

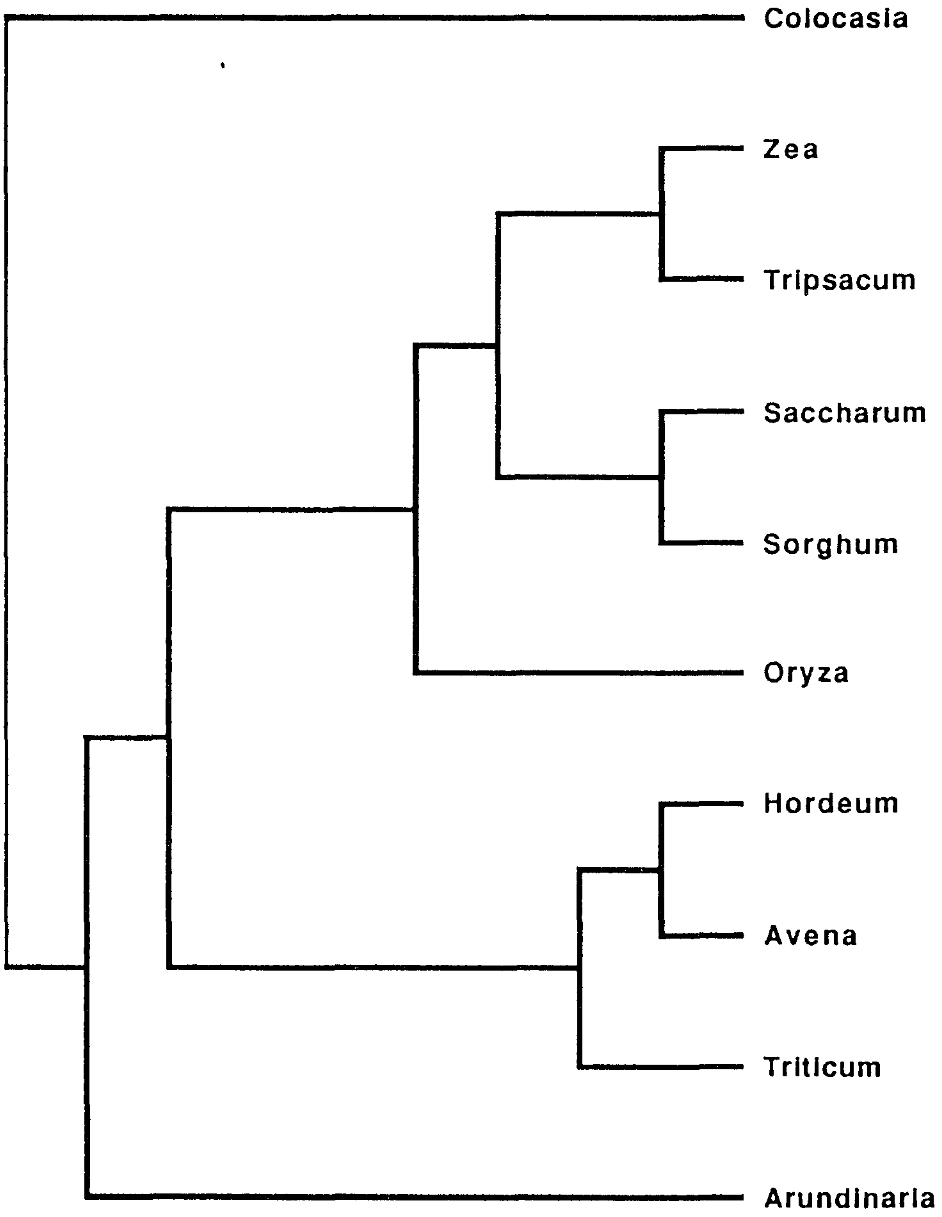


Figure 11. The distribution of trees found in an exhaustive search over the Poaceae rRNA sequence data. Numbers to the left of the vertical line are tree length; numbers to the right are the total number of trees at that length.

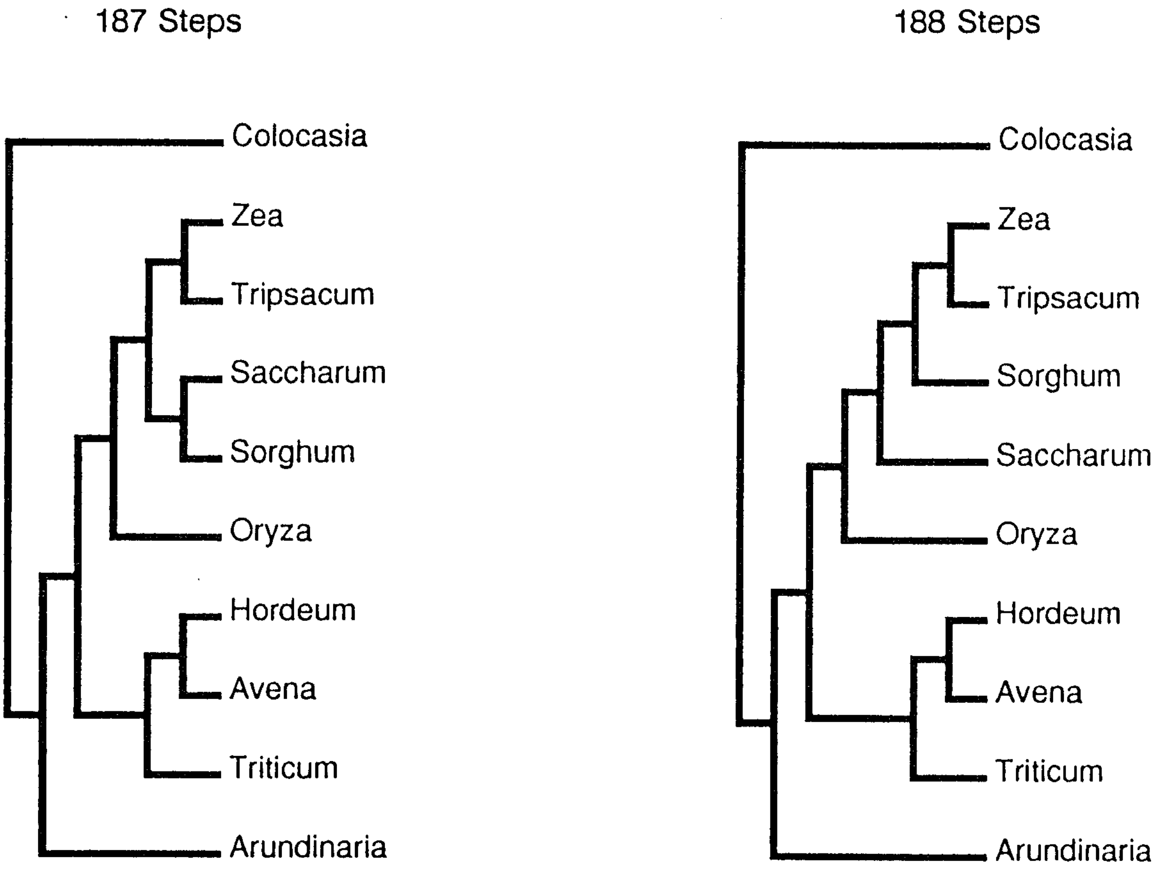
```
    +-----------------------------------------------------------
187 | (1)
188 | (1)
189 | (3)
190 | (3)
191 | (4)
192 | (15)
193 | (22)
194 | (21)
195 | (36)
196 | (55)
197 | (68)
198 | (68)
199 | (121)
200 | (118)
201 | (190)
202 | (248)
203 | (333)
204 | (436)
205 | (509)
206 | (620)
207 | (801)
208 | (871)
209 | (1085)
210 |* (1305)
211 |* (1669)
212 |* (2126)
213 |* (2756)
214 |* (3285)
215 |** (4064)
216 |** (4981)
217 |** (5912)
218 |*** (6542)
219 |*** (7109)
220 |*** (7395)
221 |*** (7459)
222 |*** (7718)
223 |*** (8379)
224 |**** (9391)
225 |**** (11319)
226 |***** (13220)
227 |****** (15675)
228 |******** (19196)
229 |********* (23203)
230 |*********** (28831)
231 |************** (34480)
232 |**************** (39129)
233 |***************** (43828)
234 |****************** (46134)
235 |******************* (46811)
236 |******************* (47160)
237 |****************** (45581)
238 |***************** (44015)
239 |**************** (39937)
240 |************** (36274)
241 |************** (36480)
242 |**************** (40797)
243 |****************** (46598)
244 |********************** (54979)
245 |************************** (64899)
246 |****************************** (75913)
247 |************************************ (89880)
248 |***************************************** (104307)
249 |********************************************** (116309)
250 |************************************************** (126646)
251 |****************************************************** (136155)
252 |**************************************************** (132034)
253 |*********************************************** (118593)
254 |************************************** (96519)
255 |**************************** (70329)
256 |****************** (45027)
257 |*********** (26962)
258 |****** (14664)
259 |*** (6574)
260 |* (2280)
261 | (514)
262 | (53)
    +-----------------------------------------------------------
```

steps differing only in the placement of *Sorghum* and *Saccharum* relative to one another: instead of forming a monophyletic group, they form a grade with *Sorghum* between *Saccharum* and the node leading to *Zea* and *Tripsacum*. This tree is compared to the most parsimonious tree in Figure 12.

The neighbor joining program of Saitou and Nei (1987) was employed to compare a phenetic (tree based on overall similarity) analysis to our cladistic one (tree based on shared derived characters). This program is insensitive to variations in the rate of evolution among different taxa. The nucleotide sequence data were converted to pairwise distances by dividing the number of variable positions by the number of sites compared between each different pair of taxa, a method considered valid for species not separated by great evolutionary time (Nei, 1987). The distances were alternatively calculated by the Jukes-Cantor method (1969) which compensates for multiple mutations at the same locus (position), and by the Kimura two-parameter model (1980) which gives more weight to less frequent transversions. Regardless of which distances were used, the topology of the resulting phenogram was the same as that of the most parsimonious cladogram.

Two hundred-fifty bootstrap replications were performed on the grass data to see which monophyletic groups were best supported by the rRNA sequence data. A majority-rule consensus tree of the 250 replications is

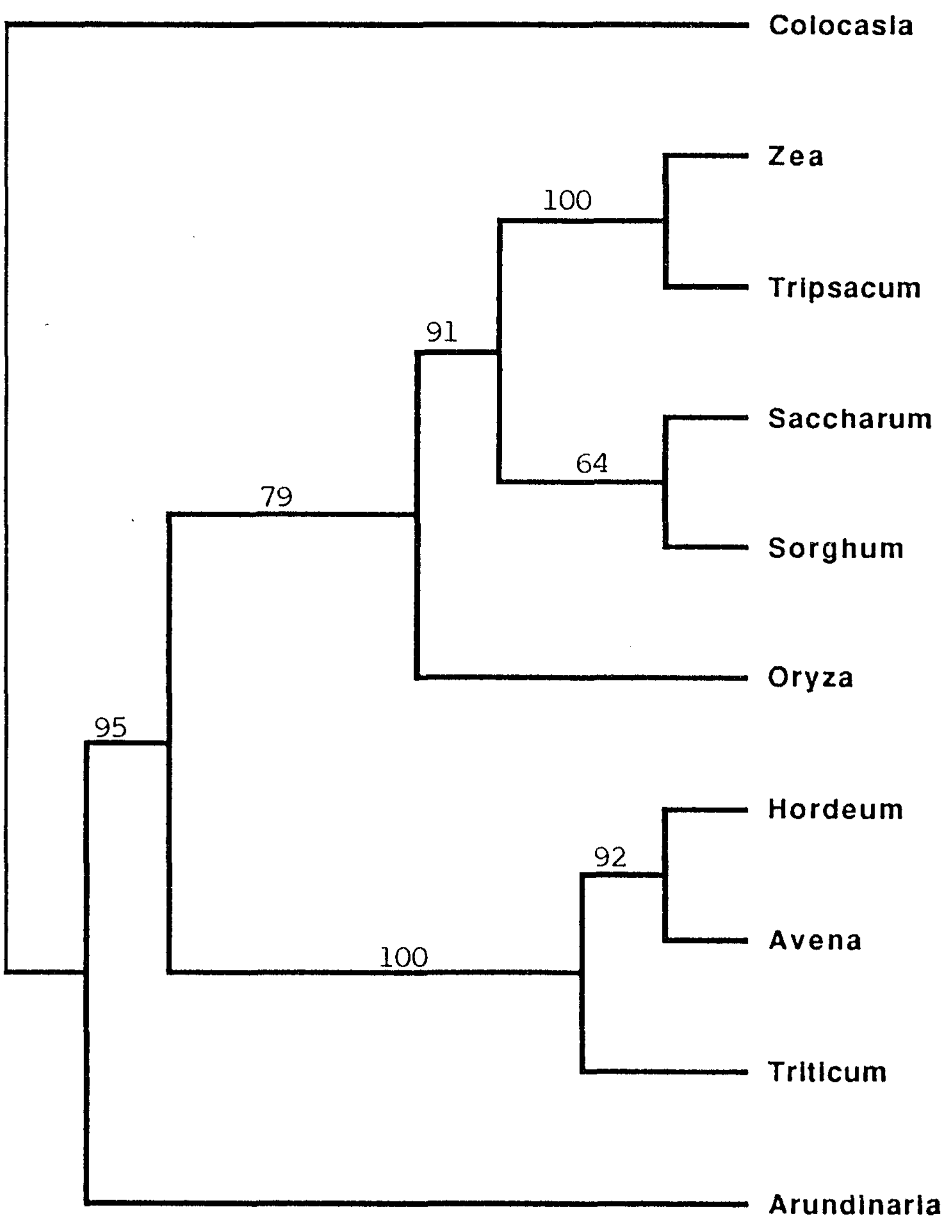
Figure 12. A comparison of the most parsimonious tree (187 steps) and the next-most parsimonious tree (188 steps) based on the Poaceae rRNA sequence data.



shown in Figure 13. A majority-rule consensus tree displays all nodes which are identical in at least 50% of the individual trees. Each node is labelled with the percentage of times out of 250 that the best tree(s) contained these nodes. In every bootstrap replication, *Zea* and *Tripsacum* were placed together as a monophyletic group and so were *Hordeum*, *Avena* and *Triticum*. Ninety-five percent of the time, *Arundinaria* was placed outside the other grasses which formed a monophyletic group. These are the only statistically significant groupings on the tree at the 95% confidence level.

In order to investigate how quickly support for various nodes on the shortest tree deteriorated as trees became less parsimonious, the trees that were one to 10 steps longer than the shortest tree of 187 steps were collected. There were a total of 227 trees within 10 steps of the most parsimonious tree. First the most parsimonious tree was combined with the one tree that was only one step longer, then these two were combined with the three that were two steps longer and so on until all 227 trees had been combined. After each new set of trees was added a majority-rule consensus tree was calculated by PAUP. As less and less parsimonious trees are added to the pool from which the consensus is calculated, support for various nodes will begin to weaken and become equivocal. The sooner that node support weakens with the addition of longer trees (as reflected by dissolution of dichotomous branching into polychotomous branching), the weaker the support for that node by the data. The series of consensus trees

Figure 13. A majority-rule consensus after 250 bootstrap replications of the grass rRNA sequence data. Nodes are labelled by the percentage of the 250 replications in which that node appears as it does on the consensus.

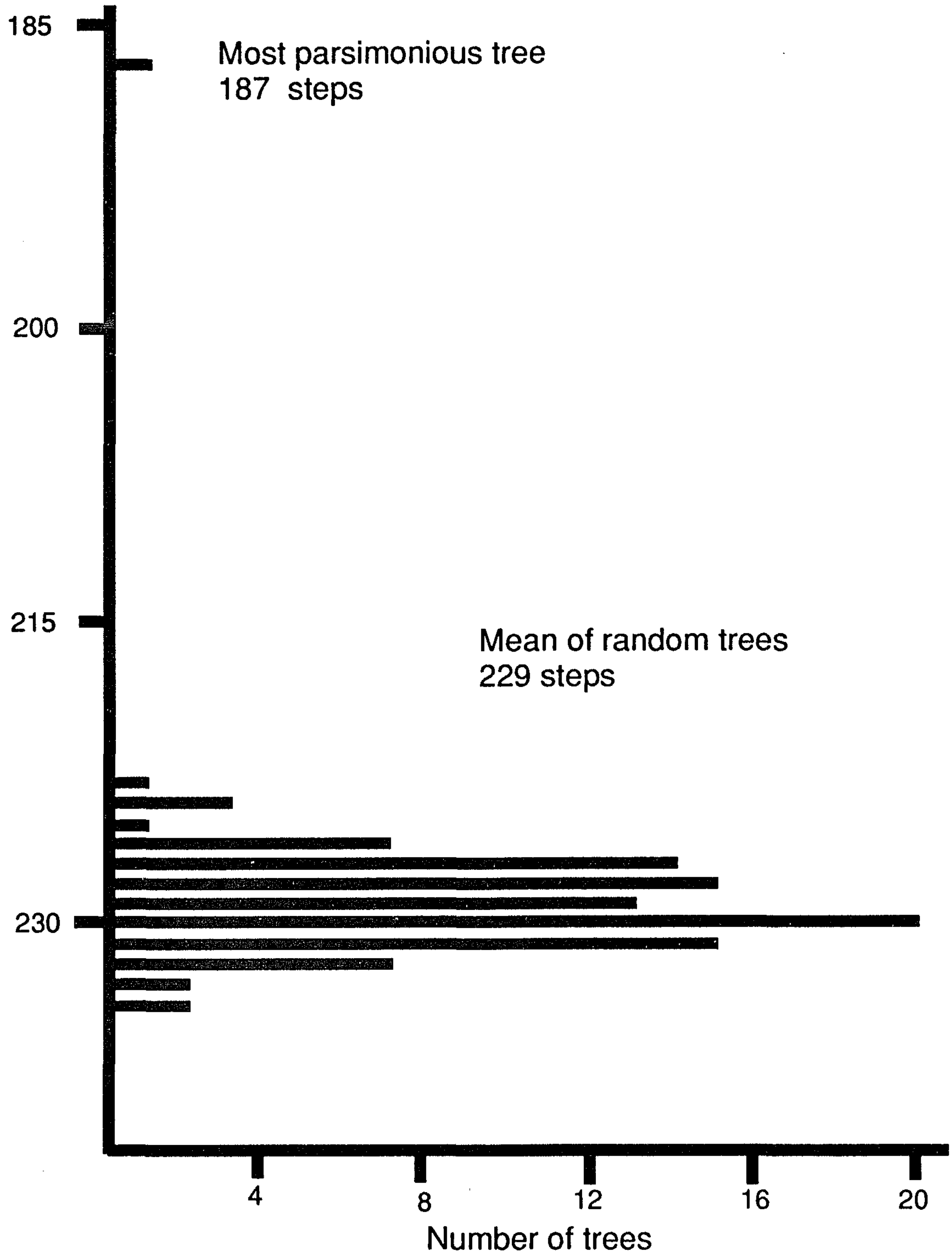


is presented in the next section.

The data were analyzed by Archie's (1989a) randomization program to test if the rRNA sequence data are informative, that is, if they are better than randomly generated sequence data. This program takes each character of the data set and looks at the distribution of character states at that site. Then the character states are randomly permuted at that site among the taxa keeping the same overall distribution of states. For example, if there are ten taxa, and the character state of character one is G for taxa 1-5, A for taxon 6, T for taxa 7-9 and C for taxon 10, the program may redistribute the character states so that character one is T for taxa 1, 6 and 10, C for taxon 2, A for taxon 5 and G for the rest. There are still five G's, one A, three T's and one C, but their arrangement among the taxa is different. Each character is independently randomly permuted and then the shortest tree is found with one of PAUP's searching algorithms. This procedure was done with the grass data 100 times and the shortest tree found each time. The results of this test are shown in Figure 14. The randomized data sets gave trees that ranged in length from 223 to 234 steps (as compared to a length of 187 for the nonrandomized data). The mean randomized tree length was 229 steps with a standard deviation of 2.2 so that the nonrandomized tree length is at least 18 standard deviations shorter.

Archie's program also allows for calculation of the homoplasy excess ratio (1989b). This is a statistic which measures the amount of homoplasy in

Figure 14. The distribution of trees found after 100 randomizations of the Poaceae rRNA data set.



the data and is different from the consistency index of Kluge and Farris (1969) which is by far the most commonly applied measure of the fitness of data. The consistency index is calculated by dividing the length of the shortest possible tree by the total length of the actual tree. The shortest possible length of any tree is calculated by subtracting the number of characters from the total number of character states in the data set. The consistency index of a tree in which there is no homoplasy, that is, no reversals or parallel changes, is 1.0 and theoretically, as data become more and more homoplaseous, the consistency index should approach 0.0. Archie (1989c) has shown, however, that the consistency index does not approach 0.0 for very homoplaseous data and, more significantly, that the consistency index is not independent of the number of characters or the number of taxa, but that it decreases with increasing numbers of taxa or characters. He proposes that the homoplasy excess ratio (HER) is an improved way to measure the relative amounts of homoplasy in different data sets. The HER is calculated by dividing the difference between the mean length of random trees and the length of the tree calculated from nonrandom data by the difference between the mean of the random trees and the minimum possible length of the nonrandom tree. If there is no homoplasy in the data set, then the HER is 1.0 and for extremely homoplaseous data, the HER approaches 0.0. For the grass data, the consistency index was found to be 0.695 and the homoplasy excess ratio 0.600.

Ribosomal RNA sequences and angiosperm radiation. The ribosomal RNA sequences from 60 different plant taxa, including the grasses mentioned above, have been used to infer the phylogenetic relationships within the flowering plants and within the seed-bearing plants. Of the 60 taxa studied, 12 are gymnosperms and 46 are angiosperms. The other two taxa are *Equisetum* (horsetail) and *Psilotum*, both of which are seedless vascular plants used as outgroups for the purpose of assigning character state polarity. The gymnosperms include *Ginkgo* and representatives of conifers and cycads as well as representatives of all three genera of Gnetales (*Ephedra*, *Welwitschia* and *Gnetum*). The angiosperms sampled here are divided into 17 monocot genera and 29 dicot genera which include members of the Nymphaeales, Piperales, Magnoliales and Aristolochiales (all orders of the subclass Magnoliidae) and representatives of the subclasses Rosidae, Hamamelidae and Caryophyllidae.

Some basic features of the sequence data for all 60 taxa are summarized in Table 7. Of the 1701 nucleotide sites from the 18S and 26S rRNA molecules, 1097 were constant and 604 were variable. Only 417 of the 604 variable sites were phylogenetically informative. The remaining variable sites were autapomorphies, the large majority of which occurred within the 58 ingroups. All 13 gap sites were variable and informative. Thirty percent of the 18S sites were variable and 20% of the 18S sites informative. Forty-five

Table 7. Summary of rRNA data over 60 taxa. Tn=transition, Tv=transversion, MH=mutilply hit.

Primer	Region	Sites	Variable	Tn	Tv	MH	Informative	Tn	Tv	MH
18E	90-308	191	98	29	14	55	77	18	8	51
18G	300-554	250	52	19	14	19	37	11	8	18
18H	910-1134	215	41	18	8	15	25	9	3	13
18J	1210-1429	214	66	22	17	27	43	15	6	22
18L	1535-1766	227	75	35	20	20	42	21	7	14
18S total		1097	332	123	73	136	224	74	32	118
26C	740-949	202	58	17	11	30	38	8	2	28
26D	1625-1836	202	108	41	25	42	80	30	14	36
26F	1960-2172	200	106	46	12	48	75	32	2	41
26S total		604	272	104	48	120	193	70	18	105
18S+26S		1701	604	227	121	256	417	144	50	223
Gaps		13	13	-	-	-	13	-	-	-
GRAND TOTAL		1714	617	227	121	256	430	144	50	223

percent of the 26S sites were variable and 32% were informative. More than half of the variable sites and more than half of the informative sites were multiply hit (both transitions and transversions had occurred at such sites). The overall ratio of transitions-to-transversions was 1.9 to 1 in the variable sites, but within the informative sites there was a transitions-to-transversions ratio of about 3 to 1. The most variable regions were those sequenced with the 18E, 26D and 26F primers.

The number of taxa in the data set is so large that the only available tree inference option in PAUP is the heuristic search. Using the tree bisecting and reconnection swapping option and the simple sequence addition option, PAUP found the shortest tree to be 1870 steps with an overall consistency index of 0.390. There were at least twenty different variations of the shortest tree. When the search was started again and an option was chosen in PAUP to save all trees that were one step longer than the shortest tree (i.e., to save the trees of length 1870 and 1871 steps), PAUP actually found seven trees that were 1869 steps long. Normally PAUP only performs branch swapping on trees of minimal length, and in the case of the first search these were trees of 1870 steps. However, the second search showed that swapping on a nonminimal tree (1871 steps) can lead ultimately to trees that are actually shorter. This second search, which was terminated after five days, found 2358 trees of 1871 steps, 259 trees of 1870 steps and seven of 1869 steps.

The data were then converted by PAUP into a format for input into

Hennig86. Surprisingly, Hennig86 found two trees that were 1867 steps long, two steps shorter than the shortest trees found by PAUP. In past experiments with as many as 57 taxa, Hennig86 had found the same shortest trees as had been found by PAUP. Hennig86 also found two trees that were 1868 steps long. One of the two trees of 1867 steps (the most parsimonious) was then used as a starting topology for branch swapping in PAUP to see if PAUP could find other trees of 1867 steps or if PAUP could rearrange the 1867-step tree to a still shorter tree. PAUP could only find the other tree of 1867 steps found by Hennig86. However when the two trees of 1868 steps were used as beginning topologies in separate searches, PAUP ultimately identified thirty trees that were 1868 steps long. Several of the 1869 trees were used as beginning swapping points in later PAUP searches and all the resulting trees combined into one large file. The condense option of PAUP was used to ensure that all the trees were unique and, after condensation, a total of 3413 trees with overall lengths between 1867 and 1871 were found. Memory limits of the Macintosh computers (4.5MB) prevented further searching, there being too many trees five steps longer than the shortest trees.

The 3413 trees break down into 2358 at 1871 steps, 666 at 1870 steps, 357 at 1869 steps, 30 at 1868 steps and two at 1867 steps. PAUP was unable to find any more trees of 1868 or 1867 steps, but there are more trees other than those already identified at lengths greater than 1868 steps.

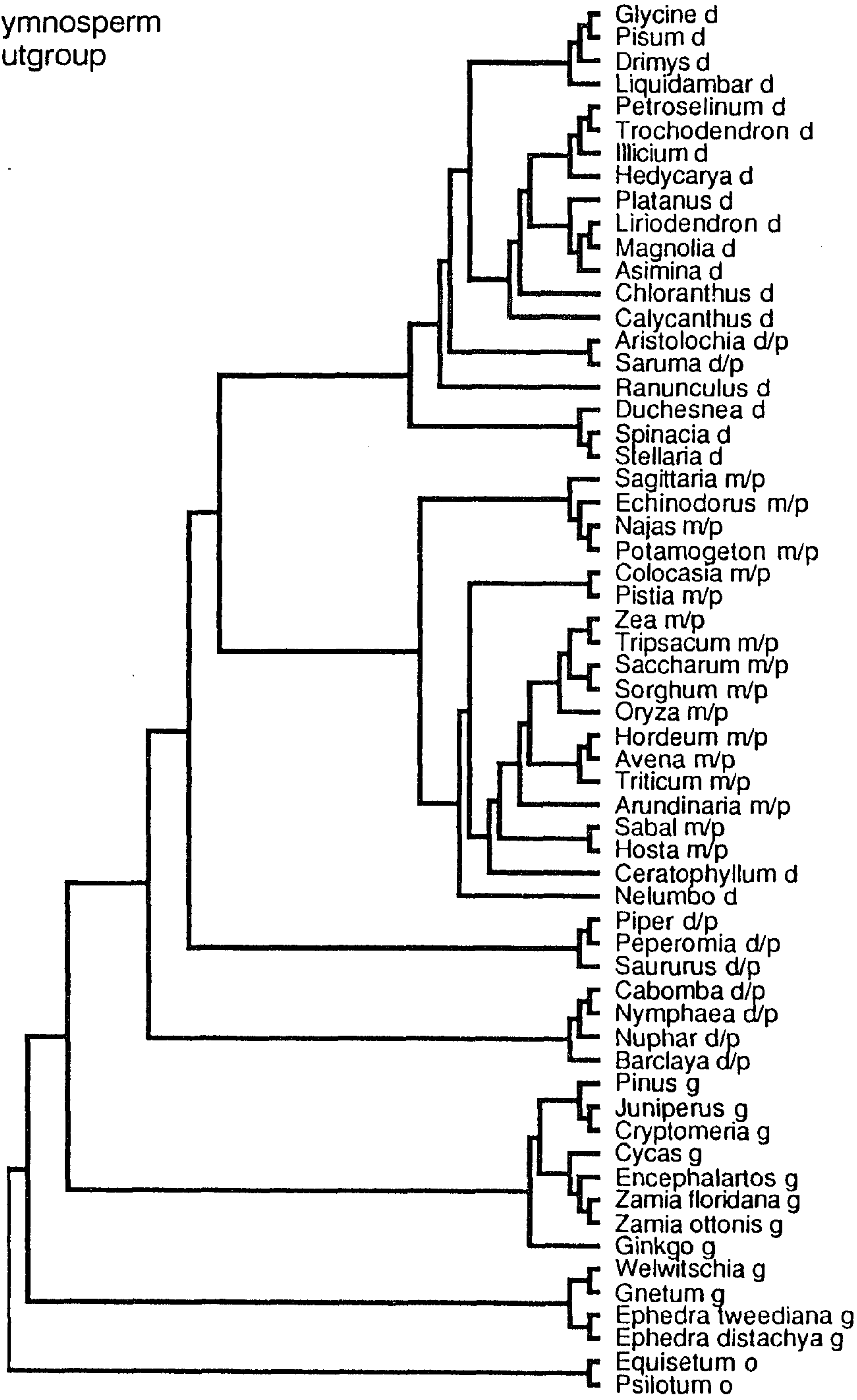
This is certain because in the search for trees less than 1871, the program was terminated while swapping on tree #748 out of the more than 3000 saved. Unfortunately, trees were being accumulated at the rate of about 500 a day, but PAUP was only able to swap on about 150 a day (the run was stopped on the fifth day) and PAUP ran the danger of running out of memory, in which case all accumulated trees would have been lost.

One of the two shortest trees of 1867 steps is shown in Figure 15. The only difference between this and the other of the shortest trees is in the placement of *Sorghum* relative to *Saccharum*. In one tree they form a monophyletic group that is the sister group to the group which contains *Zea* and *Tripsacum*. In the other, they form a grade with *Sorghum* in between *Saccharum* and the monophyletic grouping of *Zea* and *Tripsacum*. All other features of the two most parsimonious trees are identical.

In the most parsimonious trees, the gymnosperms do not form a monophyletic group, but the angiosperms are found to be a natural group. The Gnetales are shown to be the most primitive gymnosperms. Cycads, *Ginkgo* and conifers form a monophyletic group which is the sister group of the angiosperms. Within this monophyletic group, the rRNA sequence data suggest that *Ginkgo* diverged first from the common ancestor it shared with conifers and cycads. As expected, all members of the conifers form a monophyletic group as do all members of the cycads. Within the Gnetales, *Welwitschia* and *Gnetum* are indicated to have shared a common ancestor

Figure 15. One of two equally parsimonious trees found for 60 taxa based on rRNA sequence data. Length = 1867 steps.

m = monocot
d = dicot
p = paleoherb
g = gymnosperm
o = outgroup



with one another more recently than either has with *Ephedra*.

The most parsimonious trees place members of the Nymphaeales and Piperales at the base of angiosperm radiation. In these trees the Nymphaeales include the families Barclayaceae, Nymphaeaceae and Cabombaceae, but do not include Ceratophyllaceae and Nelumboaceae. The Piperales include the families Saururaceae and Piperaceae, but not Chloranthaceae. In the shortest trees, *Ceratophyllum* and *Nelumbo* are found clustered among the monocots and *Chloranthus* is found to be a more derived taxon than Piperaceae and Saururaceae.

After the Nymphaeales and Piperales, the rest of the flowering plants split into two sister groups, one of which contains all the monocots plus Ceratophyllum and Nelumbo, and the other of which is composed of the rest of the Magnoliidae and the other dicots - Caryophyllidae, Hamamelidae and Rosidae.

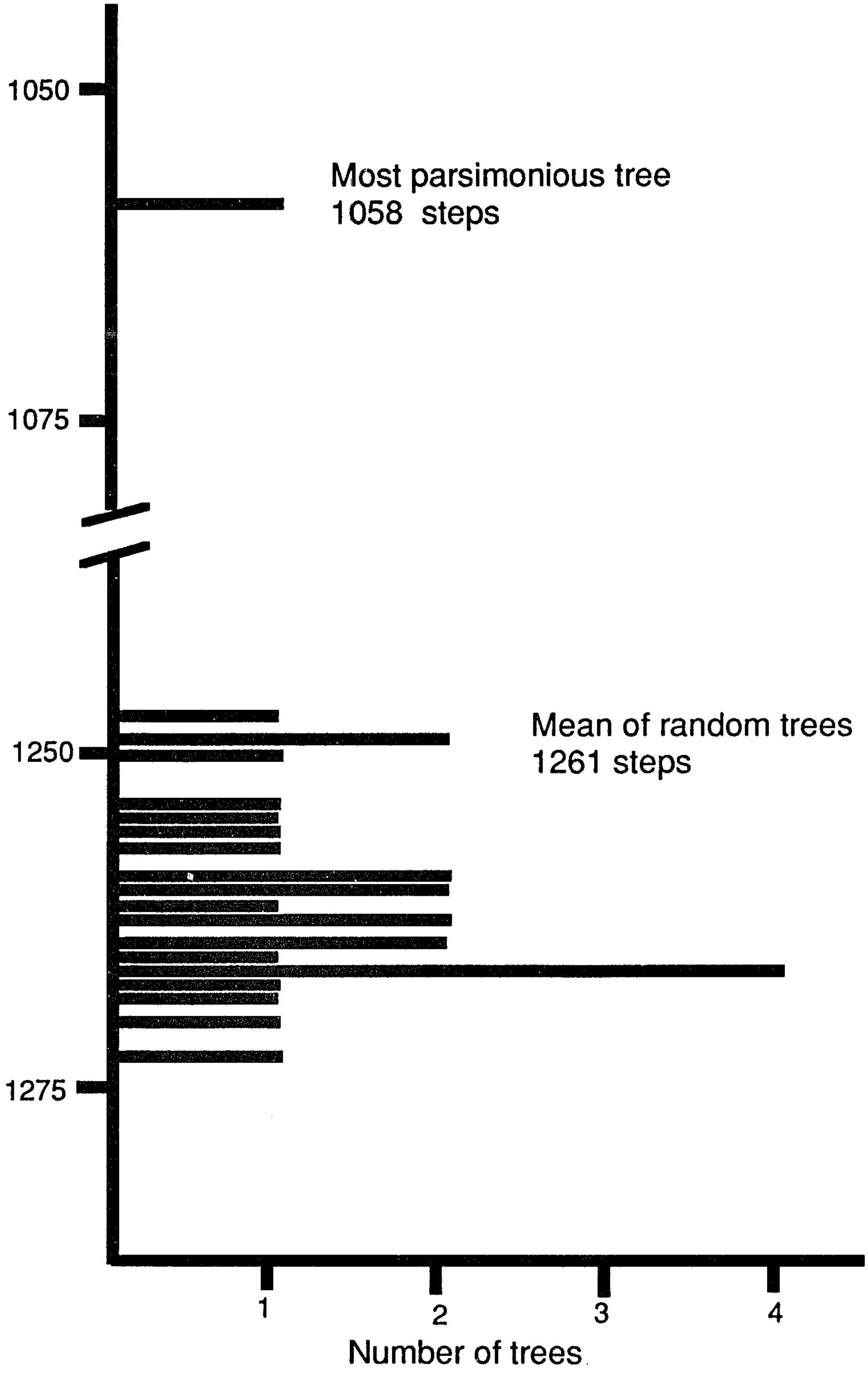
The neighbor-joining phenetic analysis was performed on the complete data set with distances calculated by the same three formulae used for the grasses. This time the results were not independent of the manner in which the distances were calculated, nor were the topologies of any of the trees identical to that of the shortest cladogram, although the topologies are very similar. The Jukes-Cantor distances gave the same tree that the Kimura distances did, but this tree was different from that based on distance calculated by the number different divided by the number compared (the overall dissimilarity). The topologies of the phenograms gave cladistic trees with lengths of 1907 steps for the one based on overall dissimilarity and 1909 steps for the Jukes-Cantor and Kimura distances. The data matrices are printed in Appendix 1.

A subset of the 60 taxa was used in a bootstrap run to test the robustness of the trees. Only 40 taxa were used because of the large amount of time required to complete 100 PAUP replications. The data were reduced to 40 taxa by eliminating duplicate members of some families (e.g., seven of the nine grasses were eliminated), and by deleting some members from consistently monophyletic groups (from earlier analyses) like *Pinus* of the conifers. With 40 taxa, it took almost 21 days on a Macintosh IIcx computer to complete 100 runs. The only nodes supported in excess of 95% of the runs were the node uniting the Gnetales (99 times out of 100) and the node uniting all the angiosperms into a monophyletic group (100 times out of 100). The other group of gymnosperms (cycads, *Ginkgo* and conifers) formed a monophyletic group 91 times out of 100.

In a second bootstrap, rather than simply eliminate taxa, another tactic was employed. Sequence data from taxa which were consistently placed into monophyletic groups in earlier analyses were condensed into one representative of the entire group. The data were condensed by choosing the consensus of each of the contributing taxa at each site. If all the taxa being condensed showed a G at a particular position, then the condensed taxon was assigned a G. If some taxa had a G and others an A at a site, then the condensed taxa was assigned the uncertain condition G or A. In the second bootstrap, the nine grasses were condensed into one taxon, three of the Piperales (*Piper*, *Peperomia* and *Saururus*) into one, four of the Nymphaeales (*Nymphaea*, *Cabomba*, *Nuphar* and *Barclaya*) into one, three of the four cycads into one, the two of the three conifers into one, the three Magnoliales (*Magnolia*, *Liriodendron* and *Asimina*) into one, the two legumes (*Glycine* and *Pisum*) into one, the two Caryophyllidae (*Stellaria* and *Spinacia*) into one, the four Alismatidae (*Echinodorous*, *Sagittaria*, *Najas* and *Potamogeton*) into one and the two Arales (*Colocasia* and *Pistia*) into one. This condensation reduced the number of taxa to 34 and the resulting data set was bootstrapped 100 times.

Archie's (1989a) randomization program is limited to thirty taxa, so another subset of the data which contained representatives of all the major groups among the 60 taxa was chosen for the randomization process and calculation of HER. The results of this randomization are summarized in Figure 16. With the 30 taxa chosen, PAUP found a most parsimonious tree of 1058 steps. When the data set was randomized 25 times and each subsequent data set analyzed by PAUP, the shortest trees ranged in length from 1247 to 1273 with a mean length of 1261.0 steps and a standard deviation of 6.6 steps. As in the analysis with the grass data, none of the randomized data sets produced a shortest tree near to that of the

Figure 16. The distribution of trees found after 25 randomizations of the rRNA data set of 30 taxa.



nonrandom data. With the larger data set, the shortest tree was more than 30 standard deviations shorter than the average randomized tree. The HER was found to be 0.274 in contrast to a consistency index of 0.49.

The robustness of the various nodes of the shortest tree were tested as in the analysis of the grass data, by constructing majority-rule consensus trees as groups of less parsimonious trees were combined with more parsimonious ones. Majority-rule consensus trees were calculated with the 1867-step and 1868-step trees, the 1867-1869 steps, 1867-1870 steps and 1867-1871 steps trees. The series of consensus trees is presented in the next section.

DISCUSSION

PATTERNS OF CHANGE. There are a few discernable patterns in the changes of the rRNA sequences throughout evolution in the seed plants, and in the more limited evolutionary study of the grasses. In both sets of data, the 18E, 26D and 26F regions are more variable than the other regions. In the complete data set, the 18H region is the most conserved and 18G the next most conserved. In the grass data, 18G is the most conserved region. It is interesting to note that secondary structure calculations (Gerbi *et al.*, 1985; Gutell and Fox, 1988) predict that the 18E region and 26F region both are within expansion segments (Clark *et al.*, 1984); the other primer regions lie completely or mostly in regions of more conserved structure. Therefore sequencing of additional regions in expansion segments offers the potential for higher resolution at lower taxonomic levels. Some of these primers (18K, 18P, 26B and 26J) are available. The primary sequence variation mirrors the secondary structure conservation patterns.

Of the variable positions, that is, those that contribute to the length of the tree (this includes autapomorphies which contribute to the length without contributing to the tree structure), about 42% had experienced both transition and transversion events. These sites are said to be multiply hit. Within the grasses, only about 13% of the variable sites were multiply hit. It is to be expected that during the differentiation of the grasses over the last 60 mya or so (Wolfe, *et al.*, 1989), fewer sites would undergo a second or third change when compared to the evolution of rRNA sequences over the 350 myr since the emergence of the gymnosperms.

In the complete data set, there were 58 variable sites that were four-state, that is, all four states of G, A, T and C were represented at that particular site among the 60 taxa. One hundred ninety-four positions had three states present, and the remaining 365 nucleotide positions were binary. The ratio of transitions-to-transversions was about 2 to 1 overall for both data sets. This represents the minimum number of changes that must have occurred to account for the present distribution of character states. The actual number of observed transition and transversion events can be determined only by looking at the true phylogenetic tree. In the most parsimonious tree of the rRNA data for all 60 taxa, there are postulated to be 1156 transition events and 691 transversion events, a ratio of 1.673 to 1.0. (The numbers of events do not sum to 1867 - the number of events in the most parsimonious tree - because some events could not be determined accurately, i.e., if a nucleotide was scored as uncertain, then it might not be possible to determine whether a transition or transversion had occurred at some terminal taxon, and these were eliminated from the calculation.) Within the circumscribed investigation of the grasses and *Colocasia*, the shortest tree suggests that 117 transition events and 70 transversion events occurred, for a total of 187 steps. Remarkably, the transition-to-transversion ratio within the grasses and *Colocasia* was 1.671 to 1.0, essentially identical to the overall ratio for all 60 taxa. Taken together, the similarities with respect to the most conserved and variable regions, and with respect to the minimum and postulated ratios of transition-to-transversion within the narrow range of the grasses and within the entire data set, suggest that the pattern of change of rRNA has been fairly consistent throughout the diversification of the seed plants.

Although the patterns of change may have been consistent, the rates at which these changes occur in different lineages may not be constant. Figure 17 is a phylogram of the shortest grass tree and Figure 18 is a phylogram of the most parsimonious tree for the entire 60 taxa. In a phylogram, the length of each branch is proportional to the number of changes that have occurred along that branch. In the grass phylogram, the number of changes which are postulated to have occurred along each branch, the branch length, is printed above each branch. If the relative rate of evolution of the rRNA molecules was constant in each lineage, then the sum of the length of each branch connecting the common ancestor of a group of taxa to the terminal taxa would be the same for each taxon in the group. More simply, if the rates of rRNA change were constant, the terminal taxa in Figures 17 and 18 would align on the right-hand side of the pages on which they are printed. That the terminal taxa do not align, then, suggests that the rRNA of seed plants is not evolving in a completely clocklike manner

Figure 17. The most parsimonious arrangement for Poaceae shown as a phylogram. The number above each branch, the branch length, represents the number of characters which changed along the branch.

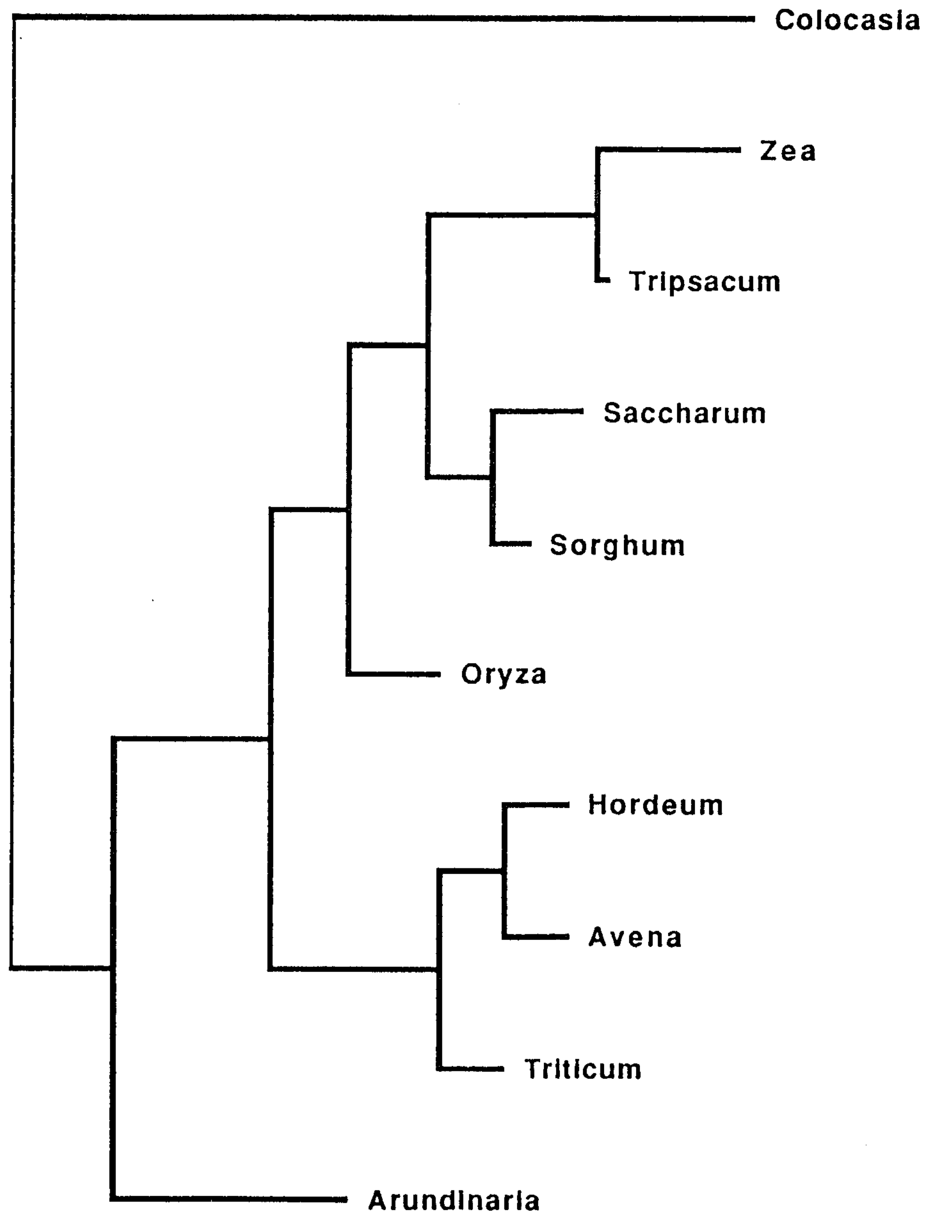
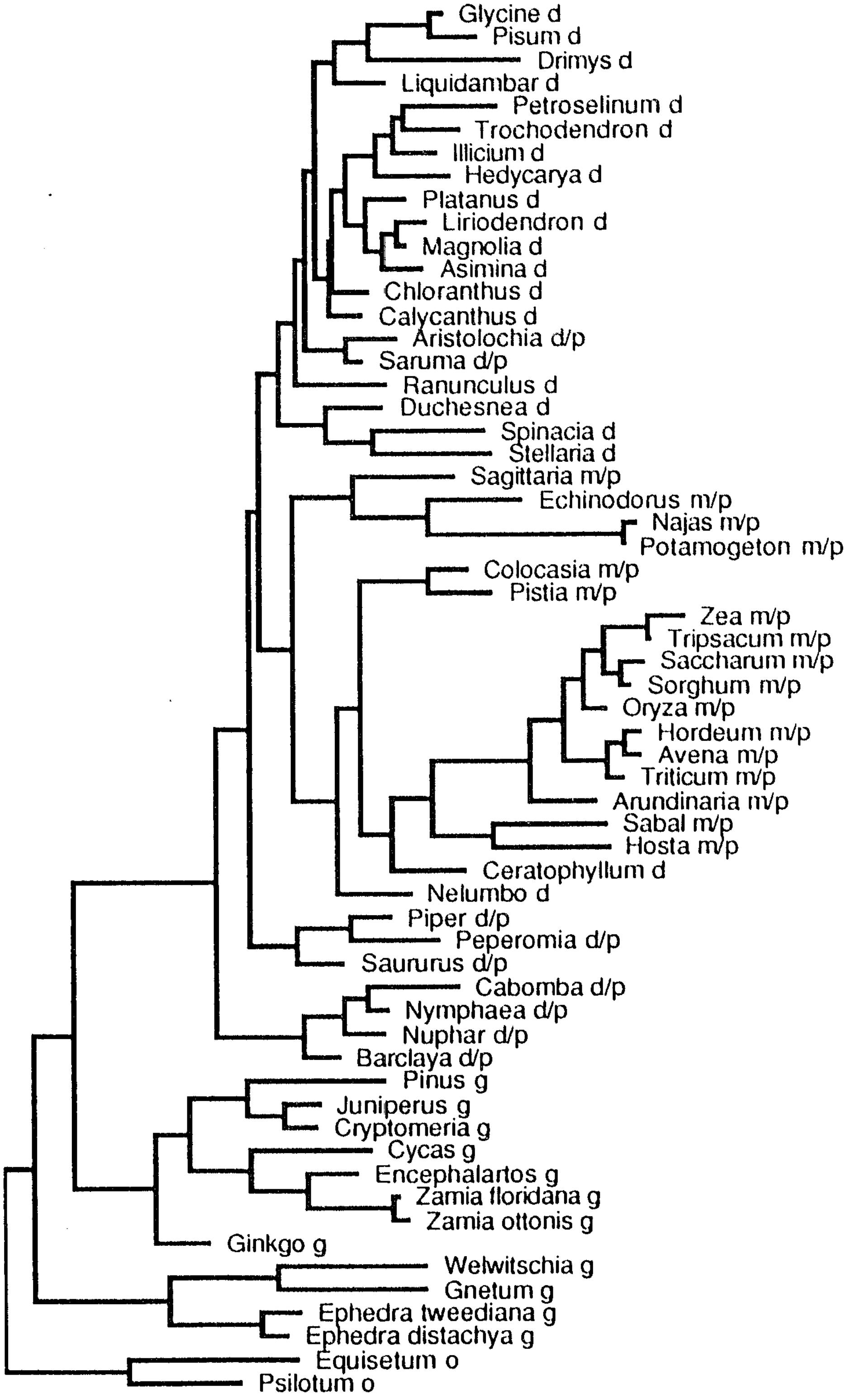


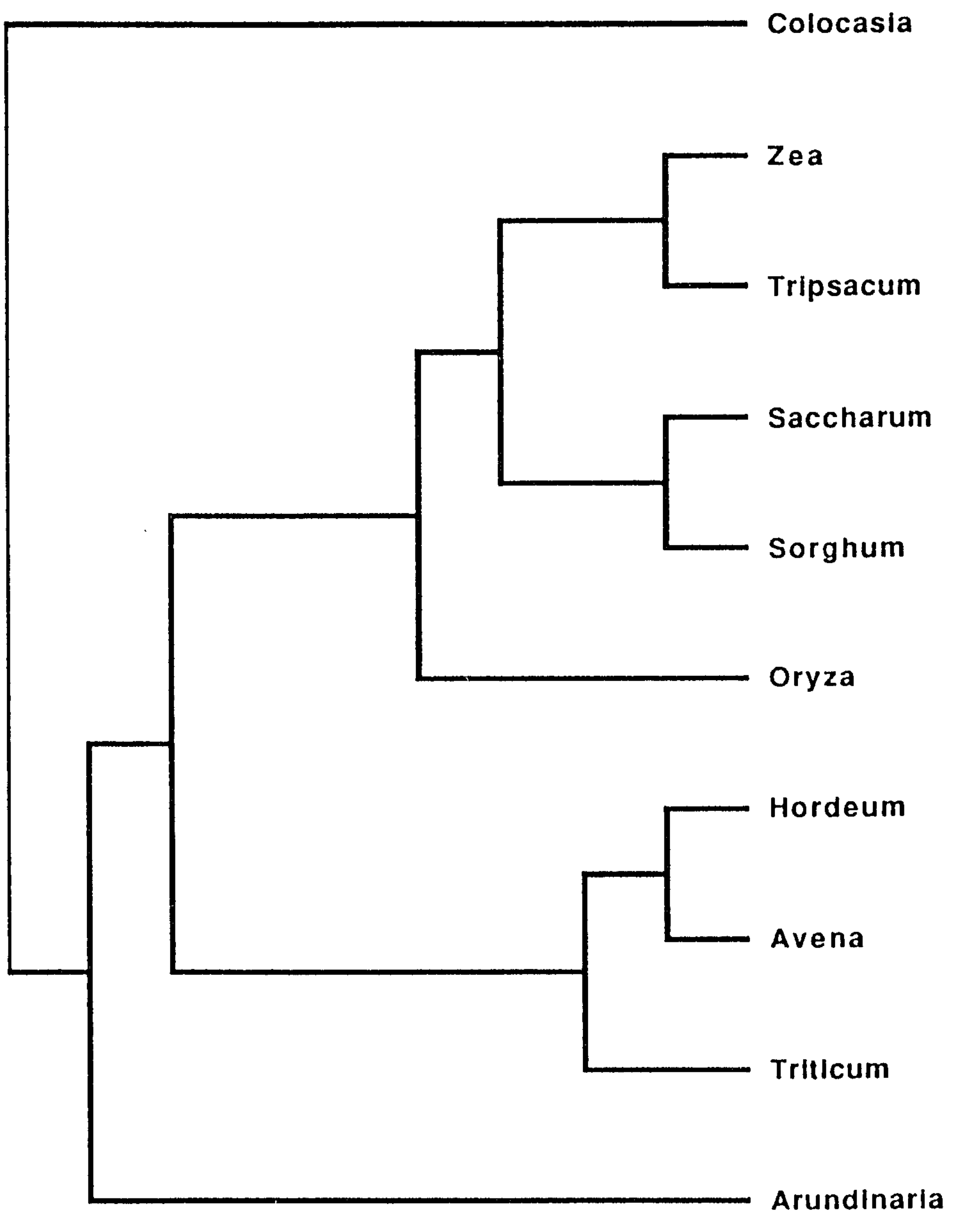
Figure 18. The most parsimonious arrangement for 60 taxa shown as a phylogram.



along all lineages. This is confirmed by looking at the number of changes necessary to go from the common ancestor of the seed plants to *Magnolia* which is 99, and to go from the same common ancestor to *Nymphaea*, *Glycine*, *Peperomia* or *Zea*, which requires 102, 111, 120 and 193 steps, respectively. It has been shown recently that the rRNA of bivalve molluscs are not evolving in a clocklike manner (Bowman, 1989), and that the cytoplasmic rRNA molecules of green algae are not evolving at similar rates (Zechman *et al.*, 1990), although these findings are not necessarily transferable to seed plant rRNA. The absolute rate of change of rRNA cannot be determined without an extensive fossil record to calibrate the molecule, and this is not available for seed plants.

RIBOSOMAL RNA SEQUENCES AND EVOLUTION IN POACEAE. The most parsimonious arrangement of nine grass genera based on their rRNA sequences is shown in Figure 19. *Colocasia*, a genus within the family Araceae, was used as the outgroup because, in preliminary analyses with a greater range of taxa, the Araceae were consistently placed as the sister group of the Poaceae. In the shortest tree, the first branch off the tree leads to *Arundinaria*, and the remaining eight taxa then split into two monophyletic groups, one of which contains *Avena* (oats), *Triticum* (wheat) and *Hordeum* (barley) while the other contains *Zea* (maize), *Tripsacum*, *Sorghum*, *Saccharum* (sugarcane) and *Oryza* (rice). Within these two monophyletic

Figure 19. The most parsimonious tree for Poaceae inferred from rRNA sequence data. (Identical to Figure 10.)



groups, several smaller natural groups can be identified: *Avena* and *Hordeum*; *Zea* and *Tripsacum*; *Sorghum* and *Saccharum*; and the latter four together.

The results of the bootstrapping indicate that the best supported nodes on the rRNA tree are the ones that unite the other eight taxa to the exclusion of *Arundinaria* (95%), the one that joins *Zea* and *Tripsacum* (100%), the one that joins *Avena*, *Hordeum* and *Triticum* (100%), the one that allies *Avena* and *Hordeum* (92%) and the one that allies *Zea* and *Tripsacum* with *Saccharum* and *Sorghum* (91%). The bootstrap results also indicate that the placement of *Saccharum* relative to *Sorghum* was questionable and that the placement of *Oryza* relative to all the other grasses except *Arundinaria* was equivocal.

Combining less parsimonious trees with the more parsimonious ones to construct majority-rule consensus trees also showed which nodes were the best supported by the data. This series of majority-rule consensus trees is presented in Figure 20. Figure 20a is the most parsimonious tree, Figure 20b is the majority-rule consensus calculated after combining the shortest tree with the one tree that was 188 steps. Figure 20c is the majority-rule consensus calculated from all trees with a length between 187 and 189 steps, etc. This series of trees showed the same pattern of conservation indicated by the bootstrapping results above. The first nodes to collapse, with the addition of less parsimonious trees to make the consensus, were the ones

Figure 20. The majority-rule consensus trees for Poaceae calculated for the indicated ranges of trees. Each node is labelled by its frequency of appearance among the trees from which the consensus was calculated, i.e., 100 means that in 100% of the trees used to calculate the consensus, the node apperars exactly as it does on the consensus tree.

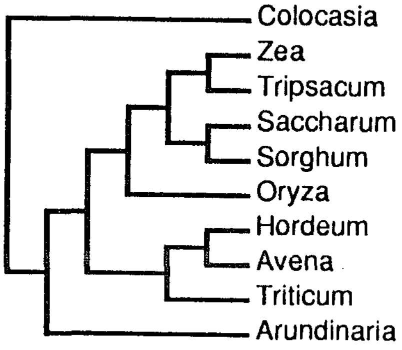


Figure 20a. The most parsimonious tree.

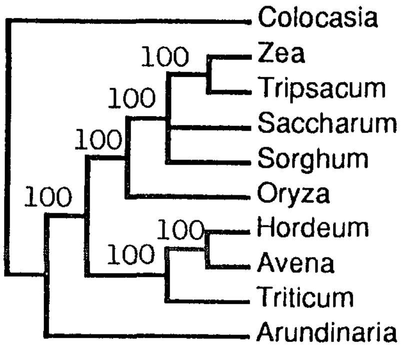


Figure 20b. Majority-rule consensus for 2 trees between 187-188 steps.

Figure 20c. Majority-rule consensus of 5 trees between 187-189 steps.

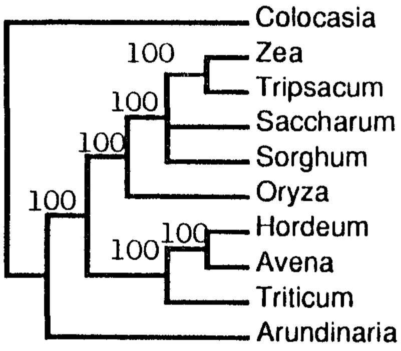


Figure 20d. Majority-rule consensus of 8 trees between 187-190 steps.

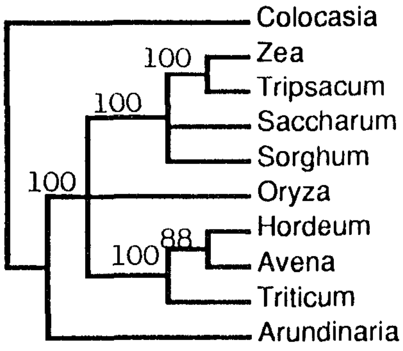


Figure 20 (con'd). N.B. The majority-rule consensus trees for the tree up to 195, 196 and 197 stipes have exactly the same topology as that of Figure 20h. The only idfferences are that the node labelled 91% in Figure 20h drops to 82, 80 and 76%, and the node labelled 74% in Figure 20h drops to 71, 62 and 59%, respectively, as the consensus includes the trees of 195, 196 and 197 steps.

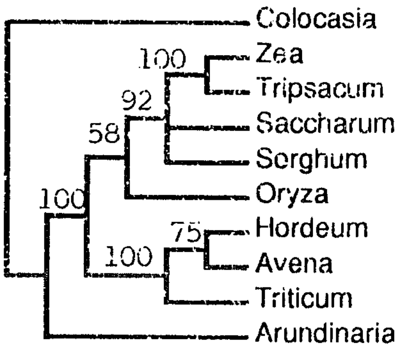


Figure 20e. Majority-rule consensus for 12 trees between 187-191 steps.

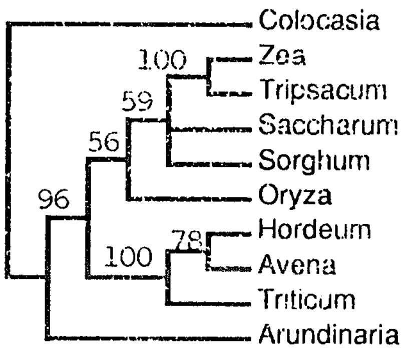


Figure 20f. Majority-rule consensus for 27 trees between 187-192 steps.

Figure 20g. Majority-rule consensus for 49 trees between 187-193 steps.

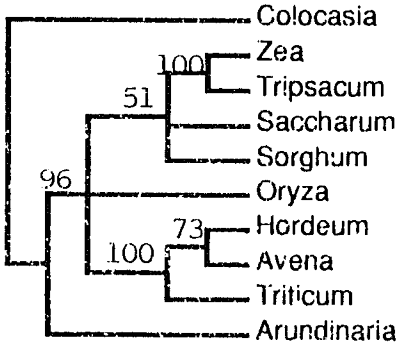
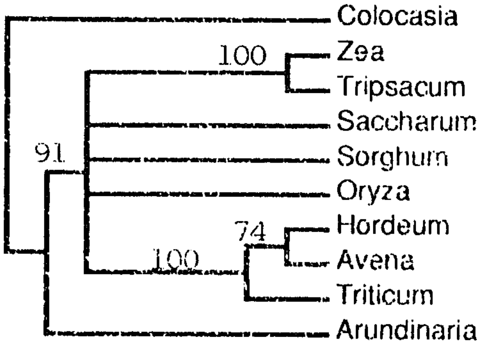


Figure 20h. Majority-rule consensus for 70 trees between 187-194 steps.



which joined *Saccharum* to *Sorghum* and the one that placed *Oryza* near *Zea*, *Tripsacum*, *Sorghum* and *Saccharum*. The most strongly supported nodes, the one that placed *Zea* and *Tripsacum* in a monophyletic group and the one that placed *Hordeum*, *Avena* and *Triticum* in a natural group were present in all 227 trees up to 10 steps longer than the shortest tree. The other nodes deteriorated at different points in the series of majority-rule trees.

Randomizing the data with Archie's (1989a) program revealed that the data are indeed more informative than random data. Although it may seem obvious that the actual sequence data should be more informative than random data, this is not always true. Archie (1989c) took the plant DNA sequence data of Martin *et al.* (1985) as analyzed by Bremer (1988) and randomized the data and then found the shortest trees with each random data set. He found that a significant fraction of those random data sets actually yielded trees that were shorter than the nonrandom data. His results did not show that the data were uninformative, necessarily, but that they were inappropriate for the level of taxonomic rank being investigated. Those DNA sequences may have been informative over a more circumscribed range of divergence. With the rRNA sequences, all 100 random data sets yielded trees that were longer than the most parsimonious tree based on the nonrandom data. The fact that all 100 random trees were at least 36 steps (or 20%) longer than the nonrandom tree and that the nonrandom tree was more than 19 standard deviations removed from the mean of the randomized trees, would seem to indicate that there is a fairly high level of information contained in rRNA sequence data of the grass family. This was confirmed by the homoplasy excess ratio (HER) determined for the rRNA data which was 0.600, and means that the data have about 40% homoplaseous characters (Archie, 1989b). This HER compares favorably with HERs of other, similarly sized (approximately the same number of taxa and/or characters) protein- and nucleotide-sequence data sets used in comparative systematics studies (Archie, 1989b).

Comparing results to other classifications. There is a high level of consistency between the most parsimonious tree based on rRNA data (Figure 19, p. 113) and the classifications of Gould and Shaw (1985) which recognize six different subfamilies within the grass family, Poaceae. Based on morphological and nonmorphological (e.g., biochemical and genetic) similarities, they propose that the Poaceae can be divided into the subfamilies Pooideae, Panicoideae, Chloridoideae, Bambusoideae, Arundinoideae and Oryzoideae. They place *Zea*, *Tripsacum*, *Sorghum* and *Saccharum* within Panicoideae, *Avena*, *Triticum* and *Hordeum* within Pooideae, *Oryza* within Oryzoideae, and *Arundinaria* within Bambusoideae. The shortest rRNA tree is consistent with this scheme except that it places *Avena* and *Hordeum* as more closely related to one another than either is to *Triticum* while Gould and Shaw place *Hordeum* and *Triticum* in the tribe Triticeae and *Avena* in Aveneae.

On the other hand, the rRNA tree does not support the classification of Watson and coworkers (1985) who recognize only five subfamilies of grasses. They place the genus *Oryza* in the tribe Oryzaneae within the subfamily Bambusoideae. If the rRNA data supported their classification, *Oryza* and *Arundinaria* would form a monophyletic group somewhere on the tree. However the rRNA data suggest that these two taxa are not closely related. The other groupings on the most parsimonious rRNA tree are consistent with Watson *et al.*'s arrangement, except as in the discussion of the Gould and Shaw classification, for the relationship of *Triticum* relative to *Hordeum* and *Avena*.

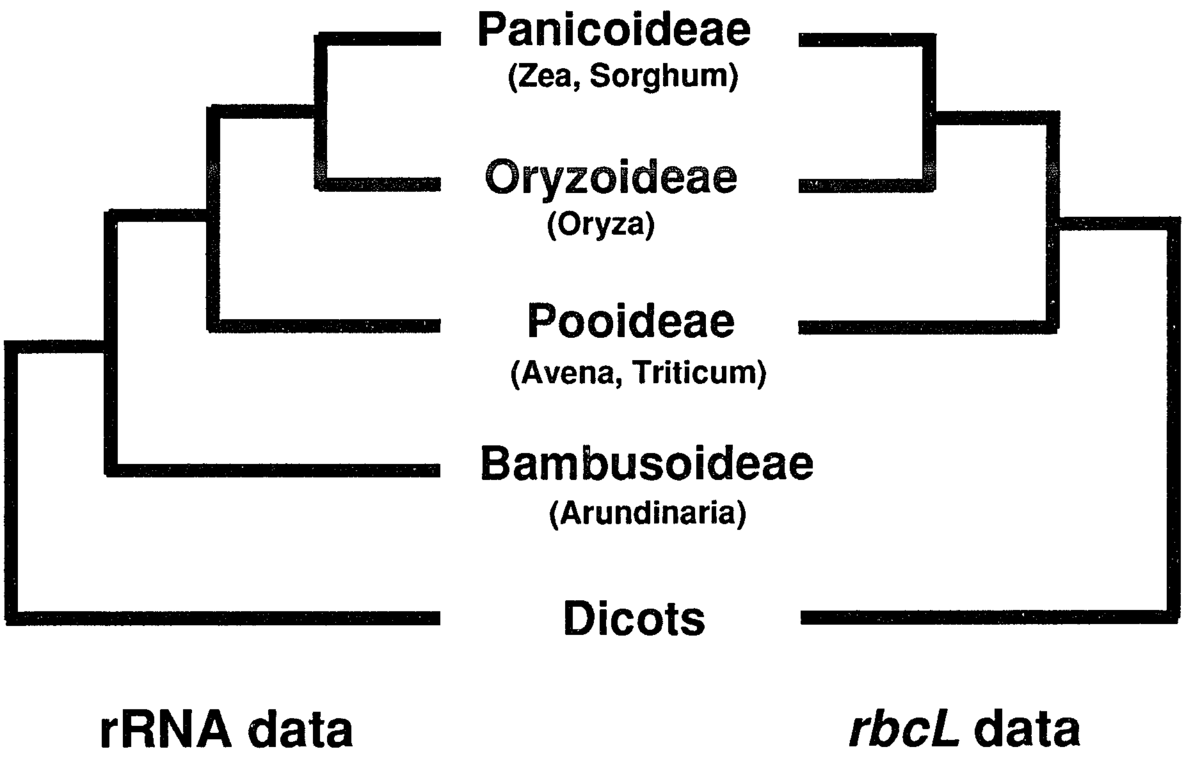
Wolfe and colleagues (1989) compared the sequences of three chloroplast genes of certain members of the grass family and found that the Panicoideae grouped together and that the Pooideae grouped together. Their analysis did not resolve the position of *Oryza* relative to the Panicoideae and Pooideae groups.

The three classifications mentioned above are all based on phenetic analyses, from which one cannot necessarily infer an evolutionary relationship. A cladistic analysis based on some of the same characters used by Watson *et al.* (1985) showed that the Pooideae, Panicoideae and Bambusoideae (including the tribe Oryzaneae) were each monophyletic assemblages (Kellogg and Campbell 1987), while the monophyly of some of the other grass subfamilies was doubtful. The rRNA data are consistent with the morphological data with respect to the Pooideae and Panicoideae, but not with respect to *Oryza* and *Arundinaria*.

The shortest tree based on rRNA data is congruent with another recent cladistic analysis of molecular sequence data within the grasses (Doebley *et al.*, 1990). In this study, sequences of the *rbcL* gene which codes for the large subunit of ribulose bisphosphate carboxylase were compared among Panicoideae, Pooideae and *Oryza* (Doebley *et al.* do not have any *Arundinaria* species in their analysis). Figure 21 is a comparison of the shortest trees from parsimony analyses of the rRNA and *rbcL* sequences. It shows that the monophyletic groups and the branching order in both trees, one based on nuclear-encoded rRNA and the other based on chloroplast-encoded *rbcL*, were identical.

For the most part, rRNA sequences have been used successfully to resolve relationships within the grass family, at least at the subfamily level, and can probably be used to resolve the subfamily relationships within any other plant family whose age is on the order of the Poaceae, about 50-70 Myr (Wolfe *et al.*, 1989). At the tribal level, the rRNA data did not group the two members of the Triticeae together relative to *Avena*. It is possible that this is due to hybridization within this tribe (Kellogg and Campbell, 1987). The bootstrapping and majority-rule consensus trees showed that while the grouping of the Pooideae was strongly supported, the alliance of *Avena* and *Hordeum* to the exclusion of *Triticum* was not so strongly supported. The

Figure 21. A comparison of an arrangement based on rRNA sequence data and an arrangement based on *rbcL* sequence data by Doebley *et al.*, 1990.



rRNA sequence data do not place *Oryza* with *Arundinaria*, though *Oryza*'s placement has been shown to be quite variable with the addition of each new taxon. It is possible that there are simply not enough informative sites yet to unequivocally place *Oryza*, or that it will be necessary to add other representatives of Oryzoideae and Bambusoideae before *Oryza*'s position can be fixed. Results after the addition of *Secale, Brachyelytrum* and *Diarrhena* (none of which are Oryzoideae or Bambusoideae) and the addition of sequences from two more regions of the 26S molecule show the positions of *Oryza* and *Arundinaria* to be unchanged and the Pooidae and Panicoidae to remain natural groups (Issel *et al.*, 1990).

RIBOSOMAL RNA SEQUENCES AND ANGIOSPERM RADIATION There were two equally parsimonious arrangements of the shortest tree constructed based on sequence data from the rRNA of 58 seed plants and two seedless plants. These trees were 1867 steps long and differed only in the placement of *Saccharum* relative to *Sorghum*: in one arrangement (Figure 22) they are sister taxa, in the other (Figure 23) *Saccharum* and *Sorghum* form a grade between *Oryza* and the monophyletic group of *Zea* and *Tripsacum*. All other features of the two topologies are identical. In the discussion to follow, I refer to the shortest tree or the most parsimonious tree as though there were only one version of this tree rather than two.

Figure 22. One of the two most parsimonious trees for 60 taxa based on rRNA sequences. Length = 1867 steps.

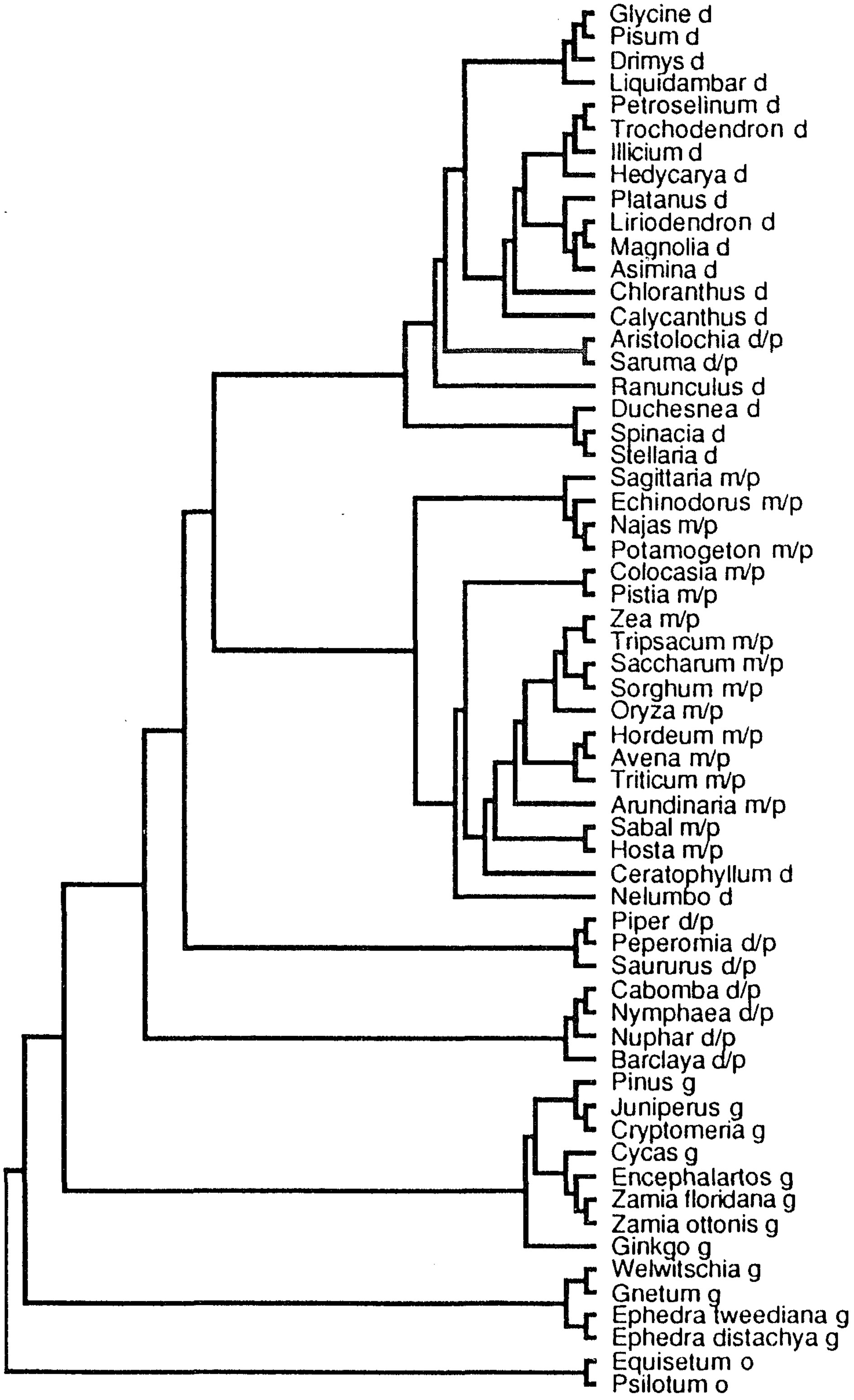
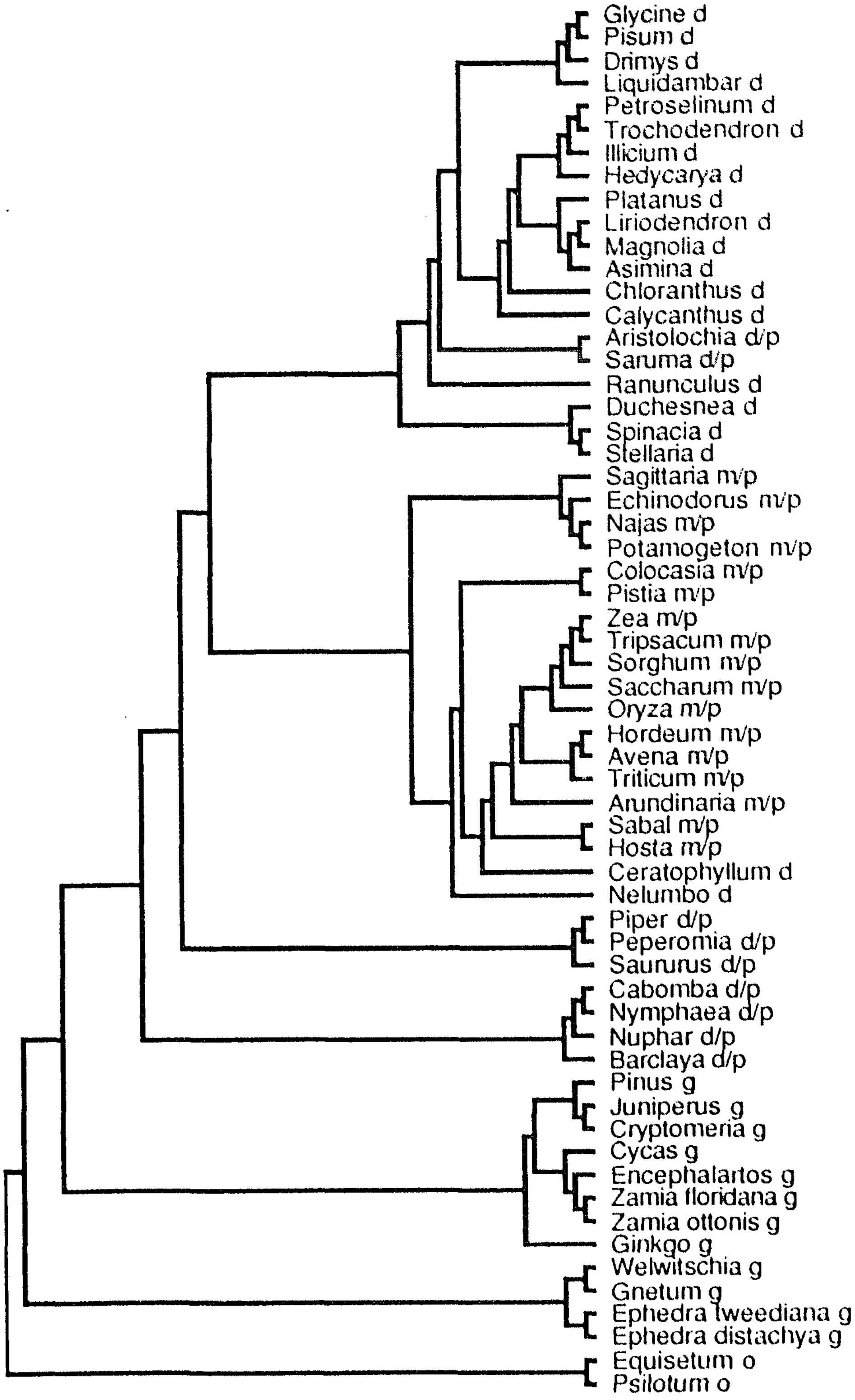


Figure 23. One of the two most parsimonious trees for 60 taxa based on rRNA sequences. Length = 1867 steps.



Parsimony trees.

The relationship between extant gymnosperms and angiosperms. In the most parsimonious tree, the gymnosperms are divided into two separate natural groups: one of these groups consists of the three genera of the order Gnetales, and the other is composed of the three conifers (*Pinus*, *Cryptomeria* and *Juniperus*), *Ginkgo*, and the four cycads (*Cycas*, *Encephalartos*, and two *Zamias*). The three conifers form a natural group, as do the four cycads, and the conifers and cycads together form another monophyletic group. According to this arrangement, the gymnosperms are not a monophyletic assemblage of taxa, because the common ancestor of all gymnosperms is also an ancestor of the angiosperms. This is not surprising, nor in conflict with most traditional views of the origin of the angiosperms which hold that the flowering plants are derived from within the gymnosperms (Cronquist, 1968; Takhtajan, 1969; Stebbins, 1974). The biological interpretation of the most parsimonious rRNA trees is in accordance with the view that the angiosperms arose from the gymnosperms.

Within the Gnetales, the rRNA data indicate that *Welwitschia* and *Gnetum* are more closely related to one another than either is to *Ephedra*, in agreement with the morphological analyses of Crane (1985) and Donoghue and Doyle (1989a). That the Gnetales themselves are a coherent natural group is unquestionably confirmed by the rRNA sequence data. In bootstrap tests with subsets of the data (some taxa were eliminated in the interest of time), the Gnetales were grouped together in 99 out of 100 replications with 40 of the 60 taxa, and 100 out of 100 replications when certain groups were merged into one, resulting in 34 taxa. Within the other gymnosperm clade, the rRNA data suggest that cycads and conifers are more closely related to one another than either is to *Ginkgo*. The morphologically-based cladistic analyses do not agree with this placement, putting *Ginkgo* and the conifers into a monophyletic group, coniferopsids. An advantage enjoyed by these morphological treatments is the inclusion of numerous fossil taxa, which has been shown to affect the placement of extant taxa (Donoghue *et al.*, 1989). A preliminary examination of the morphological data for just the extant seed plant lineages concurs with the most parsimonious rRNA trees (Donoghue, Doyle and Zimmer, unpublished results). Therefore, it is possible that the relative placement of *Ginkgo* and cycads would change in the rRNA tree if fossil sequences were available.

In the shortest trees, the Gnetales are the earliest diverging seed plants and the other gymnosperms (conifers, cycads and *Ginkgo*) are the sister group of the angiosperms. These results were not in accord with cladistic analyses of morphological data, in which Crane (1985) and Donoghue and Doyle (1989a, 1989b) separately found that of the extant gymnosperms, the Gnetales were most closely related to the flowering plants, united with them by such characteristics as reduced gametophytes and vascular structure. Omitting fossil taxa does not affect the placement of

Gnetales relative to the angiosperms (Donoghue *et al.*, 1989; Donoghue, Doyle and Zimmer, unpublished results).

The rRNA data do not support Beck's (1981) contention that the seed plants arose through two different events: one of which gave rise to the cycads, seed ferns and angiosperms, while the other event gave rise to the other gymnosperms. Nor do the rRNA data support theories that the seed plants arose once but that the cycads and angiosperms are more closely related to one another than either is to any of the other gymnosperms. If the rRNA data supported either of these proposals, the cycads and angiosperms would form a monophyletic group to the exclusion of the other gymnosperms. This is not the case in either of the most parsimonious trees, nor is this topology found in any of the 3413 trees found within 4 steps of the shortest tree. All of the trees found within four steps of the most parsimonious tree unite the conifers, cycads and *Ginkgo*.

The angiosperm radiation. The rRNA sequence data strongly support the theories of a single origin for the flowering plants. In the most parsimonious trees and all 3413 trees found within four steps of the shortest tree, the angiosperms constitute a monophyletic group. In both bootstrapping trials, one with 40 of the 60 taxa, and one with 34 collapsed taxa, the flowering plants were placed in a single clade in 100 out of 100 replications. The branch which leads to the common ancestor of all the flowering plants is supported by more characters (42) than all but two other internal branches on the phylogenetic tree; one of these is the branch which separates the seedless plants from the seed plants. The characters which support this branch have a lower level of homoplasy than any other internal branch on the tree, again except for the branch which separates the ingroups from the outgroups. The rRNA data are in strong support for a monophyletic origin of flowering plants and consequently a single origin for each of the features, like double fertilization, which are unique to the angiosperms. Clearly the rRNA data refute theories of a multiple origin for the different groups of flowering plants (Meeuse, 1967).

Within the flowering plants, cladistic analysis of the rRNA sequences places members of the order Nymphaeales at the base of the angiosperm radiation, followed next by members of the order Piperales. In the shortest tree, the genera of Nymphaeales which represent the earliest divergence of the angiosperms include *Nymphaea*, *Nuphar*, *Cabomba* and *Barclaya*, but not *Ceratophyllum* or *Nelumbo* which the rRNA data place in a different position. The former four genera constitute a natural group without *Ceratophyllum* and *Nelumbo* in 3413 trees up to four steps longer than the most parsimonious tree. In the bootstrap with 40 taxa, *Barclaya* and *Nymphaea* were included and they were placed together in 100% of the replications; *Ceratophyllum* and *Nelumbo* were also included in this bootstrap and they were never grouped with *Barclaya* and *Nymphaea*, nor were they placed with one another a significant number of times. In another cladistic analysis

(Donoghue and Doyle, 1989b) the families containing *Nymphaea* and *Cabomba* formed a natural group that did not include *Nelumbo* (*Ceratophyllum* was not examined in Donoghue and Doyle's study). A cladistic treatment of morphological characters of genera within the Nymphaeales (Ito, 1987) found *Nelumbo* to be distinct from the other members of the order and found *Cabomba* to be more closely related to *Ceratophyllum* than to any of the other genera of Nymphaeales. There are morphological characteristics which support the separation of *Nelumbo* from the Nymphaeales. Notably, the pollen of *Nelumbo* is triaperturate while the pollen of all other Nymphaeales is monosulcate, and Takhtajan (1969) does place *Nelumbo* is a separate order. The rRNA data, then, are consistent with cladistic morphological treatments and some traditional classifications in so far as placing *Nelumbo* as separate from *Nymphaea*, *Cabomba*, *Nuphar* and *Barclaya*, but not with respect to the placement of *Ceratophyllum*. It is possible that the addition of *Brasenia* will help to unite *Ceratophyllum* with the other Nymphaeales, since *Brasenia*, *Ceratophyllum* and *Cabomba* constitute a natural group in Ito's (1987) analysis.

After Nymphaeales, the next branch to diverge from the rRNA tree leads to a natural grouping of the members of the order Piperales (*sensu* Takhtajan, 1969). In the rRNA tree, the genera *Piper*, *Peperomia* and *Saururus* are united and *Chloranthus*, which is considered by Cronquist (1968) to be a member of the Piperales, is placed elsewhere in the tree. The rRNA tree supports Takhtajan (1969) and Thorne (1976) who separate *Chloranthaceae* from the rest of the Piperales. The cladistic morphological treatment of Donoghue and Doyle (1989b) also separates *Chloranthus* from Piperaceae and Saururaceae.

Subsequent divergences among the flowering plants. The remaining 39 angiosperm taxa form two monophyletic sister groups. One of these groups contains all the monocot taxa plus *Nelumbo* and *Ceratophyllum*. Were it not for the presence of these two taxa, the monocots would constitute a natural group derived from within the dicots. With these water lilies present, the monocots cannot be considered a natural group. Within the monocots, the nine grasses are placed together in the same arrangement found in the analysis of the grasses alone. *Sabal* and *Hosta* form a natural group based on 18 shared characters, but according to traditional classifications, *Sabal* is more closely related to the two members of the family Araceae (*Colocasia* and *Pistia*) which form a natural group. The four aquatic monocots (*Sagittaria*, *Echinodorus*, *Najas* and *Potamogeton*) also form a monophyletic group. *Nelumbo* and *Ceratophyllum* form a grade with the aquatic monocots placed between them. The rRNA data support the resemblance of these groups with an aquatic habit and suggest that the first monocots were aquatic. Several key monocot lineages (e.g., basal Liliales and Bromeliales) have not been sampled yet, so this remains a preliminary conjecture.

The other group of derived angiosperms (relative to Nymphaeales and

Piperales) consists of the other members of the Magnoliidae subclass, as well as those of the other dicot subclasses. The two genera from Aristolochiales (*Aristolochia* and *Saruma*) are placed together in a natural group, as are three of the four members of the Magnoliales (*Magnolia*, *Liriodendron* and *Asimina*). *Drimys*, the fourth Magnoliales, has never been placed close to any other member of its order in the rRNA trees until the recent addition of another species of *Drimys*, *D. aromatica* (Suh, pers. comm.) In phylogenetic analyses which include both *D. aromatica* and *D. winterii*, the two are allied and have moved closer to the rest of its order. The two legumes (*Glycine* and *Pisum*) form a natural group, but *Duchesnea* and *Petroselinum,* the other genera of the subclass Rosidae, do not form a natural group with the legumes. The two genera of the subclass Caryophyllidae (*Stellaria* and *Spinacia*) form a monophyletic group. Much of the resolution within the remaining dicots is poor. Many of the branches are supported by few characters and many of these characters are quite homoplaseous. The various members of the subclass Hamamelidae (*Trochodendron*, *Platanus* and *Liquidambar*) are paraphyletic according to the rRNA data, as are the members of the subclass Magnoliidae (Magnoliales, *Hedycharya* and *Calycanthus*) and the subclass Ranunculidae (*Illicium*, *Ranunculus* and *Chloranthus*). Donoghue and Doyle (1989b) also found the Magnoliidae and Ranunculidae to be paraphyletic in their anaylsis, but they did find the Hamamelidae to group together. It is possible that the addition of other taxa closely related to those whose positions are inconsistent in the present analyses (that is, with better sampling of the tree of higher dicots) a more stable topology will result.

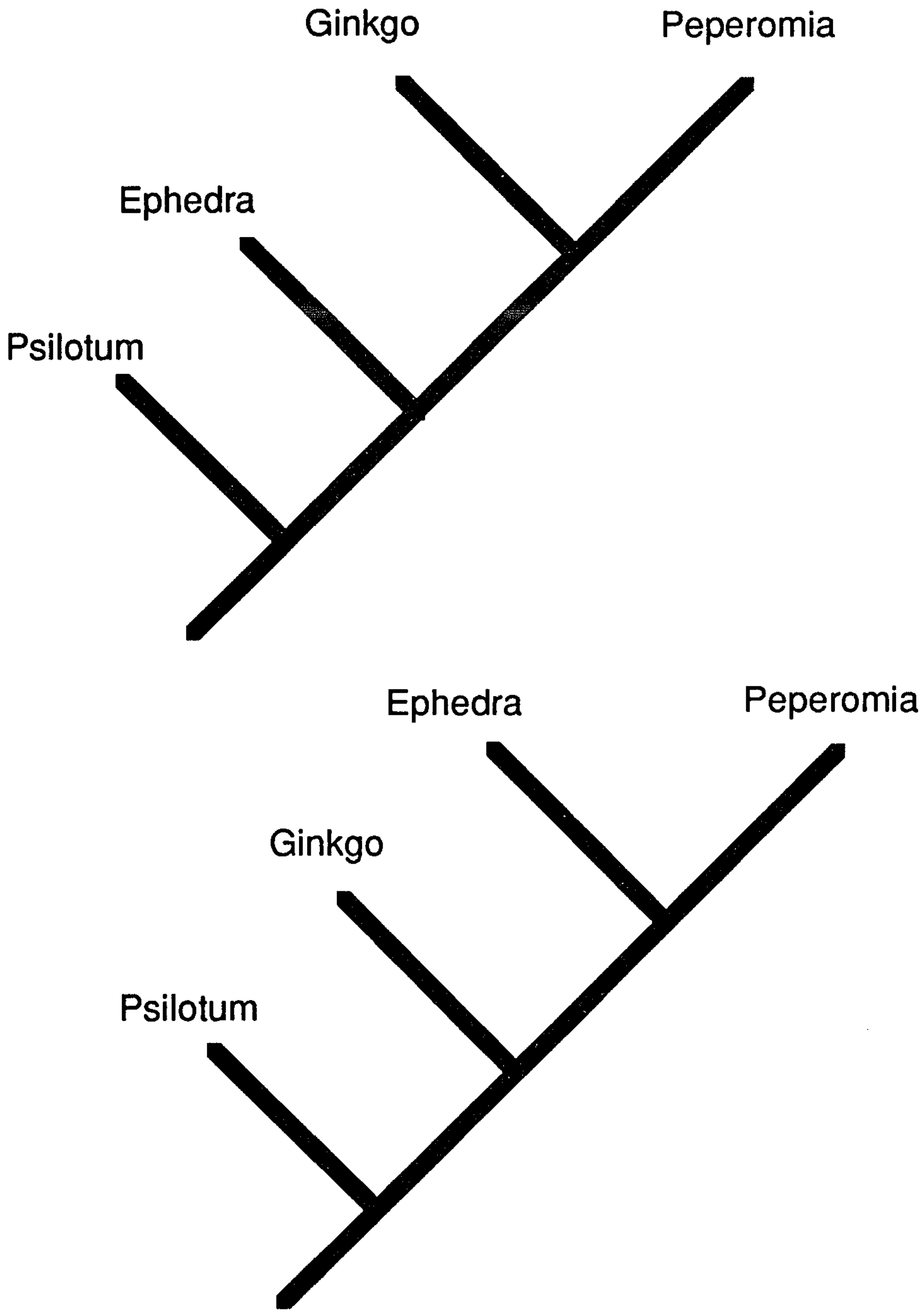
Testing alternative topologies. It is possible, with the computer program MacClade (Maddison and Maddison, 1990), to rearrange branches of phylogenetic trees and determine the "cost" as measured by additional steps to the shortest tree (creating less parsimonious arrangements of taxa). The alternative trees I chose to evaluate most closely are the ones that place the Gnetales as the sister group of the flowering plants and the ones that place the Magnoliales at the base of the angiosperm radiation. These alternatives are discussed in the text that follows.

Templeton's (1983) test can be used to compare two different phylogenies to determine if the data support one hypothesis over the other at a statistically significant level. This is a time consuming test for data sets with large numbers of characters because it is necessary to count the number of times each informative character changes in both topologies being compared. The variable but uninformative positions may be eliminated from the comparison, because they are constrained to change the same number of times and in the same location in both topologies. In this data set, it means mapping each of the 430 informative characters, one at a time, onto first one tree and then the other, and then comparing the number of changes required of that nucleotide to accomodate the particular topology. There is a simplified test for four taxa which can judge if the number of steps which separate two competing topologies is significant. However, to employ this test stringently, the sequences should be evolving in a clocklike manner (Felsenstein, 1988). As mentioned above, an assumption of clocklike evolution within seed plant rRNA does not seem stronlgy justified. The choice made here was to simplify key questions to four taxon tests and then use Templeton's test for significance (see below). If the four-taxon tests indicated strong support for one topology over another, then the test would be extended to the entire 60-taxon tree.

Alternative sister groups to the angiosperms. Placing the Gnetales as the gymnosperms most closely related to the angiosperms added 1 step to the most parsimonious rRNA tree for a total length of 1868 steps. In the shortest tree, there are 12 characters which unite the angiosperms to the conifer-cycad-*Ginkgo* group. There are, on the other hand, 17 characters which unite the Gnetales and flowering plants in the alternative tree of 1868 steps, and these characters are less homoplaseous as measured by their average consistency index (0.388 *v.* 0.410). In previous analyses with fewer taxa, the position of the Gnetales relative to the flowering plants has been variable. Although they have usually been placed as the sister group of the angiosperms, occasionally, as was the case with 60 taxa, the other gymnosperm group has been placed as most closely related to the flowering plants. In each analysis, however, it has never cost more than 3 steps, and usually only 1 step, to swap the arrangements of the two gymnosperm groups.

To test the placement of the Gnetales relative to the angiosperms, all the taxa were eliminated except a representative of the angiosperms (*Peperomia* of the order Piperales), an outgroup (*Psilotum*), a representative of Gnetales (*Ephedra*) and, one at a time, a representative of each of the three different groups of the other gymnosperms: cycads, *Ginkgo* and conifers. All the trees were rooted with *Psilotum*, so that with each combination of four taxa, there were only three different arrangements of the other taxa possible. Two alternatives to be tested with *Psilotum*, *Ephedra*, *Peperomia* and *Ginkgo* are shown in Figure 24. In one, the Gnetales are sister to the angiosperm representative and in the other, the *Ginkgo* represents the sister group to the flowering plants. The third possible topology, with the angiosperm group more closely related to *Psilotum* than either of the gymnosperm groups, was ignored. With these four taxa, there were only 32 characters that were informative (out of 1714), i.e., nucleotide sequence positions where two taxa share one nucleotide state, say G, and the other two share a nucleotide state other than G, say A. Each of the 32 variable nucleotide sequence positions favored one of the three possible trees over the other two. The tree with the Gnetalean genus as the gymnosperm most closely related to the angiosperm representative was favored by 15 sites, the tree with *Ginkgo* as the gymnosperm most closely

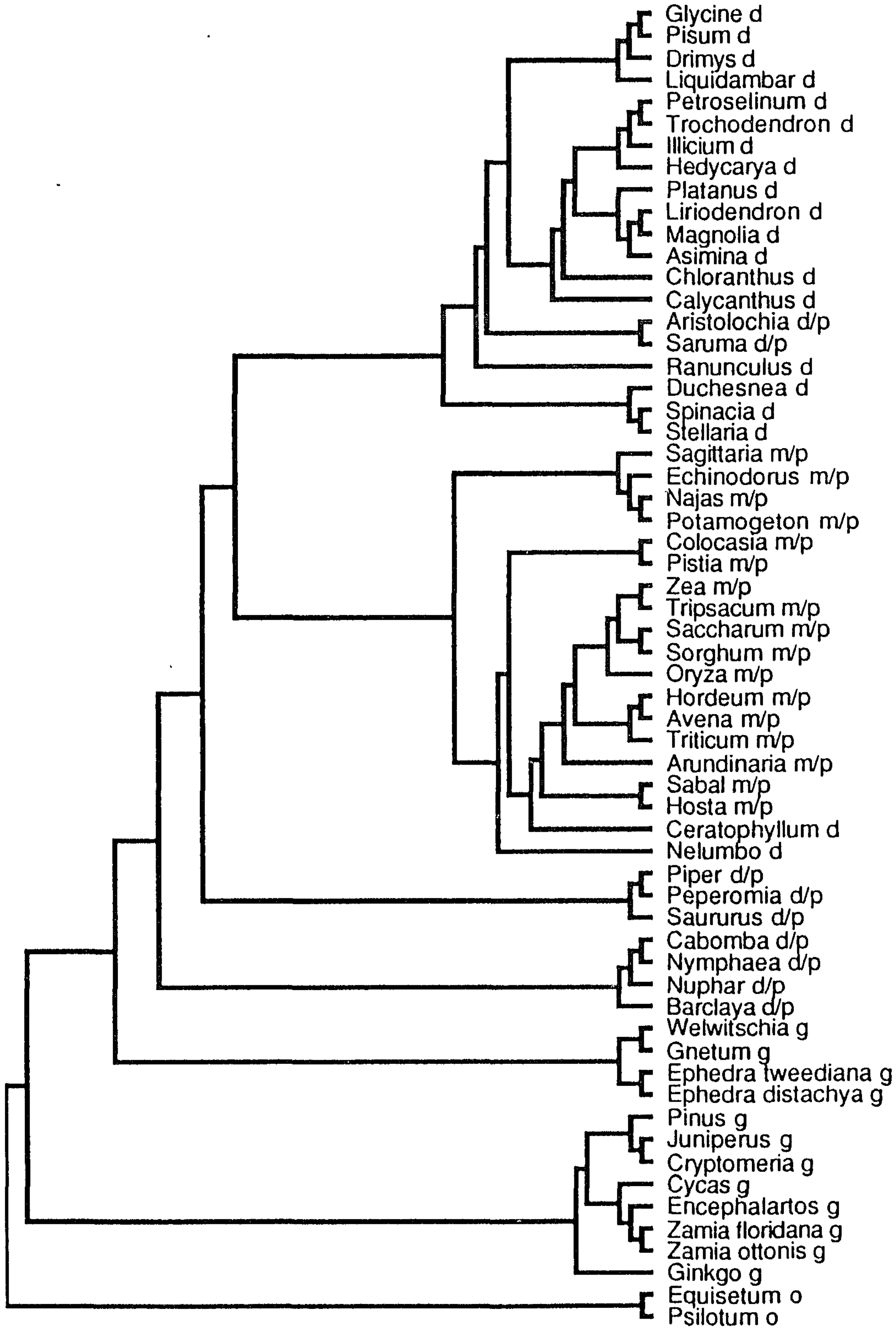
Figure 24. Two alternative topologies to test the relationships between gymnosperm lineages and flowering plants.



related to the flowering plants was favored by 7 sites. When *Ginkgo* was replaced by a cycad, *Zamia floridana*, there were again 32 informative characters and the arrangement that placed the Gnetalean representative as most closely related to the angiosperms was favored by 11 sites, and the topology placing the cycad as basal was favored by 6 sites. When the four taxa in the tree were *Psilotum*, *Peperomia*, *Ephedra* and *Juniperus*, there were 38 informative characters and the tree which placed the Gnetales as the gymnosperm most closely related to the angiosperms was favored by 14 sites and the conifer is favored as sister to the flowering plants by 7 sites. In all three tests, the Gnetales were always favored by a greater number of sites as the gymnosperm group most closely related to the angiosperms. However, only in the case when *Ginkgo* was used as the representative of the other gymnosperm groups did the confidence level of the comparisons approach 95%. The statistical significance of the results in favor of Gnetales as the sister group of the angiosperms was calculated by the winning sites test (Prager and Wilson, 1988) to be about 93%, 90% and 84% with *Ginkgo*, *Zamia* and *Juniperus* as the representative of the remaining gymnosperms, respectively.

Although the four-taxon tests indicated only weak statistical support for choosing one group of gymnosperms as sister group to the angiosperms over the other (i.e., the topology of Figure 22 or Figure 25), the results were sufficiently close to make the Templeton test over the entire 60 taxa seem

Figure 25. An alternative topology for 60 taxa in which the Gnetales are the sister group of the angiosperms. Length = 1868 steps.



worthwhile. The results of the test are listed in Table 8, and the details of the test are listed in the table legend. There were only seven characters whose number of changes varied with the two different topologies, and the test indicates that the two topologies are statistically inseparable.

Alternative basal angiosperms. In order to test whether the placement of paleoherb groups at the base of the flowering plant radiation was supported significantly over the placement of Magnoliales, PAUP was constrained to search for all shortest trees in which the Magnoliales (*Magnolia*, *Liriodendron* and *Asimina*), excluding the problematical *Drimys*, were placed as the first flowering plants. Although Hennig86 had found shorter trees than PAUP, Hennig86 cannot be constrained to a particular topology. To give PAUP a "head start" in its search, the shortest tree was first rearranged in MacClade to place the Magnoliales at the flowering plant base; this increased the length of the tree 14 steps from 1867 to 1881 steps. Then MacClade's branch swapping algorithm was invoked above the Magnoliales, that is, in the branch leading to the rest of the angiosperms, and the tree was shortened to 1880 steps. Finally, this topology was given to PAUP as a starting point, and PAUP performed its own branch swapping which is much more rigorous than MacClade's. PAUP was able to reduce the tree finally to 1877 steps (a 10-step difference) and still keep the Magnoliales at the base of the angiosperms. There were 12 equally parsimonious trees of 1877 steps with the Magnoliales as the earliest flowering plants, and a majority-rule

Table 8. Templeton's (1983) test comparing alternative topologies with either the Gnetales as sister group to the flowering plants, or the conifer-cycad-*Ginkgo* clade as sister group to the flowering plants. Characters are numbered out of the 617 variable (it was only necessary to consider the 430 informative positions, however, the data were more convenient to handle in this manner). A positive score indicates the number of changes by which the Gnetales-as-sister topology is favored. A negative score indicates the number of changes by which the conifer-cycad-*Ginkgo* placement is preferred. The rank is assigned beginning with a 1 for the smallest score, irrespective of sign, and increasing in increments of 1 afterwards. If there is more than one character with the same score, they are all assigned an average rank. For example, if there are 4 characters with an absolute value of 1 for a score, they are all assigned a rank of 2.5 (the average of 1, 2, 3 and 4). If the next lowest scores is held by 3 characters with an absolute value of 2, they would be assigned the rank of 6, i.e., the average of 5, 6 and 7. In the test shown here, all 7 characters had a score of +1 or -1, so they were assigned a rank of 4, the average of 1 through 7. The sign of the rank is the same as the sign of the score. All the positive ranks are then summed, and then the negative ranks are summed. The sum with the smaller absolute value is then used as T in the Wilcoxon signed-rank test table (Wilcoxon and Wilcox, 1964), printed in most statistics books. For the value of n, use the number of characters with differences and using the two-tailed or one-tailed chart, determine what value T must be to be significant to 0.05. If the value of T determined above is less than the value read from the chart, then the data are significant. In the test below, T is 12, and to be significant at the 95% level for n=7, T should be less than or equal to 2.

Character	Score	Rank
137	+1	+4
176	+1	+4
213	-1	-4
228	-1	-4
340	-1	-4
424	-1	-4
563	+1	+4

$n = 7$
$\Sigma T_{+} = +12$
$\Sigma T_{-} = -16$
Use $T = 12$

consensus of these 12 is shown in Figure 26. Much of the tree structure, aside from the placement of the Magnoliales, (e.g., the coherence of the Gnetales, grasses, aquatic monocots, etc.) is consistent with the shortest tree of 1867 steps. Most interesting is the fact that the paleoherb groups are still placed near the bottom of the angiosperm clade.

A four-taxon test was tried with a representative of Gnetales (*Ephedra*), Piperales (*Peperomia*), Magnoliales (*Magnolia*) and Nymphaeales (*Barclaya*) to see if there is statistical support for the placement of the Piperales or Nymphaeales as more basal in the angiosperm tree relative to Magnoliales. All three possible arrangements with the tree rooted in the Gnetalean genus are shown in Figure 27. Again Templeton's test was used to measure the significance of the results. Topology I in Figure 27 is favored by nine of the sites, topology II is also favored by nine of the sites, and topology III, in which the Magnoliales are basal, is not favored by any of the 18 sites. Templeton's test says that topology I is favored over topology III with greater than 99% confidence, and that topology II is favored over topology III by the same figure, and that topology I and topology II are indistinguishable. In other words, the four-taxon test of rRNA sequence data unequivocally supports the basal placement of Nymphaeales or Piperales over the basal placement of Magnoliales within angiosperms.

A Templeton test over the entire tree of all 60 taxa (Table 9) indicated that the difference between the two trees, the tree with Nymphaeales and

Figure 26. The majority-rule consensus of 12 shortest trees with the Magnoliales as the basal angiosperms. All nodes except those labelled were 100% conserved within all 12 trees.

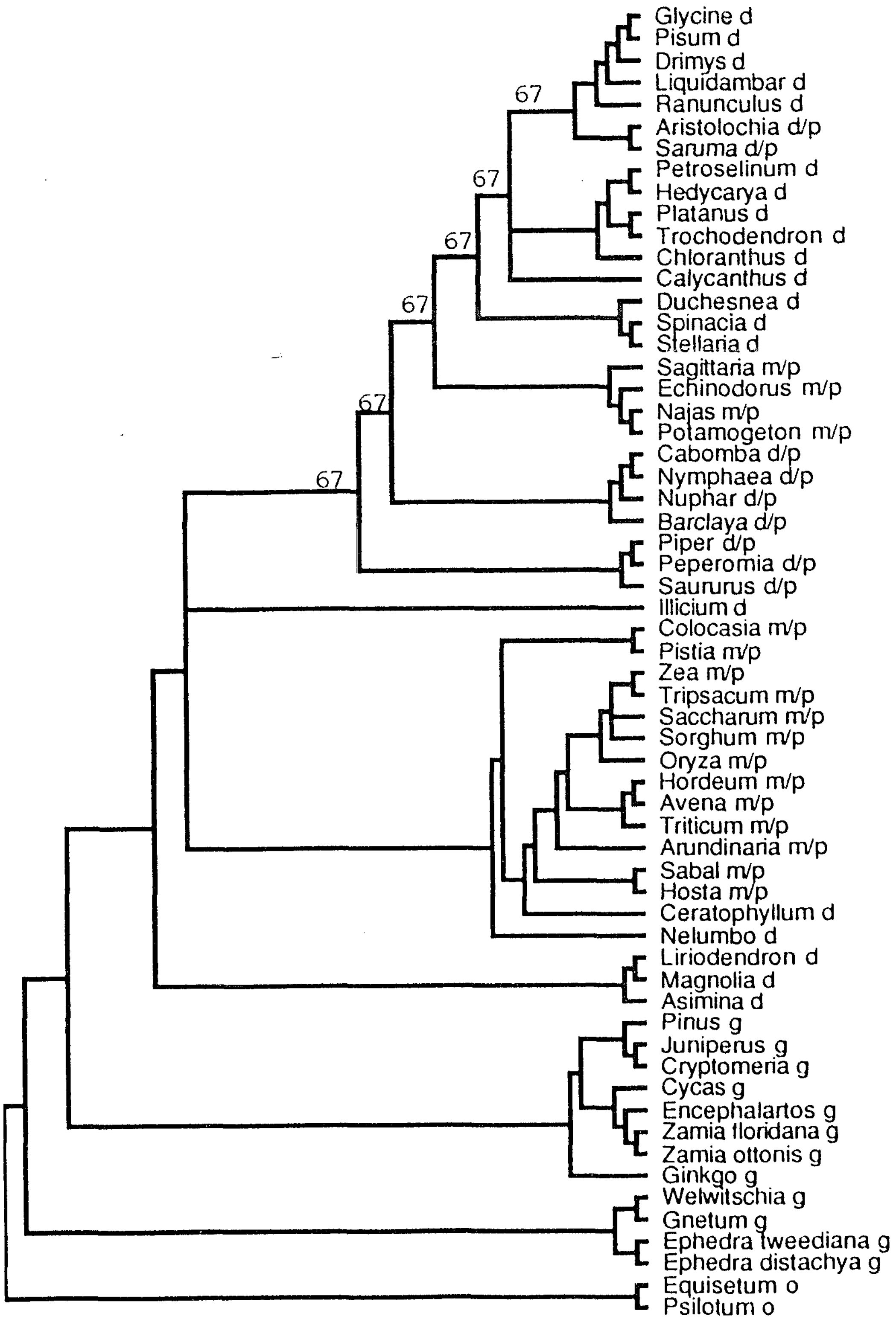


Figure 27. Three alternative arrangements at the base of the angiosperm radiation.

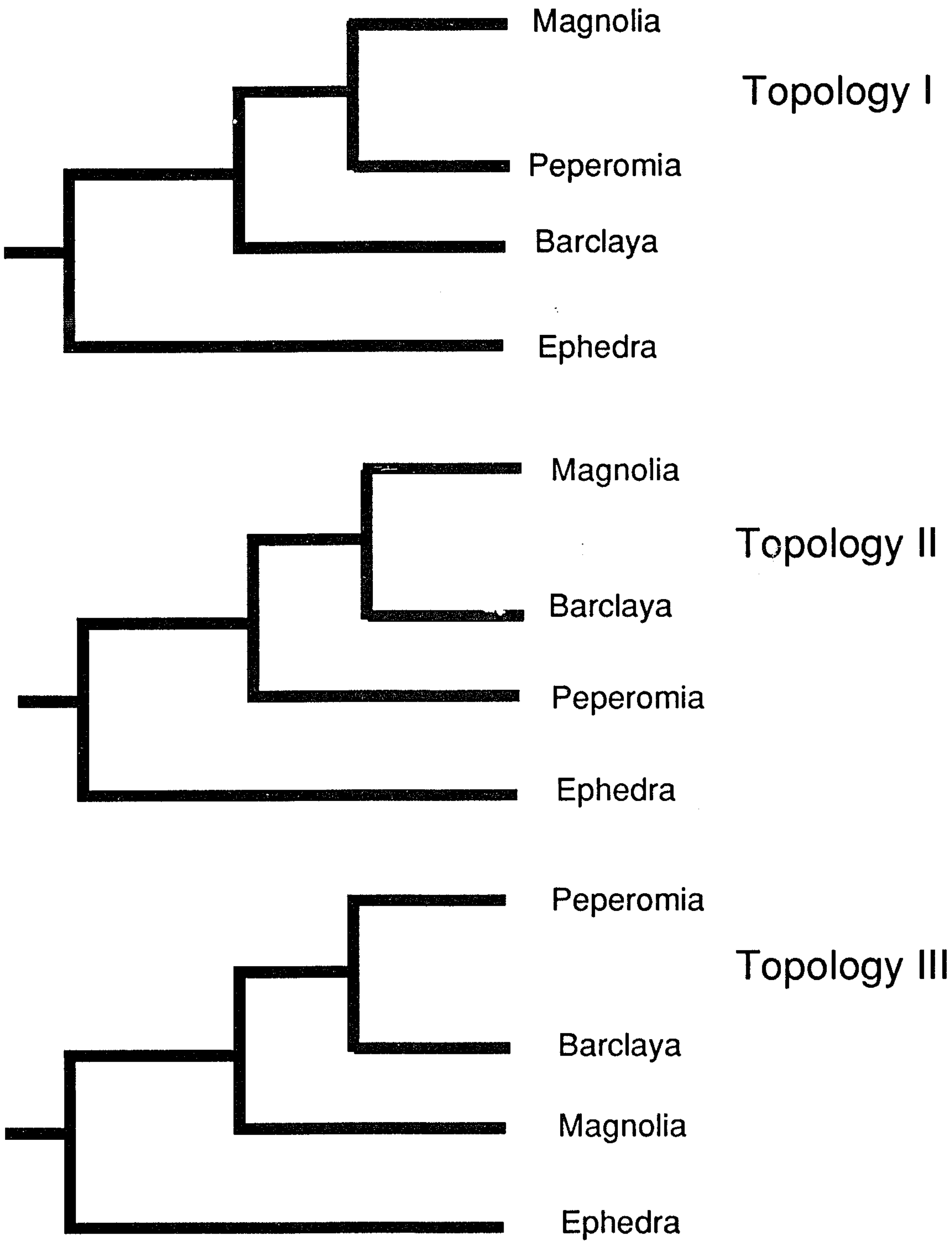


Table 9. Results of a Templeton test comparing the alternative arrangements at the base of the phylogenetic tree of flowering plants. The character number is relative to the 617 variable rRNA sequence positions. The negative scores indicate that the character underwent fewer changes in the topology with Magnoliales as the most primitive angiosperm group. The positive scores indicate that the character underwent fewer changes in the most parsimonious tree, i.e., the tree with Nymphaeales as the basal angiosperm. Other details are in the legend of Table 8.

Character	Score	Rank
10	-1	-15.5
11	+1	+15.5
12	+2	+33
30	+2	+33
50	+1	+15.5
58	-1	-15.5
66	+1	+15.5
70	+1	+15.5
71	+1	+15.5
72	+1	+15.5
93	+1	+15.5
106	+2	+33
125	-1	-15.5
155	-1	-15.5
168	+1	+15.5
183	+1	+15.5
223	-1	-15.5
229	+3	+36.5
230	+1	+15.5
236	+1	+15.5
267	-1	-15.5
268	-1	-15.5
269	-1	-15.5
325	+1	+15.5
326	+1	+15.5
333	+1	+15.5
345	-1	-15.5
330	-1	-15.5
336	-2	-33
424	+1	+15.5
425	+2	+33
430	+1	+15.5
450	-1	-15.5
487	-1	-15.5
517	-1	-15.5
553	+1	+15.5
562	-3	-36.5

$n = 37$

$\Sigma T_+ = +440.0$

$\Sigma T_- = -271.0$

Piperales basal at 1867 steps and the tree with Magnoliales basal at 1877 steps, was not significant with 95% confidence. If, as suggested by Templeton (1983), the Wilcoxon signed rank test is applied as a one-tailed test, the data are found to be significant at the 90-95% level. If instead, the Wilcoxon test is applied as a two-tailed test as argued by Felsenstein (1988), the data are significant at a level of about 85%. A two-tailed test is required if there is no reason to otherwise differentiate between the two hypotheses being tested. The fact that parsimony supports the Nymphaeales over Magnoliales as basal by 10 steps may give sufficient support for choosing the shortest tree as "correct" and applying the one-tailed test to determine if it is significantly different from the other. In this case, the support across the entire tree for rejecting the Magnoliales as basal comes very close to significance at a high level, with probabilities of between 90 and 95%.

Testing the data. The sequence data for all 60 taxa were tested to determine if they contained any information. As before, the data were tested by randomizing them, and then inferring the shortest phylogenetic tree with the randomized sequences (Archie, 1989a). The randomization program can handle only 30 taxa at a time, so half the taxa were deleted, while keeping the overall range of taxa the same. This was accomplished by eliminating duplicate members of some families (e.g., deletion of eight of the nine Poaceae), or by eliminating some members of particularly strong clades, e.g., the one which unites all three conifers. When the 30 remaining taxa were analyzed by PAUP, the shortest tree inferred was 1058 steps long. Because each replication required significant mainframe computer time, the number of randomizations was reduced to 25 from the recommended 100, a tactic that Archie suggests is valid since the variance in the lengths of random trees is usually low (1989b). The trees calculated based on the randomly generated data ranged in length from 1249 to 1273 steps with a mean of 1261, about 20% greater than the tree based on the actual data, and a standard deviation of 6.6 steps. The shortest tree with the actual sequence data is more than 30 standard deviations removed from the mean of the randomized trees, indicating strongly that the data are more significant than random data.

The HER was calculated to be 0.274, indicating that there are many homoplaseous characters (roughly 70%) in the data set. These results are not surprising considering that within the narrow divergence of the grasses, the HER indicated about 40% randomness for the rRNA data. It should not be considered contradictory to say that the data are informative, but possess many homoplasies. Homoplasy is to be expected in DNA data sets, especially over long evolutionary times since DNA is subject to back mutations and these DNA characters are confined to only five possible character states (G, A, T, C or absent).

The randomization test should allow for identification of those characters which are particularly homoplaseous. As these characters are eliminated, the HER will increase. All the characters which exhibited all four character states were eliminated from the data set to see if these characters were contributing significantly to the homoplasy. When the 58 positions with at least three changes were eliminated, PAUP found eight versions of the most parsimonious tree of 819 steps. The majority-rule consensus tree is shown in Figure 28. In this tree, the coniferopsids and cycads are the sister group to the flowering plants and the paleoherb groups are still placed at the base of the angiosperm radiation, while *Ceratophyllum* and *Nelumbo* are no longer placed among the monocots. Instead, *Aristolochia* and *Saruma*, other members of the paleoherbs, move down to within the monocots. In this tree, then, all the paleoherb groups are together at the base of the angiosperm radiation separate from all the other angiosperms. Surprisingly, when the HER was recalculated, after 25 randomizations on the modified data set, it was actually reduced to 0.233. This indicates that the homoplaseous data may be distributed fairly evenly among the data, and that it will be difficult to remove them simply by eliminating those characters that change the most. The distribution of the randomized trees over 30 taxa when the four-state characters were removed is shown in Figure 29. When all characters with more than two states were eliminated, it reduced the data set such that PAUP could no longer converge on a most parsimonious solution; there were hundreds of equally parsimonious trees.

In another attempt at identifying the most homoplaseous data, the rRNA sequences from the most variable primers, 18E, 26D and 26F were

Figure 28. The majority-rule consensus of 8 shortest trees with the four-state characters eliminated from the data set.

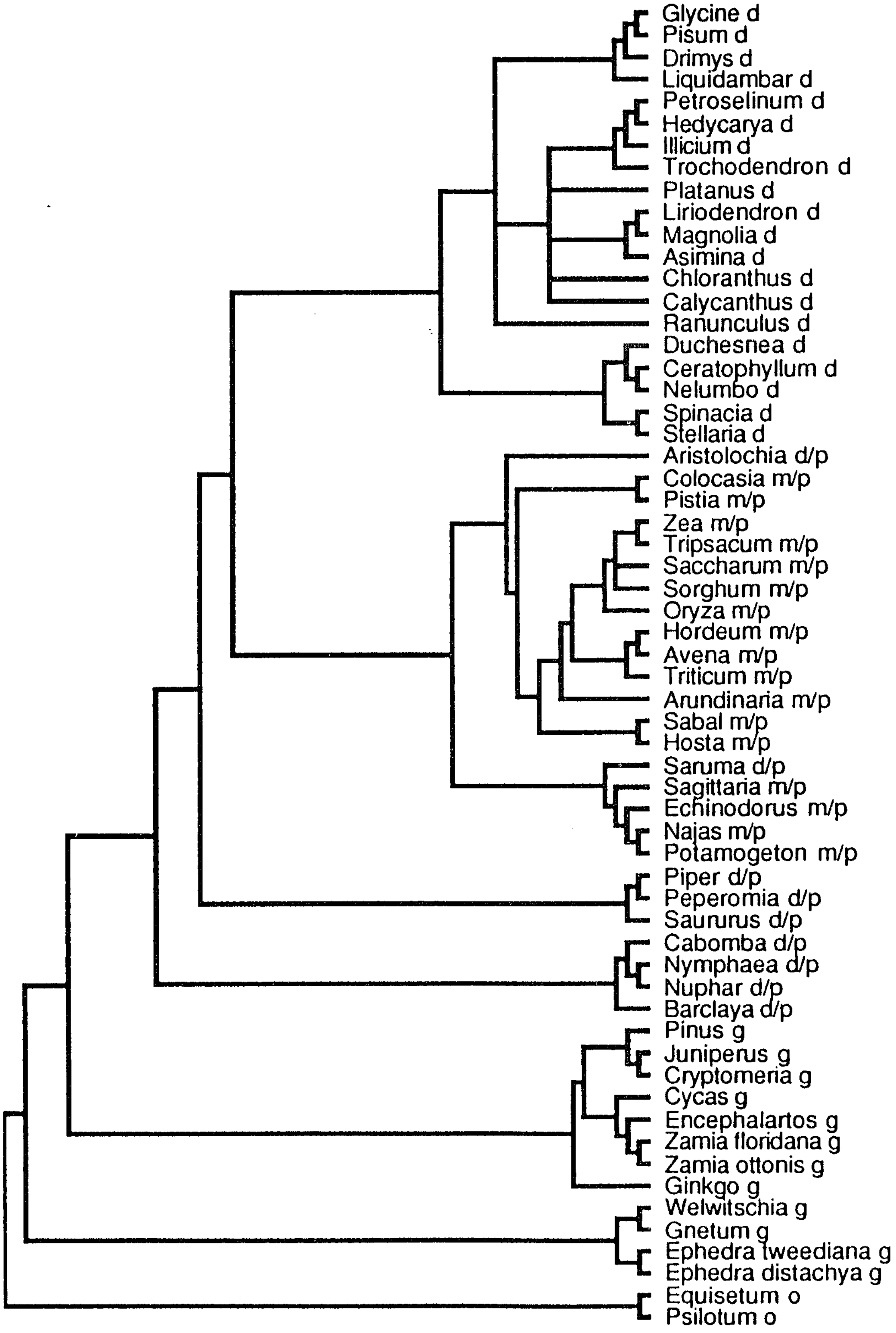
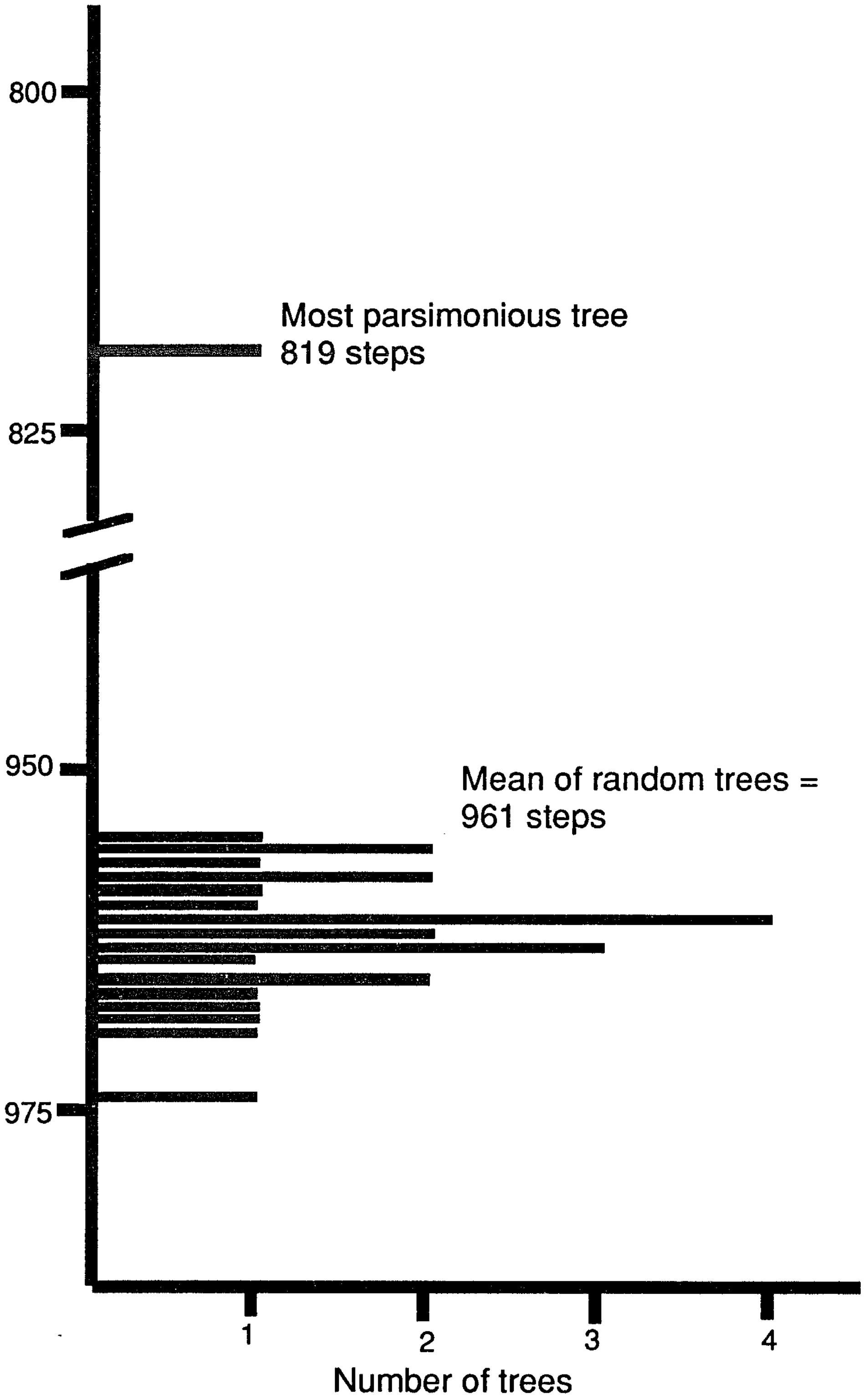


Figure 29. The distribution of trees after 25 randomizations of the data set for 30 taxa when the four-state characters were removed from the data set.



excluded. Again there was insufficient information for PAUP to resolve the relationships among the 60 taxa. When either the 18S data or the 26S data alone were analyzed, there was not enough sequence information to converge to a most parsimonious solution.

Majority-rule consensus trees. A series of majority-rule consensus trees was constructed by PAUP by the addition of less parsimonious trees to the most parsimonious ones. The number of trees within four steps of the most parsimonious tree was so large that it was impractical to try to collect trees longer than 1871 steps. This series of trees is shown in Figure 30, and key nodes are labelled with the percentage of trees in which that node was found. Figure 30a is the consensus of the two versions of the shortest tree, Figure 30b is the majority-rule consensus of the two trees that are 1867 steps and the 30 that are 1868 steps. Figure 30c is a majority-rule consensus of the 32 trees less than or equal to 1868 steps and the 357 trees found at 1869 steps. The majority-rule consensus trees shown in Figures 30d and 30e were calculated from 1055 and 3413 trees, respectively. The first nodes to disintegrate with the addition of less parsimonious trees, as reflected by the dissolution of dichotomous branching into polychotomous branching, were among the higher dicots. This is consistent with the apparently poor resolution in this part of the tree. There does not appear to be sufficient information contained within the current rRNA data set to competely resolve the relationships within the higher dicots. On the other hand, within the

Figure 30. A series of majority-rule consensus trees for 60 taxa.
Figure 30a. The majority-rule consensus of 2 trees at 1867 steps.

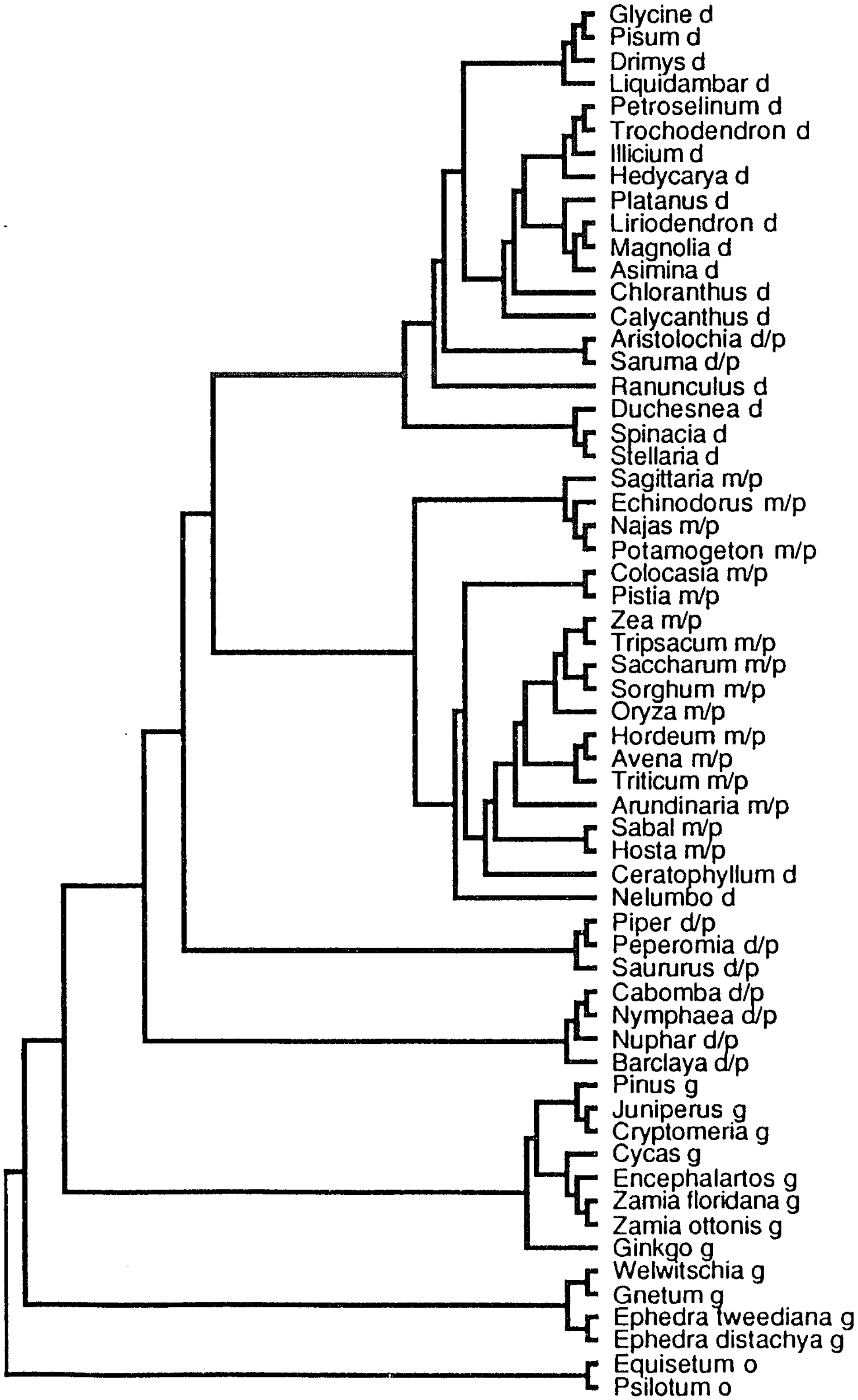


Figure 30b. The majority-rule consensus of 32 trees of 1867-1868 steps.

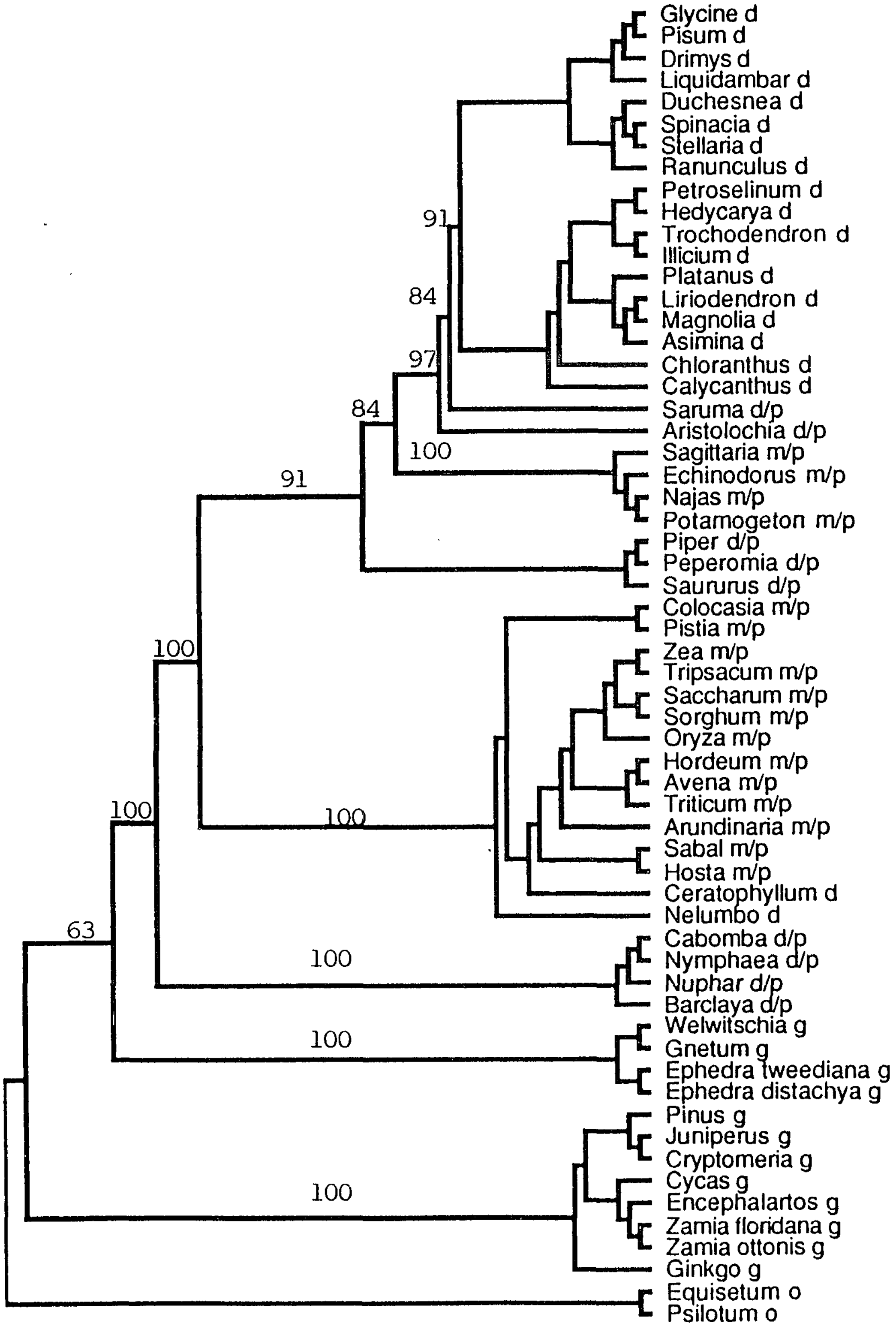


Figure 30c. The majority-rule consensus of 389 trees of 1867-1869 steps.

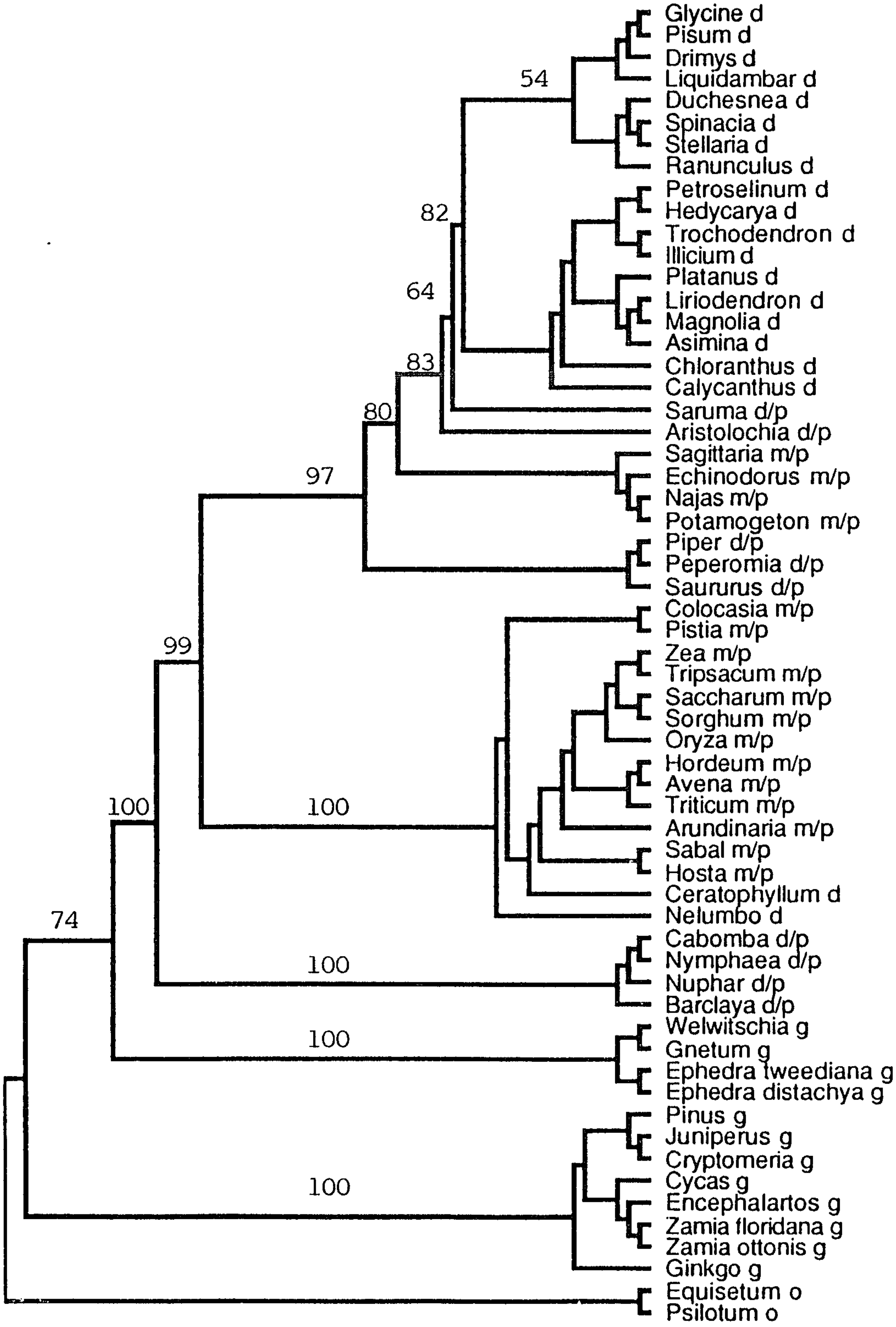


Figure 30d. The majority-rule consensus of 1055 trees of 1867-1870 steps.

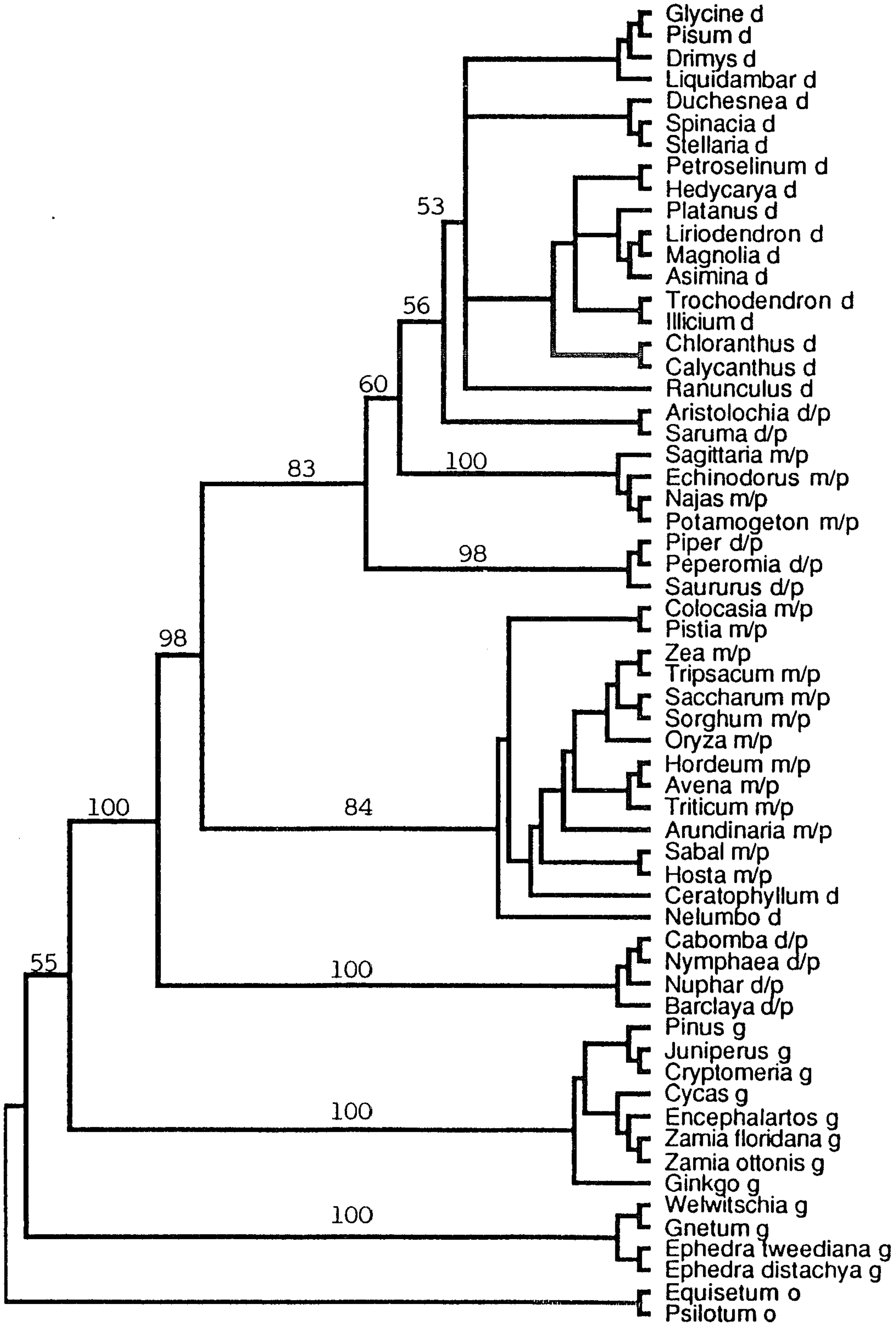
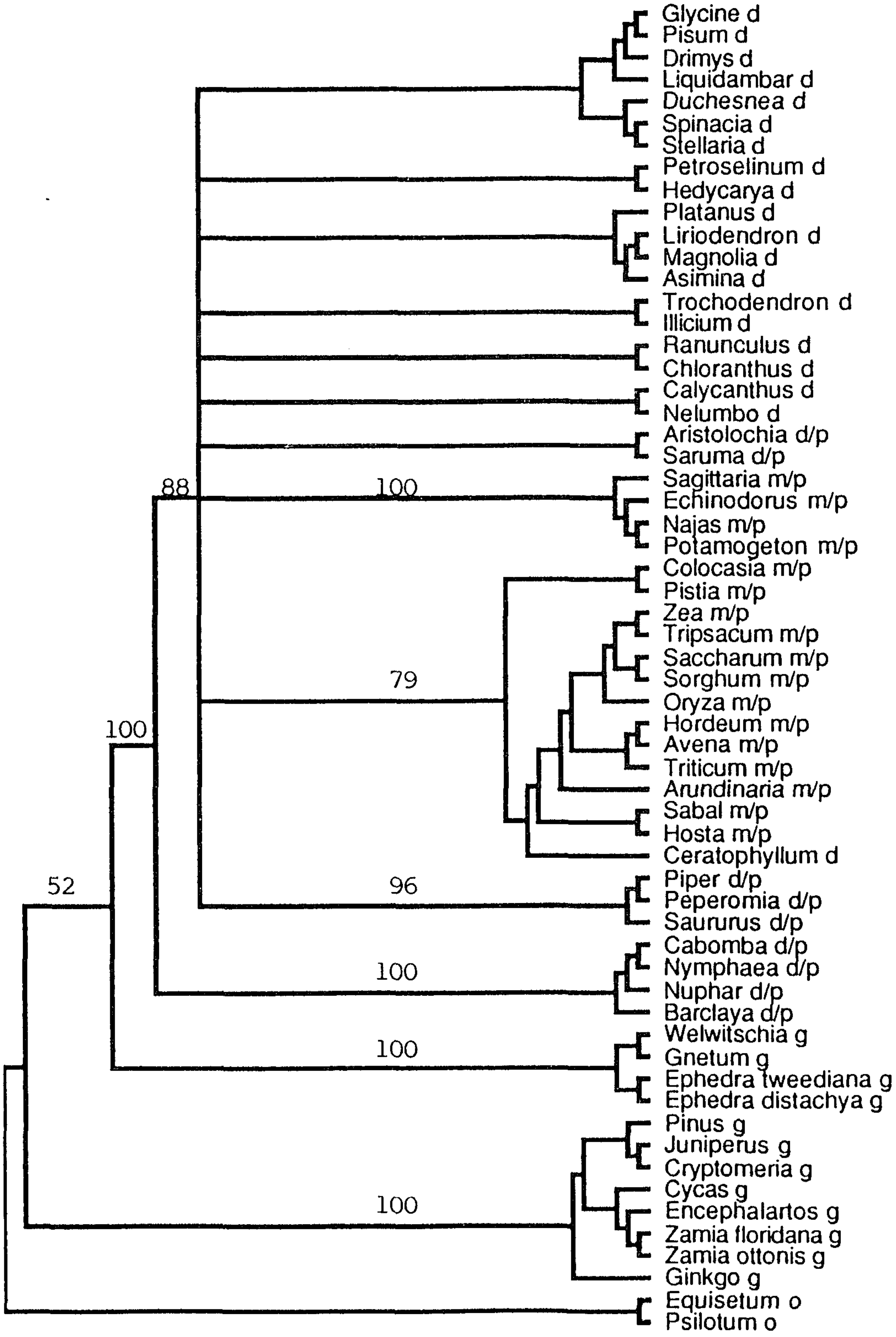


Figure 30e. The majority-rule consensus of 3413 trees of 1867-1871 steps.

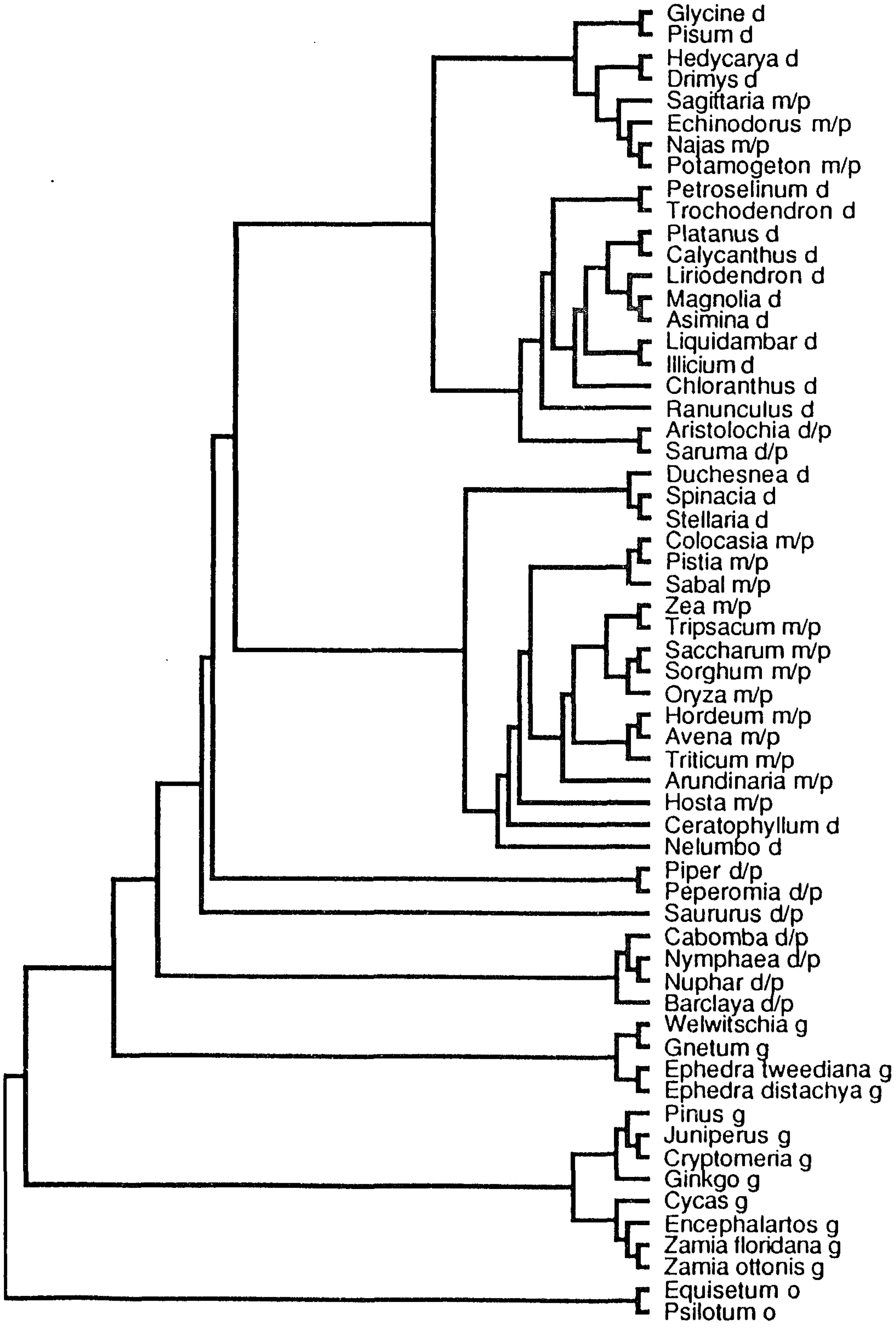


angiosperms, the nodes which are best supported are the ones that place the Nymphaeales, Piperales and the monocot group (all members of the paleoherbs) near the base of flowering plant evolution. The position of the Gnetales relative to the angiosperms shifts about depending on which set of trees are used for the consensus, again indicating the weakness of the placement of either gymnosperm group as sister to the flowering plants.

Distance analysis. The neighbor-joining analysis (Saitou and Nei, 1987) yielded different results based on the manner in which the sequences were converted to distances. When the distance was simply equivalent to the dissimilarity (number different/number compared), the phenogram in Figure 31 was inferred. When the data were corrected for possible multiple changes at individual loci (Jukes and Cantor, 1969), and adjusted to give more weight to transversions (Kimura, 1980) the resultant phenograms shared the topology shown in Figure 32.

In comparing the two topologies, there are slight differences among the more derived taxa, depending on how the distances were determined, but several features of the two phenograms are consistent. In both topologies, the Gnetales are placed as the sister group of the flowering plants, and the remaining gymnosperms form an older, monophyletic group. At the base of the flowering plants lie the Nymphaeales and Piperales, with the Piperales split into two separate lineages, one consisting of *Saururus* alone, and the other comprised of *Piper* and *Peperomia*. The phenetic analysis, then,

Figure 31. Results of a phenetic analysis of rRNA sequences from 60 taxa. Distance based on overall dissimilarity.



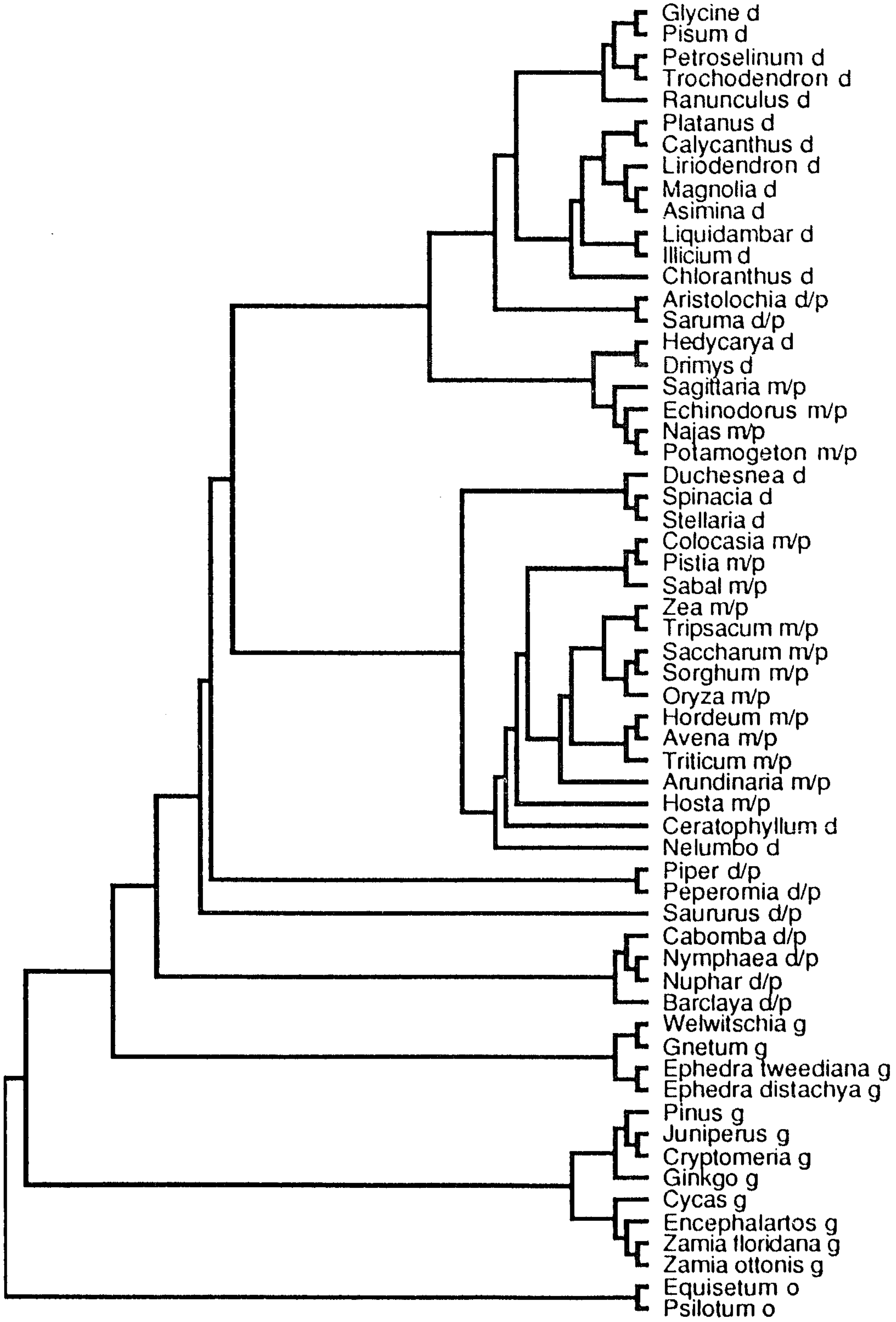


Figure 32. Results of a phenetic analysis when the distances are calculated with Jukes-Cantor (1969) or Kimura (1980) formulae.

supports the placement of some members of the paleoherbs as the first flowering plants, specifically the Nymphaeales and Piperales. It also supports the placement of the Gnetales as the sister group of the angiosperms. There is no provision for calculating phenograms other than deriving the shortest one, so it is not possible to investigate alternative arrangements of taxa with distance data.

SUMMARY

The rRNA data support a monophyletic origin of the seed plants and do not support Beck's hypothesis that the seed plants arose separately from two different progymnosperm lineages. The rRNA sequences suggest that the flowering plants are not more closely related to any one group of extant gymnosperms over the others, but that the flowering plants did arise from within the gymnosperms. The rRNA data are consistent with a derivation of the flowering plants from one of the extinct seed fern lineages. However, if the flowering plants were derived from a seed fern group, it was not the same seed fern group which gave rise to cycads (unless all seed plants, or all seed plants except Gnetales, are descended from the same seed fern ancestor).

The rRNA data give strong support for the coherence of the Gnetales as a natural group. The most parsimonious rRNA trees do not place the Gnetales as the sister group of the angiosperms, although an insignificant penalty of 1 step is all that is required to reverse the position of the Gnetales and the remaining gymnosperms. In this alternative tree, the branch which unites the angiosperms with their most closely related gymnosperm lineage (the Gnetales in this case) is supported by more characters and with less homoplasy than is the analogous branch in the most parsimonious tree. When the data set was reduced to four taxon tests, statistical tests favored the placement of the Gnetales as the sister group of the flowering plants, though the differences by which the Gnetales were favored only approached statistical significance at the 95% level in one case. Templeton's test over the complete tree indicated that the two topologies were indistinguishable. The Gnetales were often placed as the gymnosperms most closely related to angiosperms in preliminary analyses of rRNA sequences with fewer taxa, and they are also placed as sister to the flowering plants when there are 64 taxa (Suh, pers. comm.) and 72 taxa (Bult, pers. comm.). In addition, the distance analyses indicated that the Gnetales were the sister group of the angiosperms. Clearly, the placement of the Gnetales relative to the other gymnosperms and the angiosperms cannot be resolved unequivocally by the rRNA sequence data from eight primers alone.

The Nymphaeales and Piperales lie at the base of the angiosperm diversification according to rRNA sequence analysis. Along with the monocots and Aristolochiales, the Nymphaeales and Piperales make up the "paleoherbs" clade of Donoghue and Doyle (1989a). The rRNA sequences suggest that some members of the paleoherbs are the earliest diverging flowering plants, and that the rest of the angiosperms arose from within these paleoherbs. The basal arrangement of these paleoherb groups was supported by the most parsimonious tree, and the majority of all trees up to four steps longer than the shortest tree. It was also supported by the distance analyses and the analysis in which the four-state characters were eliminated. Even in the tree in which the Magnoliales were forced to the base of the tree, the paleoherb groups were the next groups to evolve according to the rRNA data.

The paleoherbs, according to the rRNA data, should no longer be referred to as a group, because they are not a natural assemblage. If they were a natural assemblage, and therefore, an appropriate group for plant classification systems, there would be one ancestor common to all paleoherbs which had no other descendants other than the paleoherbs. This is not the case in the rRNA tree, because the common ancestor of the paleoherbs is also the common ancestor of the remaining flowering plants. By the same reasoning, the rRNA data also suggest that the monocotyledonous and dicotyledonous condition are not appropriate for classification systems in the Hennigian sense. The rRNA data show that the dicots and monocots are both paraphyletic groups.

The alternative topology with the Magnoliales emerging first during differentiation of the flowering plants was rejected in a four-taxon test with greater than 99% confidence. When the statistical test was applied to trees containing all 60 taxa, the tree with the Magnoliales basal could be rejected with a confidence level between 90 and 95%.

Thus the rRNA data suggest that the first flowering plant lineages were herbaceous and perhaps aquatic, in contrast to the traditional views which hold that the first angiosperms were woody plants. The rRNA sequence data also support hypotheses that the first angiosperms had monosulcate pollen similar to that of gymnosperms. Finally, analysis of the trees based on rRNA sequences suggests that the first monocots were also aquatic plants, and that many other groups recognized today in traditional classification systems may not be natural groups.

CONCLUSIONS

Phylogenetic analyses were performed on more than 1700 sites of nuclear ribosomal RNA sequence from the 18S and 26S molecules of 58 seed plants and two outgroup taxa. Based on these analyses the following conclusions can be made:

1. The rRNA sequence data are informative as compared to randomly-generated data, although there is a high level of homoplasy in the rRNA sequence data.
2. The seed plants arose only once during evolution. Theories proposing multiple origins of the seed plants are not supported by the rRNA sequence data.
3. The gymnosperms are not a natural group. The extant gymnosperms can be divided into two separate natural groups, the Gnetales and a clade consisting of cycads, conifers and *Ginkgo*.
4. The angiosperms are a natural group that arose once from within the gymnosperms. Although the most parsimonious rRNA tree indicates that the conifer-cycad-*Ginkgo* gymnosperm clade is the sister group of the angiosperms, there is no statistical significance for this placement over the slightly less parsimonious arrangement of the Gnetales as the sister group of the flowering plants.

5. The most basal of angiosperms are the lineages leading to Nymphaeales (*sensu* Takhtajan, 1969) with the exception of the family Ceratophyllaceae. The next most basal lineage is represented by the Piperales (*sensu* Takhtajan, 1969). The alternative placement of the Magnoliales as the first divergence within the angiosperms can be rejected with a level of confidence approaching 95%.
6. Neither dicots nor monocots constitute natural groups according to the rRNA trees.

RECOMMENDATIONS TOWARD FUTURE WORK

Future directions of this project to elucidate flowering plant genealogies should proceed along several parallel tracks. The first is to add sequences from other molecules. This will best be accomplished by utilizing the polymerase chain reaction (PCR) (Saiki *et al.*, 1985; Mullis and Faloona, 1987) for either cloning and sequencing, or, for sequencing directly from single-standed (asymmetric) amplifications. Sequencing from a cloned PCR product has the advantage of allowing one to sequence both strands of the gene(s) of interest, though asymmetric amplification and sequencing may be a more rapid means to acquire sequence data. We already have the necessary primers to amplify and sequence almost the complete chloroplast 16S rRNA gene. There is a also a set of primers which contains restriction sites within the sequence of the primer to aid in subsequent cloning of the PCR product. The current protocols for PCR amplification and cloning are in the appendix.

There are good reasons to continue sequencing other regions of the 18S and 26S rRNA molecules. It will be necessary to have complete sequences to propose and test secondary structure models which may help to identify those regions which are more conserved relative to other regions. New sequences also may add more information for the resolution of the seed plant evolution questions, although when two more regions of the 26S molecule were sequenced for the grasses, little resolution was added to the problematical placement of *Oryza*. These two regions were highly conserved among the grasses even though one of them (26J) is within one of the purported expansion segments of the 26S molecule.

Instead of adding new sequences, it has in the past often been more beneficial to add more taxa to increase resolution in the phylogenetic trees. The shifting of the Gnetales relative to the flowering plants was discussed previously. Another good example of a volatile taxon is the monocot *Sagittaria*. When there were only 37 taxa in the analysis, *Sagittaria* was placed as more closely related to the relatively advanced legumes, *Glycine* and *Pisum*, than it was to any other monocot. The terminal branch connected to *Sagittaria* in that first tree was very long, and it is well known that parsimony can fail when there are very long branch lengths. The best way to handle very long branches is to add related taxa (Swofford and Olsen, 1990), and when more aquatic monocots were added to the analysis, *Sagittaria* eventually settled into place within the aquatic monocots.

More representatives of the higher dicots are presently being added to the data set, as well as more members of the Magnoliidae and Hamamelidae to maximize the overlap between the rRNA data set and the morphological data set of Donoghue. Eventually, the goal is to combine the molecular and morphological data sets to see if they are complementary. There may be certain features of seed plant evolution that can only be resolved by one data set or the other, and the combination of the two could be very powerful. Combining the two data sets into one is not going to be trivial, though, because consideration must be given as to how to weight the morphological data in comparison to the molecular data. Donoghue and Doyle's analysis has on the order of 60 characters, while ours presently has more than 400 informative characters; the molecular data could overwhelm the morphological data. Possibly the best solution will be some *a priori* weighting of the data to give more importance to the morphological characters.

The other track should be a more thorough analysis of the character of the data. The randomization test of Archie (1989a) can be a powerful tool in the identification and elimination of the more homoplaseous characters from the data set. A simple test of deleting the four-tate characters revealed that more than a cursory examination of the patterns of change of each character will be necessary to effectively eliminate especially noisy characters. Deletion of these noisy characters is necessary for two reasons; the obvious one is that they intefere with the inference of the best trees, the second one is that the rRNA data set is growing so rapidly that it will soon overwhelm most if not all phylogenetic programs currently available. Many programs already cannot handle data sets as large as this one. While PAUP can theoretically handle data sets with many more taxa or characters than we presently have, it has had difficulty converging on the shortest tree since the number of taxa has exceeded 57. Hennig86 is presently the best option for finding the shortest tree, but it is limited to 990 characters, and with 72 taxa, the number of informative sites has risen to almost 700. It will not be too much longer before Hennig86 will also be overwhelmed.

REFERENCES

Appels, R., and Dvorak, J. (1982a) The wheat ribosomal DNA spacer region: its structure and variation in populations and among species. *Theor. Appl. Genet.*, **63**, 337-48.

Appels, R., and Dvorak, J. (1982b) Relative rates of divergence of spacer and gene sequences within the rDNA regions of the species in the Triticeae: implications for the maintenance of homogeneity of a repeated gene family. *Theor. Appl. Genet.*, **63**, 361-5.

Appels, R., and Honeycutt, R.L. (1986). rDNA: evolution over a billion years. *in* DNA Systematics Volume II: Plants (ed S.K. Dutta), CRC Press, Boca Raton, Florida, pp. 81-135.

Arber, E.A.N., and Parkin, J. (1907) On the origin of angiosperms. *J. Linn. Soc. Bot.*, **38**, 29-80.

Archie, J.W. (1989a) A randomization test for phylogenetic information in systematic data. *Syst. Zool.*, **38**, 239-52.

Archie, J.W., (1989b) Homoplasy excess ratios: new indices for measuring levels of homoplasy in phylogenetic systematics and a critique of the consistency index. *Syst. Zool.*, **38**, 253-69.

Archie, J.W., (1989c) Phylogenies of plant families: a demonstration of phylogenetic randomness in DNA sequence data derived from proteins. *Evol.*, **43**, 1796-1800.

Arnheim, N. (1983) Concerted evolution of multigene families. *in* Evolution of Genes and Proteins (eds M. Nei and R.K. Koehn), Sinauer Assoc., Inc., Sunderland, MA., pp. 38-61.

Beaucage, S.L., and Caruther, M.H. (1981) Deoxynucleoside phosporamidites - a new class of key intermediates for deoxypolynucleotide synthesis. *Tet. Lett.*, **22**, 1859-62.

Beck, C.B. (1974) Origin and early evolution of angiosperms: a perspective. *in* Origin and early evolution of angiosperms (ed C.B. Beck), Columbia University Press, New York, pp 1-10.

Beck, C.B. (1981) *Archeopteris* and its role in vascular plant evolution. *in* Paleobotany, paleoecology, and evolution, Volume I (ed K.J. Niklas), Praeger, New York, pp. 193-230.

Bessey, C.E. (1897) Phylogeny and taxonomy of the angiosperms. *Bot. Gazette*, **24**, 145-78.

Birky, C.W., Jr., and Skavaril, R.V. (1976) Maintenance of genetic homogeneity in systems with multiple genomes. *Genet. Res.*, **27**, 249-65.

Bowman, B.H. (1989) Non-clocklike evolution in the ribosomal RNAs fo bivalve molluscs. Ph.D. dissertation, Univeristy of California, Berkeley.

Bremer, K. (1988) The limits of amino acid sequence data in angiosperm phylogenetic reconstruction. *Evol.*, **42**, 795-803.

Britten, R.J., and Kohne, D.E. (1968) Repeated sequences in DNA. *Science*, **161**, 529-40.

Brown, D.D., Wensink, P.C., and Jordan, E. (1972) Comparison of the ribosomal DNA's of *Xenopus laevis* and *Xenopus mulleri*: the evolution of tandem genes. *J. Mol. Biol.*, **63**, 57-73.

Brown, D.D., and Sugimoto, K. (1974) The structure and evolution of ribosomal and 5S DNAs in *Xenopus laevis* and *Xenopus mulleri*. *Cold Spring Harbor Symp. Quant. Biol.*, **38**, 501-5.

Brown, R.W. (1956) Palm-like plants from the Dolores Formation (Triassic) in southwestern Colorado. *U.S. Geological Survey Professional Papers* **274H**, 205-9.

Buchheim, M.A., Turnel, M., Zimmer, E.A., and Chapman, R.L. (1990) Phylogeny of *Chlamydomonas* (Chlorophyta): an investigation based on cladistic analysis of nuclear 18S rRNA sequence data. *J. Phycol.*, **26**, in press.

Burger, W.C. (1977) The piperales and the monocots. *Bot. Rev.*, **43**, 345-93.

Burger, W.C. (1981) Heresy revived: the monocot theory of angiosperm origin. *Evol. Theory*, **5**, 189-225.

Callan, H.G. (1967) The organization of genetic units in chromosomes. *J. Cell Sci.*, **2**, 1-7.

Camin, J.H., and Sokal, R.R. (1965) A method for deducing branching sequences in phylogeny. *Evol.*, **19**, 311-26.

Cech, T.R. (1983) RNA splicing: three themes with variations. *Cell*, **34**, 713-6.

Chapman, R.L., and Avery, D.W. (1989) Nuclear ribosomal RNA genes and the phylogeny of the Trentepohliales. *J. Phycol.*, **25** (suppl), 25.

Chirgwin, J.M., Przybyla, A.E., MacDonald, R.J., and Rutter, W.J. (1979) Isolation of biologically active rebonucleic acid from source enriched in ribonucleases. *Biochem.*, **18**, 5294-9.

Clark, C.G., Tague, B.W., Ware, V.C. and Gerbi, S.A. (1984) *Xenopus laevis* 28S ribosomal RNA: A secondary structure model and its evolutionary and functional implications. *Nucl. Acids Res.*, **12**, 6197-220.

Coen, E., Thoday, J.M., and Dover, G. (1982) Rate of turnover of structural variants in the rDNA gene family of *Drosophila melanogaster*. *Nature*, **295**, 564-8.

Crane, P.R. (1985) Phylogenetic analysis of seed plants and the origin of angiosperms. *Ann. Missouri Bot. Gard.*, **72**, 716-93.

Cronquist, A. (1968) The evolution and classification of flowering plants, Houghton Mifflin, Boston.

Dahlberg, A.E. (1989) The functional role of ribosomal RNA in protein synthesis. *Cell*, **57**, 525-9.

DeBorde, D.C., Naeve, C.W., Herlocher, M.L., and Maassab, H.F. (1986) Resolution of a common RNA sequencing ambiguity by terminal deoxynucleotidyl transferase. *Annal. Biochem.*, **157**, 275-82.

Devereux, J., Haeverli, P., and Smithies, O. (1984) A comprehensive set of sequence analysis programs for the VAX. *Nucl. Acids Res.*, **12**, 387-95.

DeWinter, R.F.J., and Moss, T. (1986) The ribosomal spacer in *Xenopus laevis* is transcribed as part of the primary ribosomal RNA. *Nucl. Acids Res.*, **14**, 6041-51.

Doebley, J., Durbin, M., Golenberg, E.M., Clegg, M.T., and Ma, D.P. (1990) Evolutionary analysis of the large subunit of carboxylase (*rbcL*) nucleotide sequence among the grasses (Gramineae). *Evol.*, in press.

Donoghue, M.J. (1989) Phylogenies and the analysis of evolutionary sequences with examples from seed plants. *Evol.*, **43**, 1137-556.

Donoghue, M.J., and Doyle, J.A. (1989a) Phylogenetic studies of seed plants and angiosperms based on morphological characters. *in* The Hierarchy of

Life (eds B. Fernholm, K. Bremer and H. Jörnvall), Elsevier Science Publishers B.V., Amsterdam, pp. 181-95.

Donoghue, M.J., and Doyle, J.A. (1989b) Phylogenetic analysis of angiosperms and the relationships of Hamamelidae. *in* Evolution, systematics, and fossil history of the Hamamelidae (eds P.R. Crane and S. Blackmore) Clarendon Press, Oxford.

Donoghue, M.J., Doyle, J.A., Gauthier, J., Kluge, A.G., and Rowe, T. (1989) The importance of fossils in phylogeny reconstruction. *Ann. Rev. Ecol. Syst.*, **20**, 431-60.

Dover, G. (1982) Molecular drive: A cohesive mode of species evolution. *Nature*, **299**, 111-7.

Dover, G.A., and Flavell, R.B. (1984) Molecular coevolution: DNA divergence and the maintenance of function. *Cell*, **38**, 622-3.

Doyle, J.A., and Donoghue, M.J. (1986) Seed plant phylogeny and the origin of angiosperms: an experimental cladistic approach. *Bot. Rev.*, **52**, 321-431.

Doyle, J.J, and Beachy, R.N. (1985) Ribosomal gene variation in soybean (*Glycine*) and its relatives. *Theor. Appl. Genet.*, **70**, 369-76.

Dvorak, J. (1989) Evolution of multigene families: The ribosomal RNA loci of wheat and related species. *In* Plant Population Genetics, Breeding, and Genetic Resources (eds. A.H.D. Brown, M.T. Clegg, A.L. Kahler, and B.S. Weir), Sinauer Assoc., Sunderland, Mass, pp.83-97.

Eckenrode, V.K., Arnold, J., and Meagher, R.B. (1985) Comparison of the nucleotide sequence of soybean 18S rRNa with the sequences of other small-subunit rRNAs. *J. Mol. Evol.*, **21**, 259-69.

Edman, J.C., Kovacs, J.A., Masur, H., Santi, D.V., Elwood, H.J., and Sogin, M.L. (1988) Ribosomal RNA sequence shows *Pneumocystis carinii* to be a member of the fungi. *Nature* **334**, 519-22.

Endress, P.K. (1987) The early evolution of the angiosperm flower. *Tree*, **2**, 117-25.

Farris, J.S. (1970) Methods for computing Wagner trees. *Syst. Zool.*, **19**, 83-92.

Farris, J.S. (1977) Phylogenetic analysis under Dollo's law. *Syst. Zool.*, **26**, 77-88.

Farris, J.S. (1986) Hennig86 Manual, 41 Admiral Street, Port Jefferson Station, New York, New York 11776.

Felsenstein, J. (1982) Numerical methods for inferring evolutionary trees. *Quart. Rev. Biol.*, **57**, 379-404.

Felsenstein, J. (1985) Confidence limits on phylogenies: an approach using the bootstrap. *Evol.*, **39**, 783-91.

Felsenstein, J. (1988) Phylogenies from molecular sequences: inference and reliability. *Ann. Rev. Genet.*, **22**, 521-65.

Field, K.G., Olsen, G.J., Lane, D.J., Giovannoni, S.J., Ghiselin, M.T., Raff, E.C., Pace, N.R., and Raff, R.A. (1988) Molecular phylogeny of the animal kingdom. *Science*, **239**, 748-53.

Fitch, W.M. (1971) Toward defining the course of evolution: minimum change for a specific tree topology. *Syst. Zool.*, **20**, 406-16.

Flavell, R.B., O'Dell, M., and Thompson, W.F. (1983) Cytosine methylation of ribosomal RNA genes and nucleolus organizer activity in wheat. *in* Kew Chromosome Conference II (eds P.E. Brandham and M.D. Bennett), George Allen and Unwin, London, pp. 11-7.

Flavell, R.B. (1986) The structure and control of expression of ribosomal RNA genes. *Oxf. Surv. Pl. Mol. Cell Biol.*, **3**, 251-74.

Fogel, S., Mortimer, R., Lusnak, K., and Tavares, F. (1978) Meiotic gene conversion: a signal of the basic recombination event in yeast. *Cold Spring Harbor Symp. Quant. Biol.*, **43**, 1325-41.

Friedman, W.E. (1990) Double fertilization in *Ephedra*, a nonflowering seed plant: its bearing on the origin of angiosperms. *Science*, **247**, 951-4.

Friis, E.M, Chaloner, W.G., and Crane, P.R. (1986) Introduction to angiosperms. *in* The origin of angiosperms and their biological consequences, (eds E.M. Friis, W.G. Chaloner and P.R. Crane), Cambridge University Press, Cambridge, pp. 1-49.

Gerbi, S.A. (1985) Evolution of ribosomal RNA. *In* Molecular Evolutionary Genetics (ed R.J. MacIntyre), Plenum Press, New York, pp. 419-518.

Givens, J.F., and Phillips, R.L. (1976) The nucleolus organizer region of maize (*Zea mays* L.). *Chromosoma*, **57**, 103-17.

Glisin, V., Crkvenjakov, R., and Byus, C. (1974) Ribonucleic acid isolated by cesium chloride centrifugation. *Biochem.*, **13**, 2633-7.

Gould, F.W., and Shaw, R.B. (1983). Grass Systematics, Texas A&M University Press, College Station, Texas, pp. 111-30.

Gouy, M. and Li, W.-H. (1989a) Phylogenetic analysis based on rRNA sequences supports the archaebacterial rather than the eocyte tree. *Nature*, **339**, 145-7.

Gouy, M., and Li, W.-H. (1989b) Molecular Phylogeny of the kingdoms animalia, plantae, and fungi. *Mol. Biol. Evol.*, **6**, 109-22.

Grummt, I., Roth, E., and Paule, M.R. (1982) Ribosomal RNA transcription *in vitro* is species specific. *Nature*, **296**, 173-4.

Gutell, R., Weiser, B., Woese, C., and Noller, H. (1985) Comparative anatomy of 16S-like ribosomal RNA. *Prog. Nucl. Acid Res. Mol Biol.*, **32**, 155-216.

Gutell, R.R., and Fox, G.E. (1988) A compilation of large subunit RNA sequences presented in a structural format. *Nucl. Acids Res.*, **16**s, r175-203.

Hall, T.C., Ma, Y., Buchbinder, B.U., Pyne, J.W., Sun, S.M., and Bliss, F.A. (1978) Messenger RNA for G1 protein of french bean seeds: cell-free translation and product characterization. *Proc. Natl. Acad. Sci., USA*, **75**, 3196-3200.

Hamby, R.K., Issel, L.E., and Zimmer, E.A. (1987) Nuclear and organellar evolution in higher plants - ribosomal RNA as a marker molecule. *Genetics*, **116**, s20.

Hamby, R.K. and Zimmer, E.A. (1988) Ribosomal RNA sequences for inferring phylogeny within the grass family (Poaceae). *Pl. Syst. Evol.*, **160**, 29-37.

Hamby, R.K, Sims, L.E., Issel, L.E., and Zimmer, E.A. (1988) Direct ribosomal RNA sequencing: optimization of extraction and sequencing methods for work with higher plants. *Plant Mol. Biol. Rep.*, **6**, 175-92.

Hamby, R.K., and Zimmer, E.A. (1991) Ribosomal RNA as a phylogenetic tool in plant systematics. *in* Plant Molecular Systematics (eds D. Soltis, P. Soltis and J. Doyle), in press.

Hemleben, V., Ganal, M., Gerstner, J., Schiebel, K., and Torres, R.A. (1988) Organization and length heterogeneity of plant ribosomal RNA genes. *in*

Architecture of Eukaryotic Genes (ed G. Kahl), VCH, Weinheim, Fed. Rep. Germ., pp. 371-83.

Hendy, M.D., and Penny, D. (1982) Branch and bound algorithms to determine minimal evolutionary tree. *Math. Biosci.*, **59**, 277-90.

Hennig, W. (1950) Grundzute einer theorie der phylogenetischen systematik. Deutscher Zentralverlag, Berlin.

Hennig, W. (1965) Phylogenetic systematics. *Ann. Rev. Entomol.*, **10**, 97-116.

Hennig, W. (1966) Phylogenetic systematics, University of Illinois Press, Urbana, Ill.

Hickey, L.J., and Doyle, J.A. (1977) Early Cretaceous fossil evidence for angiosperm evolution. *Bot. Rev.*, **43**, 2-104.

Hillis, D.M., and Dixon, M.T. (1989) Vertebrate phylogeny: evidence from 28S ribosomal DNA sequences. *in* The Hierarchy of Life (eds B. Fernholm, K. Bremer and H. Jörnvall), Elsevier Science Publishers B.V., Amsterdam, pp. 355-67.

Hood, L., Campbell, J.H., and Elgin, S.C.R. (1975) The organization, expression, and evolution of antibody genes and other multigene families. *Ann. Rev. Genet.*, **9**, 305-53.

Hori, H., Lim, B.-L., and Osawa, S. (1985) Evolution of green plants as deduced from 5S rRNA sequences. *Proc. Natl. Acad. Sci. USA*, **82**, 820-3.

Hori, H., and Osawa, S. (1987) Origin and evolution of organisms as deduced from 5S ribosomal RNA sequences. *Mol. Biol. Evol.*, **4**, 445-72.

Hughes, N.F. (1976) Palaeo-succession of earliest angiosperm evolution. *Bot. Rev.*, **43**, 105-27.

Hui, A.S., Eaton, D.H., and de Boer, H.A. (1988) Mutagenesis at the mRNA decoding site in the 16S ribosomal RNA using the specialized ribosome system in *Escherichia coli*. *EMBO J.*, **7**, 4383-8.

Issel, L.E., Hamby, R.K., and Zimmer, E.A., (1990) Further development of a ribosomal RNA phylogeny for the grasses. *in* Proc. IV Int. Congress Syst. Evol. Biol., in press.

Ito, M. (1987) Phylogenetic systematics of the Nymphaeales. *Bot. Mag. Tokyo*, **100**, 17-35.

Jacob, W.F., Santer, M., and Dahlberg, A.E. (1987) A single base change in the Shine-Dalgarno region of 16S rRNA of *Escherichia coli* affects translation of many proteins. *Proc. Natl. Acad. Sci. USA*, **84**, 4757-61.

Jorgensen, R.A., and Cluster, P.D. (1988) Modes and tempos in the evolution of nuclear ribosomal DNA: new characters for evolutionary studies and new markers for genetic and population studies. *Ann. Missouri Bot. Gard.*, **75**, 1238-47.

Jukes, T.H. and Cantor, C.R., (1969) Evolution of protein molecules. *in* Mammalian Protein Metabolism (ed H.N. Munro), Academic Press, New York, pp. 21-123.

Jupe, E.R. (1988) DNA methylation, chromatin structure and expression of maize ribosomal genes. Ph.D. dissertation. Louisiana State University.

Jupe, E.R., Chapman, R.L., and Zimmer, E.A. (1988) Nuclear ribosomal RNA genes and algal phylogeny - the *Chlamydomonas* example. *Biosystems*, **21**, 223-30.

Jupe, E.R., and Zimmer, E.A. (1990) Unmethylated regions in the intergenic spacer of maize and teosinte ribosomal RNA genes. *Pl. Mol. Biol.*, in press.

Kantz, T.S., Theriot, E.C., Zimmer, E.A., and Chapman, R.L. (1990) The Pleurostrophyceae and Micromonadophyceae: a cladistic analysis of nuclear rRNA sequence data. *J. Phycol.*, **26,** in press.

Kellogg, E.A., and Campbell, C.S. (1987) Phylogenetic analysis of the Gramineae. *in* Grass Systematics and Evolution (eds T.R. Soderstrom, K.W. Hilu, C.S. Campbell and M.E. Barkworth), Smithsonian Institution Press, Washington, D.C., pp. 310-22.

Kimura, M. (1980) A simple method for estimating evolutionary rates of vase substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.*, **16**, 111-20.

Klein, H.L., and Petes, T.D., (1981) Intrachromosomal gene conversion in yeast. *Nature*, **289**, 144-8.

Kluge, A.G., and Farris, J.S. (1969) Quantitative phyletics and the evolution of anurans. *Syst. Zool.*, **18**, 1-32.

Knaak, C., Hamby, R.K., Arnold, M.L., LeBlanc, M.D., Chapman, R.L., and Zimmer, E.A. (1990) Ribosomal DNA variation and its use in plant biosystematics. *in* Proceedings of the 4th International Symposium on Plant Biosystematics, Academic Press, New York, in press.

Kumazaki, T., Hori, H., and Osawa, S. (1983) Phylogeny of protozoa deduced from 5S rRNA sequences. *J. Mol. Evol.*, **19**, 411-9.

Labhart, P., and Reeder, R.H. (1986) Characterization of three sites of RNA 3' end formation in the *Xenopus* ribosomal spacer. *Cell*, **45**, 431-3.

Lake, J.A. (1988) Origin of the eukaryotic nucleus determined by rate-invariant analysis of rRNA sequences. *Nature*, **331**, 184-6.

Lake, J.A. (1989) Origin of the eukaryotic nucleus determined by rate-invariant analyses of ribosomal RNA genes. *in* The Hierarchy of Life (eds B. Fernholm, K. Bremer and H. Jörnvall), Elsevier Science Publishers B.V., Amsterdam, pp. 87-101.

Lane, D.J., Pace, B., Olsen, G.J., Stahl, D.A., Sogin, M.L., and Pace, N.R. (1985) Rapid determination of 16S ribosomal sequences for phylogentic analyses. *Proc. Natl. Acad. Sci., USA*, **82**, 6955-9.

Li, W.-H., Luo, D.-C., and Wu, C.-I. (1985) Evolution of DNA sequences. *in* Molecular Evolutionary Genetics (ed R.J. MacIntyre), Plenum Press, New York, pp. 1-94.

Long, E.O., and Dawid, I.B. (1980) Repeated genes in eukaryotes. *Ann. Rev. Biochem.*, **49**, 727-64.

Maddison, W.P., and Maddison, D.R. (1990) MacClade v. 3.0: interactive analysis of phylogeny and character evolution, Sinauer Assoc., Sunderland, MA, in prep.

Martin, S., Zimmer, E.A., Davidson, W., Wilson, A.C., and Kan, Y.W. (1981) The untranslated regions of β-globin mRNA evolve at a functional rate in higher primates. *Cell*, **25**, 737-41.

Martin, P.G., Boulter, D., and Penny, D. (1985) Angiosperm phylogeny studied using sequences of five macromolecules. *Taxon.*, **34**, 393-400.

Martin, P.G., and Dowd, J.M. (1989) Phylogeny among the flowering plants as derived from amino acid sequence data. *in* The Hierarchy of Life (eds B. Fernholm, K. Bremer and H. Jörnvall), Elsevier Science Publishers B.V., Amsterdam, pp. 195-204.

Mascia, P.N., Rubenstein, I., Phillips, R.L., Wang, A.S., and Xiang, L.Z. (1981) Localization of the 5S rRNA genes and evidence for diversity in the 5S rDNA region of maize. *Gene*, **14**, 205-15.

Matteucci, M.D., and Caruthers, M.H. (1981) Synthesis of deoxyoligonucleotides on a polymer support. *J. Am. Chem. Soc.*, **103**, 3185-91.

McCarroll, R., Olsen, G.J., Stahl, Y.D., Woese, C.R., and Sogin, M.L. (1983) Nucleotide sequence of the *Dictyostelium discoideum* small-subunit ribosomal ribonucleic acid inferred from the gene sequence: evolutionary implications. *Biochem.*, **22**, 5858-68.

McClintock, B. (1934) The relationship of a particular chromosomal element to the development of the nucleoli in *Zea mays*. *Z. Zellforsch. Mikrosk. Anat.*, **21**, 294-328.

McMullen, M.D., Hunter, B., Phillips, R.L., and Rubenstein, I. (1986) The structure of the maize ribosomal DNA spacer region. *Nucl. Acids Res.*, **14**, 4953-68.

Meeuse, A.D.J. (1967) Again: the growth habit of the early angiosperms. *Acta Bot. Neerl.*, **16**, 33-41.

Messing, J., Carlson, J, Hagen, G., Rubenstein, I., and Oleson, A. (1984) Cloning and sequencing of the ribosomal RNA genes in maize: the 17S region. *DNA*, **3**, 31-40.

Meyen, S.V. (1984) Basic features of gymnosperm systematics and phylogeny as evidenced by the fossil record. *Bot. Rev.*, **50**, 1-112.

Moazed, D. and Noller, H.F. (1987) Interaction of antibiotics with functional sites in 16S ribosomal RNA. *Nature*, **327**, 389-94.

Moazed, D. and Noller, H.F. (1989) Interaction of tRNA with 23S rRNA in the ribosomal A, P, and E sites. *Cell*, **57**, 585-97.

Mullis, K.B., and Faloona, F.A. (1987) Specific synthesis of DNA *in vitro* via a polymerase-catalyzed chain reaction. *Meth. Enzym.*, **155**, 335-50.

Murgola, E.J., Hijazi, K.A., Goringer, H.U., and Dahlberg, A.E. (1988) Mutant 16S ribosomal RNA: a codon-specific translational suppressor. *Proc. Natl. Acad. Sci. USA*, **85**, 4162-5.

Nagylaki, T., and Petes, T.D. (1982) Intrachromosomal gene conversion and the maintenance of sequence homogeneity among repeated genes. *Genetics*, **100**, 315-37.

Nairn, C.J., and Ferl, R.J. (1988) The complete nucleotide sequence of the small-subunit ribosomal RNA coding region for the cycad *Zamia pumila*: phylogenetic applications. *J. Mol. Evol.*, **27**, 133-41.

Nickrent, D.L. and Franchina, C.R. (1989) Phylogenies of parasitic flowering plants (Santalales) using ribosomal RNA sequences. *Am. J. Bot.*, **76** (suppl.), 262.

Noller, H.F., Stern, S., Moazed, D., Powers, T., Svensson, P., and Changchien , L.-M. (1987) *Cold Spring Harbor Symp. Quant. Biol.*, **52**, 695-708.

Ohta, T. (1983) On the evolution of multigene families. *Theor. Popul. Biol.*, **23**, 216-40.

Ohta, T. (1984). Some models of gene conversion for treating the evolution of multigene families. *Genetics*, **106**, 517-28.

Pace, N.R., Olsen, G.J., and Woese, C.R. (1986) Ribosomal RNA phylogeny and the primary lines of evolutionary descent. *Cell*, **45**, 325-6.

Penny, D., Hendy, M.D., Zimmer, E.A., and Hamby, R.K. (1990) Trees from sequences: panacea or Pandora's box? *Australian Syst. Bot.*, **3**, in press.

Perasso, R., Baroin, A., Qu, L.H., Bachellerie, J.P., and Adoutte, A. (1989) Origin of the algae. *Nature*, **339**, 142-4.

Petes, T.D. (1980) Unequal meiotic recombination within tandem arrays of yeast ribosomal DNA genes. *Cell*, **19**, 765-74.

Phillips, R.L. (1978) Molecular cytogenetics of the nucleolus organizer region. *in* Maize Breeding and Genetics (ed D.B. Walden), John Wiley and Sons, New York, pp.711-41.

Prager, E.M., and Wilson, A.C. (1988) Ancient origin of lactalbumin from lysozyme: analysis of DNA and amino acid sequences. *J. Mol. Evol.*, **27**, 326-35.

Qu, L.H., Michot, B., and Bachellerie, J.-P. (1983) Improved methods for structure probing in large RNAs: a rapid 'heterologous' sequending approach is coupled to the direct mapping of nuclease accessible sites. *Nucl. Acids Res.*, **11**, 5903-20.

Raff, R.A., Field, K.G., Olsen, G.J., Giovannoni, S.J., Lane, D.J., Ghiselin, M.T., Pace, N.R., and Raff, E.C. (1989) Metazoan phylogeny based on analysis of 18S ribosomal RNA. *in* The Hierarchy of Life (eds B. Fernholm, K. Bremer and H. Jörnvall), Elsevier Science Publishers B.V., Amsterdam, pp. 247-60.

Raven, P.H., Evert, R.F., and Eichhorn, S.E. (1986) Biology of Plants, Worth Publishers, New York.

Razin, A., and Riggs, A.D. (1980) DNA methylation and gene function. *Science*, **210**, 604-10.

Reeder, R.H. (1984) Enhancers and ribosomal gene spacers. *Cell*, **38**, 349-51.

Rivin, C.J. , Cullis, C.A., and Walbot, V. (1986) Evaluating quantitative variation in the genome of *Zea mays*. *Genetics*, **113**, 1009-19.

Rogers, S.O., Honda, S., and Bendich, A.J. (1986) Variation in the ribosomal RNA genes among individuals of *Vicia faba*. *Pl. Mol. Biol.*, **6**, 339-45.

Rogers, S.O., and Bendich, A.J. (1987) Ribosomal RNA genes in plants: variability in copy number and in the intergenic spacer. *Pl. Mol. Biol.*, **9**, 509-20.

Rothwell, G.W. (1982) New interpretations of the earliest conifers. *Rev. Palaeobot. Palynol.*, **37**, 7-28.

Rubstov, P., Musakhanov, M., Zakharyev, V., Krayev, A., Skryabin, K., and Bayev, A. (1980) The structure of the yeast ribosomal RNA genes. I. The complete nucleotide sequence of the 18S ribosomal RNA gene from *Saccharomyces cerevisiae*. *Nucl. Acids Res.*, **8**, 5779-94.

Saghai-Maroof, M.A., Soliman, K.M., Jorgensen, R.A., and Allard, R.W. (1984) Ribosomal DNA spacer-length polymorphisms in barley: Mendelian inheritance, chromosomal location, and population dynamics. *Proc. Natl. Acad. Sci. USA*, **81**, 8014-8.

Saiki, R.K., Scharf, S.J., Faloona, F., Mullis, K.B., Horn, G.T., Erlich, H.A., and Arnheim, N. (1985) Enzymatic amplification of β-globin genomic sequences and restriction site analysis of sickle cell anemia. *Science*, **230**, 1350-4.

Saitou, N, and Nei, M. (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Bio. Evol.*, **4**, 406-25.

Salim, M., and Maden, B. (1981) Nucleotide sequence of *Xenopus laevis* 18S ribosomal RNA inferred from gene sequence. *Nature*, **291**, 205-8.

Sarich, V. M. and Wilson, A.C. (1967) Immunological time scale for hominid evolution. *Science*, **158**, 1200-3.

Schaal, B.A., and Learn, Jr., G.H. (1988) Ribosomal DNA variation within and among plant populations. *Ann. Missouri Bot. Gard.*, **75**, 1207-16.

Scherer, S. and Davis, R.W., (1980) Recombination of dispersed repeated DNA sequences in yeast. *Science*, **209**, 1380-4.

Schleifer, K.H., and Ludwig, W. (1989) Phylogenetic relationships among bacteria. *in* The Hierarchy of Life (eds B. Fernholm, K. Bremer and H. Jörnvall), Elsevier Science Publishers B.V., Amsterdam, pp. 103-17.

Shermoen, A.W., and Kiefer, B.I. (1975) Regulation in rDNA-deficient *Drosophila melanogaster*. *Cell*, **4**, 275-80.

Smith, G.P. (1974) Unequal crossover and the evolution of multigene families. *Cold Spring Harbor Symp. Quant. Biol.*, **38**, 507-13.

Smith, G.P. (1976) Evolution of repeated DNA sequences by unequal crossover. *Science*, **191**, 528-35.

Smith, J.M. (1989) Evolutionary genetics, Oxford University Press, Oxford.

Sogin, M.L., Edman, U., and Elwood, H. (1989) A single kingdom of eukaryotes. *in* The Hierarchy of Life (eds B. Fernholm, K. Bremer and H. Jörnvall), Elsevier Science Publishers B.V., Amsterdam, pp. 133-43.

Stebbins, G.L. (1974) Flowering plants evolution above the species level, Belknap Press, Cambridge, Mass.

Steffensen, D.M., and Patterson, E.B. (1979) Using translocations to map the 5S rRNA genes to chromosome 2L in maize. *Genetics*, **9** (suppl.), s123.

Stringer, S.L., Hudson, K., Blase, M.A., Walzer, P.D., Cushion, M.T., and Stringer, J.R. (1989) Sequence from ribosomal RNA of *Pneumocystis carinii* compared to those of four fungi suggests an ascomycetous affinity. *J. Protozool.*, **36**, 14S-6S.

Swofford, D.L. (1989) PAUP 3.0. Illinois Natural History Survey, 607 E. Peabody Dr., Champaign, IL 61820.

Swofford, D.L., and Olsen, G.J. (1990) Phylogeny construction. *in* Molecular systematics (eds D.M. Hillis and C. Moritz), Sinauer Assoc., Sunderland, MA, pp. 411-501.

Sytsma, K.J., and Schaal, B.A. (1985) Phylogenetics of the *Lisianthius skinneri* (Gentianaceae) species complex in Panama utilizing DNA restriction fragment analysis. *Evol.*, **39**, 594-608.

Takaiwa, F., Oono, K., and Sugiura, M. (1984) The complete nucleotide sequence of a rice 17S ribosomal RNA gene. *Nucl. Acids Res.*, **12**, 5441-8.

Takaiwa, F., Oono, K., Iida, Y., and Sugiura, M. (1985a) The complete nucleotide sequence of a rice 25S ribosomal RNA gene. *Gene*, **37**, 255-89.

Takaiwa, F., Oono, K., and Sugiura, M. (1985b) Nucleotide sequence of the 17S-25S spacer region from rice rDNA. *Plant Mol. Biol.*, **4**, 355-64.

Takhtajan, A. (1969) Flowering plants origin and dispersal, Smithsonian Institution Press, Washington, D.C.

Tanaka, Y., Dyer, T.A., and Brownlee, G.G. (1980) An improved direct RNA sequence method: its application to *Vicia faba* 5.8S ribosomal RNA. *Nucl. Acids Res.*, **8**, 1259-72.

Tartof, K.D. (1975) Redundant genes. *Ann. Rev. Genet.*, **9**, 355-85.

Taylor, D.W., and L.J. Hickey (1990) An Aptian plant with attached leaves and flowers: implications for angiosperm origin. *Science*, **247**, 702-4.

Taylor, T.N. (1981) Paleobotany an introduction to fossil plant biology, McGraw-Hill, New York.

Templeton, A.R. (1983) Phylogenetic inference from restriction endonuclease cleavage site maps with particular reference to the evolution of humans and the apes. *Evol.*, **37**, 221-44.

Thorne, R.F. (1976) A phylogenetic classification of the Angiospermae. *Evol. Biol.*, **9**, 35-106.

Tohdoh, N., and Sugiura, M. (1982) The complete nucleotide sequence of a 16S ribosomal RNA molecule from tobacco chloroplasts. *Gene*, **17**, 213-8.

Tolan, D., Amsden, A.B., Putney, S., Urdea, M., and Penhoet, E.E. (1984) The complete nucleotide sequence for rabbit muscle aldolase A messenger RNA. *J. Biol. Chem.*, **259**, 1127-31.

Torczynski, R., Bollon, A., and Fuke, M. (1983) The complete nucleotide sequence of the rat 18S ribosomal RNA gene and comparison with the respective yeast and frog genes. *Nucl. Acids Res.*, **11**, 4879-90.

Trifonov, E.N. (1987) Translation framing code and frame-monitoring mechanism as suggested by the analysis of mRNA and 16S rRNA nucleotide sequences. *J. Mol. Biol.,* **194**, 643-52.

Troitsky, A.V., Melekhovets, Y.F., Rakhimove, G.M., Bobrova, V.K., Valiejo-Roman, K.M., and Antonov, A.S. (1990) Angiosperms origin and early stages of seed plant evolution deduced from rRNA sequence comparisons. Ms. submitted.

Turner, S., Burger-Wiersma, T., Giovannoni, S.J., Mur, L.R., and Pace, N.R. (1989) The relationship of a prochlorophyte *Prochlorothrix hollandica* to green chloroplasts. *Nature*, **337**, 380-2.

Vincentz, M., and Flavell, R.B. (1989) Mapping of ribosomal RNA transcripts in wheat. *Plant Cell*, **1**, 579-89.

Vossbrinck, C.R., Maddox, J.V., Friedman, S., Debrunner-Vossbrinck, B.A., and Woese, C.R. (1987) Ribosomal RNA sequence suggests microsporidia are extremely ancient eukaryotes. *Nature*, **326**, 411-4.

Walker, J.W., and Walker, A.G. (1984) Ultrastructure of lower Cretaceous angiosperm pollen and the origin and early evolution of flowering plants. *Ann. Missouri Bot. Gdn.*, **71**, 464-521.

Watanabe, J.-I., Hori, H., Tanabe, K., and Nakamura, Y. (1989) 5S ribosomal RNA sequence of *Pneumocystis carinii* and its phylogenetic association with "Rhizopoda/Myxomycota/Zygomycota group" *J. Protozool.*, **36**, 16S-7S.

Watson, L., Clifford, H.T., and Dallwitz, M.J. (1985) The classification of Poaceae: subfamilies and supertribes. *Austral. J. Bot.*, **33**, 433-84.

Weiss, R.B., Dunn, D.M., Atkins, J.F., and Gesteland, R.F. (1987) Slippery runs, shifty stops, backward steps, and forward hops: -2, -1, +1, +2, +5, and +6 ribosomal frameshifting. *Cold Spring Harbor Symp. Quant. Biol.,* **52**, 687-93.

Weiss, R.B., Dunn, D.M., Dahlberg, A.E., Atkins, J.F., and Gesteland, R.F. (1988) Reading frame switch caused by base-pair formation between the 3' end of 16S rRNA and the mRNA during elongation of protein synthesis in *Escherichia coli*. *EMBO J.*, **7**, 1503-7.

Wettstein, R.R. von (1907) Handbuch der systematischen Botanik, Second edition. Franz Deuticke, Leipzig.

Wheeler, W.C., and Honeycutt, R.L. (1988) Paired sequence difference in ribosomal RNAs: evolutionary and phylogenetic implications. *Mol. Biol. Evol.*, **5**, 90-6.

Wilcoxon, F., and Wilcox, R.A. (1964) Some rapid approximate statistical procedures, Lederle Laboratories, Pearle River, New York.

Wiley, E.O. (1981) Phylogenetics, John Wiley & Sons, New York.

Woese, C.R. (1987) Bacterial evolution *Microbiol. Rev.*, **51**, 221-71.

Wolfe, K.H., Gouy, M., Yang. Y.-W., Sharp, P.M., and Li, W.-H. (1989) Date of the monocot-dicot divergence estimated from chloroplast DNA sequence data. *Proc. Natl. Acad. Sci. USA*, **86**, 6201-5.

Yakura, K., Kato, A., and Tanifuji, S. (1984) Length heterogeneity of the large spacer of *Vicia faba* is due to the differing number of a 325 bp repetitive sequence elements. *Mol. Gen. Genet.*, **193**, 400-5.

Youvan, D., and Hearst, J. (1981) A sequence from *Drosophila melanogaster* 18S rRNA bearing the conserved hypermodified nucleoside amψ. *Nucl. Acids Res.*, **9**, 1723-41.

Zechman, F.W., Theriot, E.C., Zimmer, E.A., and Chapman, R.L., (1990) Phylogeny of the Ulvophyceae (Chlorophyta): cladistic analysis of nuclear-encoded rRNA sequence data. *J. Phycol.*, **26**, in press.

Zimmer, E.A., Martin, S.L., Beverley, S.M, Kan, Y.W., and Wilson, A.C. (1980) Rapid duplication and loss of genes coding for the α chains of hemoglobin. *Proc. Natl. Acad. Sci. USA*, **77**, 2158-62.

Zimmer, E.A. and Sims, L.E. (1985) Ribosomal gene yardsticks of plant molecular evolution: direct RNA sequencing analysis. *ICPMB Abstr.*, **1**, 88.

Zimmer, E.A., Jupe, E.R., and Walbot, V. (1988) Ribosomal gene structure, variation, and inheritance in maize and its ancestors. *Genetics*, **120**, 1125-36.

Zimmer, E.A., Hamby, R.K., Arnold, M.L., LeBlanc, D.A., and Theriot, E.C. (1989) Ribosomal RNA phylogenies and flowering plant evolution. *in* The Hierarchy of Life (eds B. Fernholm, K. Bremer and H. Jörnvall), Elsevier Science Publishers B.V., Amsterdam, pp. 205-14.

Zuckerkandl, E. and Pauling, L. (1962) Molecular disease, evolution and genic heterogeneity. *in* Horizons in Biochemistry (eds M. Kasha and B. Pullman), Academic Press, New York.

APPENDIX 1

A computer program to convert sequence data into distances.

This program will read sequence files that have standard DNA format, i.e., G, A, T, C, R, Y, W, S, M, K and X, but requires that the sequence data are not interleaved (PAUP 2.4 format). It is formatted for up to 100 taxa and up to 2000 characters; the taxon names cannot exceed 15 characters. This program cannot run on a PC unless perhaps the PC has tremendous amounts of memory; it is currently executing on the VAX. It requires a FORTRAN compiler wherever it runs. The program prompts for the number of characters and taxa and allows gap data to be placed at the end of the sequence data. It will ignore the gap data when calculating distances. It will convert the sequence data into distances according to three formulae: dissimilarity, Jukes-Cantor (1969) and Kimura's (1980) two-parameter distances. The program creates a data matrix of distances and places it in a file on the mainframe. The files are called simple.nj, jukes.nj and kimura.nj. These files can then be used as input files for the neighbor-joining program of Saitou and Nei (1987). The program also calculates the number of sites compared between each pair of taxa and tallies the number of transition and transversion events between each pair of taxa. It also counts the number of missing or ambiguous positions for each taxon. The pairwise information and the missing and ambiguous counts are placed in a file called Summary.dat.

```
      CHARACTER*25 INFILE
      INTEGER NUMCOMP(100,100),SITE(100,2000),NAME(100,15)
      INTEGER NTN(100,100),NTV(100,100),NKNOWN(100)
      INTEGER PROD,DIFF,NUNKNOWN(100)
      REAL DATA(100,2000),COMP(100,100),SIMPLE(100,100)
      REAL JUKES(100,100),KIMURA(100,100)
      REAL P(100,100),Q(100,100)
      WRITE (*,100)
 100  FORMAT (1X,'HOW MANY TAXA?')
      READ (5,110) NTAX
 110  FORMAT (1I4)
      WRITE (*,120)
 120  FORMAT (1X,'HOW MANY CHARACTERS (INCLUDING GAPS)?')
      READ (5,130) NCHAR
 130  FORMAT (1I5)
      WRITE (*,131)
 131  FORMAT(1X,'HOW MANY GAP SITES ARE AT THE END OF THE DATA SET?')
      READ(5,132)NGAP
 132  FORMAT(1I5)
      NCOMP=NCHAR-NGAP
      WRITE (*,140)
 140  FORMAT(1X,'INPUT FILE?')
      READ (5,150)INFILE
 150  FORMAT (1A25)
      OPEN (UNIT=10,FILE=INFILE,STATUS='OLD')
      DO 160 I=1,NTAX
C     The format of the input file must be changed to accommodate
C     files different from those with 78 characters per line
      READ(10,170)(NAME(I,J),J=1,15),(DATA(I,K),K=1,NCHAR)
 170  FORMAT (15A1,/,21(78A1,/),76A1)
 160  CONTINUE
      CLOSE (10)
      DO 777 J=1,NTAX
        NKNOWN(J)=0
        DO 780 K=1,NCHAR
          IF (DATA(J,K).EQ.'G') GO TO 790
          IF (DATA(J,K).EQ.'A') GO TO 791
          IF (DATA(J,K).EQ.'T') GO TO 792
          IF (DATA(J,K).EQ.'C') GO TO 793
          IF (DATA(J,K).EQ.'R') GO TO 794
          IF (DATA(J,K).EQ.'Y') GO TO 795
          SITE(J,K)=5
          GO TO 780
 790      SITE(J,K)=0
          NKNOWN(J)=NKNOWN(J)+1
          GO TO 780
 791      SITE(J,K)=1
          NKNOWN(J)=NKNOWN(J)+1
          GO TO 780
```

```
792       SITE(J,K)=8
          NKNOWN(J)=NKNOWN(J)+1
          GO TO 780
793       SITE(J,K)=9
          NKNOWN(J)=NKNOWN(J)+1
          GO TO 780
794       SITE(J,K)=3
          GO TO 780
795       SITE(J,K)=6
780     CONTINUE
777   CONTINUE
180   DO 190 J=1,NTAX-1
       DO 200 K=2,NTAX
        IF (K.LE.J) GO TO 200
        NTV(J,K)=0
        NTN(J,K)=0
        COMP(J,K)=0.
        DO 210 I=1,NCOMP
         IF (SITE(J,I).EQ.5) GO TO 210
         IF (SITE(K,I).EQ.5) GO TO 210
         DIFF=ABS(SITE(J,I)-SITE(K,I))
         PROD=SITE(J,I)*SITE(K,I)
         IF (DIFF.EQ.0) GO TO 220
         IF (DIFF.EQ.1) GO TO 230
         IF (DIFF.GE.5) GO TO 240
         IF (PROD.EQ.18) GO TO 240
         GO TO 210
220      IF (PROD.GT.2.AND.PROD.LT.63) GO TO 210
         COMP(J,K)=COMP(J,K)+1.
         GO TO 210
230      COMP(J,K)=COMP(J,K)+1.
         NTN(J,K)=NTN(J,K)+1
         GO TO 210
240      COMP(J,K)=COMP(J,K)+1.
         NTV(J,K)=NTV(J,K)+1
210     CONTINUE
       P(J,K)=NTN(J,K)/COMP(J,K)
       Q(J,K)=NTV(J,K)/COMP(J,K)
       SIMPLE(J,K)=P(J,K)+Q(J,K)
       JUKES(J,K)=-0.75*LOG(1.-4./3.*(P(J,K)+Q(J,K)))
       KIMURA(J,K)=-0.5*LOG((1.-2.*P(J,K)-Q(J,K))*SQRT(1.-2.*Q(J,K)))
       NUMCOMP(J,K)=COMP(J,K)
200    CONTINUE
190   CONTINUE
      OPEN (UNIT=10, FILE='SIMPLE.NJ',STATUS='NEW')
      DO 211 M=1,NTAX
        WRITE(10,213)(NAME(M,I),I=1,15)
213     FORMAT(1X,15A1)
211   CONTINUE
```

```
      WRITE(10,217)
217   FORMAT(1X)
       J=1
505    DO 510 K=J+1,NTAX
        WRITE(10,500) SIMPLE(J,K)
500     FORMAT (1X,F8.5)
510    CONTINUE
      WRITE(10,515)
515   FORMAT(1X)
      J=J+1
      IF (J.LT.NTAX) GO TO 505
      CLOSE (10)
      OPEN (UNIT=10,FILE='JUKES.NJ',STATUS='NEW')
      DO 223 M=1,NTAX
        WRITE(10,221)(NAME(M,I),I=1,15)
221     FORMAT(1X,15A1)
223   CONTINUE
      WRITE(10,227)
227   FORMAT(1X)
      J=1
630    DO 635 K=J+1,NTAX
        WRITE(10,640) JUKES(J,K)
640     FORMAT(1X,F8.5)
635    CONTINUE
      WRITE (10,645)
645   FORMAT(1X)
      J=J+1
      IF (J.LT.NTAX) GO TO 630
      CLOSE (10)
      OPEN (UNIT=10,FILE='KIMURA.NJ',STATUS='NEW')
      DO 231 M=1,NTAX
       WRITE(10,233)(NAME(M,I),I=1,15)
233    FORMAT(1X,15A1)
231   CONTINUE
      WRITE(10,237)
237   FORMAT(1X)
      J=1
725    DO 730 K=J+1,NTAX
        WRITE (10,735)KIMURA(J,K)
735     FORMAT(1X,F8.5)
730    CONTINUE
      WRITE(10,740)
740   FORMAT(1X)
      J=J+1
      IF (J.LT.NTAX) GO TO 725
      CLOSE (10)
      OPEN (UNIT=10,FILE='SUMMARY.DAT',STATUS='NEW')
      DO 901 L=1,NTAX
        NUNKNOWN(L)=NCHAR-NKNOWN(L)
        WRITE(10,903)(NAME(L,I),I=1,15),NUNKNOWN(L)
903     FORMAT(1X,'For ',15A1,2X,I5,' sequence positions
     * are missing or uncertain.')
901   CONTINUE
      WRITE(10,764)
      WRITE(10,287)
287   FORMAT(7X,'SPECIES COMPARED',14X,'SITES',2X,'TN',3X,'TV',
     *2X,'SIMPLE',3X,'J-C',6X,'KIM 2-P')
764   FORMAT(1X)
      J=1
813    DO 811 K=J+1,NTAX
         WRITE(10,812)(NAME(J,I),I=1,15),(NAME(K,I),I=1,15),
     *NUMCOMP(J,K),NTN(J,K),NTV(J,K),SIMPLE(J,K),JUKES(J,K),
     *KIMURA(J,K)
812      FORMAT(1X,15A1,' v. ',15A1,1X,I5,1X,I4,1X,I4,2X,F7.5,
     *2X,F7.5,2X,F7.5)
811    CONTINUE
      WRITE(10,814)
814   FORMAT(1X)
      J=J+1
      IF (J.LT.NTAX) GO TO 813
      STOP
      END
```

APPENDIX 2

Nucleotide sequence data converted to distances for 60 taxa.

Distance = overall dissimilarity. d = #different/#compared

Soy
Pea
Strawberry
Spinach
Chickweed
Saururus
Peperomia
Magnolia
Hedycarya
Illicium
Drimys
Sycamore
Sweetgum
Ranunculus
Parsley
Chloranthus
Sagittaria
Colocasia
Echinodorus
Najas
Potamogeton
Pistia
Zea
Tripsacum
Sugarcane
Sorghum
Rice
Barley
Oats
Wheat
Bamboo
Nymphaea
Cabomba
Welwitschia
Gnetum
E.Tweediana
E.Distachya
Pine

Juniper
Cryptomeria
Cycad
Encephalartos
Zamia
ZamiaO
Ginkgo
Equisetum
Psilotum
Ceratophyllum
Sabal
Nelumbo
Hosta
Nuphar
Liriodendron
Asimina
Calycanthus
Piper
Saruma
Trochodendron
Aristolochia
Barclaya

```
0.01117 0.03444 0.04330 0.05488 0.02860 0.04570 0.03156 0.03387
0.03465 0.04131 0.02814 0.02607 0.03508 0.03659 0.02644 0.03545
0.03780 0.05109 0.06529 0.06520 0.05322 0.05941 0.06022 0.05775
0.05734 0.05718 0.05701 0.05836 0.05825 0.06089 0.04174 0.05672
0.09689 0.07954 0.07595 0.07776 0.07265 0.06562 0.06325 0.07363
0.06688 0.07503 0.07072 0.06328 0.09543 0.08722 0.04013 0.04407
0.03333 0.05114 0.04476 0.03390 0.04134 0.02901 0.03413 0.02939
0.03041 0.02903 0.04854

0.04123 0.05078 0.06211 0.03383 0.05216 0.03387 0.04014 0.03987
0.05058 0.03481 0.03340 0.03966 0.04526 0.03678 0.04300 0.04935
0.05988 0.07598 0.07587 0.06414 0.07277 0.06985 0.07188 0.06564
0.06919 0.06726 0.07138 0.07021 0.07133 0.04840 0.05913 0.10752
0.08580 0.08512 0.08674 0.07867 0.06993 0.06958 0.08028 0.07046
0.08384 0.07588 0.07088 0.10420 0.09677 0.04416 0.05299 0.03914
0.06620 0.05216 0.03607 0.04619 0.03458 0.03952 0.03611 0.03414
0.03510 0.05351

0.03879 0.03737 0.02712 0.04409 0.02895 0.02969 0.02944 0.04479
0.02888 0.02715 0.03301 0.04014 0.02350 0.03889 0.04188 0.05145
0.05879 0.05871 0.05205 0.05874 0.05752 0.05375 0.05422 0.05423
0.05056 0.05072 0.05062 0.05816 0.03519 0.05004 0.09511 0.07637
0.06899 0.07309 0.06842 0.06743 0.06654 0.07248 0.06332 0.07474
0.06744 0.05666 0.08758 0.07617 0.03108 0.04934 0.02794 0.04840
```

```
0.04034 0.02867 0.03312 0.02602 0.03291 0.02424 0.03150 0.03106
0.04469

0.04333 0.04294 0.05186 0.03924 0.04783 0.03629 0.05278 0.04335
0.03806 0.04264 0.03969 0.03675 0.05204 0.05107 0.06098 0.06722
0.06713 0.05887 0.06485 0.06482 0.05912 0.05540 0.05564 0.05749
0.05967 0.05734 0.06300 0.05300 0.06449 0.10960 0.09064 0.08414
0.08616 0.08014 0.07114 0.07084 0.08622 0.07686 0.08316 0.07823
0.06582 0.10204 0.09294 0.04264 0.05899 0.04733 0.06303 0.05217
0.04364 0.04523 0.04380 0.04355 0.03953 0.04420 0.04260 0.05198

0.03993 0.05909 0.04530 0.04261 0.05090 0.06182 0.04897 0.04429
0.04994 0.05289 0.04334 0.05988 0.05513 0.06364 0.06863 0.06854
0.06118 0.07085 0.06870 0.06756 0.06318 0.06691 0.05994 0.06183
0.06423 0.06983 0.04907 0.06708 0.10491 0.08897 0.08274 0.08651
0.08384 0.07717 0.07198 0.08587 0.06899 0.08022 0.07616 0.06594
0.09555 0.08668 0.04608 0.06441 0.04422 0.06034 0.05259 0.04960
0.04891 0.04388 0.05292 0.04542 0.04633 0.04222 0.05603

0.03833 0.03212 0.04090 0.02902 0.04295 0.02257 0.02976 0.02857
0.04255 0.02410 0.03710 0.03539 0.04326 0.05964 0.05964 0.04472
0.05352 0.05728 0.05519 0.05381 0.05638 0.04735 0.05053 0.05327
0.05530 0.03866 0.04989 0.07719 0.07644 0.06365 0.06745 0.06094
0.05740 0.05470 0.06052 0.06311 0.06431 0.06362 0.05142 0.07928
0.06298 0.03330 0.04374 0.03226 0.04314 0.03850 0.03204 0.03452
0.02914 0.02620 0.02376 0.03560 0.02617 0.03794

0.03562 0.04352 0.03902 0.05429 0.03954 0.04309 0.04590 0.04749
0.03672 0.05168 0.04959 0.05950 0.06930 0.06921 0.05722 0.06059
0.05947 0.05940 0.05745 0.06080 0.06095 0.06168 0.06052 0.06842
0.04876 0.05689 0.09718 0.07747 0.07189 0.07687 0.07571 0.07215
0.06996 0.08240 0.06939 0.07943 0.07757 0.06250 0.08555 0.07439
0.04112 0.05563 0.04636 0.05722 0.05263 0.03578 0.04088 0.03711
0.02951 0.03416 0.04513 0.03887 0.04981

0.02294 0.01688 0.04228 0.01413 0.02093 0.02710 0.02664 0.01685
0.03420 0.03369 0.04452 0.05834 0.05827 0.04388 0.05219 0.05113
0.04755 0.04529 0.04748 0.05087 0.04956 0.05046 0.05521 0.03661
0.04914 0.09052 0.07915 0.06276 0.07059 0.06419 0.06179 0.06140
0.06820 0.06513 0.06539 0.06935 0.05316 0.09097 0.08080 0.02837
0.03986 0.02766 0.04682 0.04041 0.00981 0.01404 0.01550 0.02381
0.01753 0.02203 0.02957 0.03962

0.03421 0.03560 0.02005 0.02794 0.02828 0.03527 0.02106 0.03616
0.03094 0.04526 0.04750 0.04750 0.04181 0.05575 0.05888 0.05391
0.05146 0.05800 0.04908 0.04829 0.05541 0.05773 0.04395 0.05904
0.08993 0.07791 0.06566 0.06969 0.07279 0.06672 0.06111 0.06903
0.06977 0.06470 0.06915 0.05536 0.08799 0.07628 0.02885 0.04819
0.02766 0.04170 0.04757 0.02057 0.03172 0.03093 0.03361 0.02468
```

```
0.02718 0.02634 0.04549

0.04853 0.01718 0.02021 0.02917 0.03228 0.01882 0.03977 0.03213
0.04683 0.06149 0.06149 0.04204 0.04675 0.04996 0.04545 0.04285
0.04312 0.04444 0.04535 0.04490 0.05285 0.03996 0.04618 0.08692
0.07862 0.05627 0.06574 0.06326 0.05836 0.05632 0.06782 0.06437
0.06641 0.06694 0.05153 0.09121 0.07341 0.03020 0.04432 0.02907
0.04146 0.03922 0.02352 0.02174 0.01994 0.02880 0.02067 0.02215
0.03315 0.03569

0.04178 0.03824 0.04528 0.05058 0.03190 0.05041 0.05100 0.06552
0.06564 0.06555 0.05949 0.07779 0.07395 0.07446 0.07062 0.07244
0.06843 0.06794 0.07099 0.07347 0.04845 0.05898 0.10572 0.08650
0.08981 0.08707 0.08618 0.08286 0.08673 0.08952 0.07213 0.09019
0.08503 0.07500 0.10703 0.09441 0.04315 0.06311 0.04351 0.06343
0.05223 0.03493 0.04528 0.03520 0.04219 0.03653 0.03713 0.03666
0.05527

0.01889 0.02880 0.03085 0.01719 0.03849 0.03043 0.04745 0.05635
0.05631 0.04092 0.05849 0.05506 0.05166 0.04858 0.05201 0.05466
0.05544 0.05314 0.05534 0.03345 0.04605 0.09302 0.07460 0.06700
0.07381 0.07281 0.06241 0.06193 0.07216 0.05970 0.06713 0.06196
0.05646 0.09541 0.08238 0.02862 0.04748 0.02227 0.04523 0.03814
0.01547 0.02133 0.01143 0.02502 0.01862 0.01702 0.02286 0.04009

0.02939 0.03159 0.01703 0.03468 0.03562 0.04997 0.06047 0.06039
0.04868 0.05067 0.04895 0.05010 0.04711 0.04816 0.05170 0.05107
0.05174 0.05866 0.03811 0.04901 0.09577 0.07980 0.07037 0.07292
0.07249 0.06620 0.06401 0.07284 0.06115 0.07081 0.06464 0.05574
0.09032 0.07935 0.03236 0.04068 0.02847 0.04110 0.04145 0.02083
0.02506 0.02137 0.03014 0.01807 0.02513 0.02713 0.04067

0.03800 0.02276 0.04365 0.04074 0.05628 0.06254 0.06246 0.05479
0.06259 0.06110 0.05765 0.05242 0.05745 0.05577 0.05830 0.05730
0.06065 0.04199 0.05626 0.09966 0.08079 0.07687 0.08208 0.07358
0.07065 0.07054 0.07966 0.06703 0.07900 0.07132 0.06362 0.10059
0.08824 0.03675 0.04819 0.02971 0.05400 0.04329 0.02464 0.03277
0.02273 0.03341 0.02613 0.02597 0.02683 0.04878

0.03062 0.04330 0.04620 0.06054 0.07220 0.07210 0.05479 0.05956
0.06054 0.05874 0.05352 0.05604 0.05582 0.05574 0.05820 0.06470
0.04914 0.05719 0.10602 0.08365 0.07850 0.08165 0.07986 0.07407
0.07104 0.08274 0.07543 0.08245 0.07698 0.06483 0.10592 0.08877
0.03929 0.05597 0.04276 0.05878 0.05377 0.03388 0.03793 0.03081
0.04224 0.03495 0.03052 0.03701 0.04997

0.03391 0.03163 0.04625 0.05284 0.05284 0.04041 0.05068 0.05123
0.04796 0.04502 0.04672 0.04701 0.04716 0.04975 0.05324 0.03445
0.04796 0.09337 0.07303 0.06623 0.07062 0.07143 0.06323 0.06106
```

```
0.06988 0.05900 0.06612 0.06452 0.05216 0.08667 0.07505 0.02789
0.03918 0.02312 0.03976 0.03751 0.01498 0.02201 0.01293 0.02552
0.01476 0.01760 0.02159 0.03597

0.03947 0.04348 0.06275 0.06266 0.05014 0.06216 0.06107 0.06027
0.05810 0.06192 0.06135 0.06262 0.06138 0.06597 0.04615 0.05752
0.09358 0.08029 0.08057 0.08718 0.08179 0.07545 0.07065 0.08014
0.06667 0.08095 0.07427 0.06284 0.09732 0.09097 0.04180 0.04727
0.04069 0.05567 0.04825 0.03454 0.03516 0.03291 0.03730 0.03400
0.03583 0.03869 0.05052

0.04943 0.06154 0.06146 0.01889 0.05056 0.05024 0.04997 0.04545
0.04681 0.05163 0.05052 0.05306 0.04952 0.04641 0.05707 0.09344
0.07638 0.07297 0.07824 0.07279 0.06704 0.06417 0.07049 0.05860
0.06865 0.06616 0.05578 0.08752 0.07756 0.03287 0.03107 0.03237
0.04510 0.04936 0.03078 0.03715 0.02826 0.03543 0.02761 0.03486
0.02857 0.04908

0.05465 0.05393 0.05349 0.06974 0.07197 0.06741 0.06631 0.06855
0.06725 0.06788 0.06948 0.07237 0.05046 0.06830 0.10652 0.08705
0.08718 0.09298 0.08294 0.07967 0.07466 0.08581 0.06778 0.07730
0.07646 0.07027 0.09845 0.09320 0.04598 0.05764 0.04517 0.06174
0.05236 0.04501 0.04824 0.04305 0.05156 0.04138 0.05003 0.04705
0.06105

0.00062 0.06134 0.08206 0.08745 0.08305 0.08077 0.08699 0.07400
0.07785 0.08062 0.08702 0.06734 0.07237 0.11149 0.08537 0.09552
0.09633 0.08705 0.08707 0.08306 0.09334 0.07338 0.09142 0.08371
0.07878 0.10461 0.09908 0.05164 0.07477 0.05725 0.07498 0.06894
0.05201 0.05896 0.04944 0.05506 0.04882 0.05996 0.05238 0.07434

0.06125 0.08262 0.08732 0.08293 0.08066 0.08686 0.07391 0.07775
0.08051 0.08690 0.06734 0.07237 0.11149 0.08524 0.09539 0.09627
0.08693 0.08696 0.08296 0.09321 0.07338 0.09131 0.08359 0.07867
0.10447 0.09961 0.05160 0.07467 0.05716 0.07488 0.06894 0.05201
0.05887 0.04935 0.05497 0.04874 0.05988 0.05231 0.07424

0.05081 0.05445 0.04981 0.04789 0.04823 0.05330 0.05207 0.05539
0.05199 0.05464 0.05861 0.09886 0.08108 0.07425 0.08082 0.07791
0.07124 0.06774 0.07887 0.06515 0.07429 0.07543 0.05794 0.08872
0.07661 0.03672 0.03870 0.04079 0.04351 0.05667 0.04199 0.04631
0.03874 0.04751 0.04024 0.04580 0.04292 0.05270

0.00768 0.02069 0.01927 0.02230 0.02787 0.02722 0.03016 0.03397
0.06329 0.06942 0.10109 0.09311 0.08201 0.08480 0.08716 0.07587
0.07518 0.09377 0.07358 0.08542 0.08171 0.07048 0.09858 0.09091
0.04495 0.05582 0.05535 0.05777 0.06581 0.04679 0.05012 0.05023
0.05648 0.05112 0.05757 0.05716 0.06304
```

```
0.01635 0.01318 0.01901 0.02985 0.02917 0.02752 0.03268 0.06469
0.07105 0.10482 0.09531 0.08559 0.08842 0.08903 0.07870 0.07635
0.09397 0.07378 0.08523 0.08159 0.07001 0.09875 0.09027 0.04635
0.05368 0.05302 0.05768 0.07083 0.04603 0.04936 0.05190 0.05600
0.04914 0.05603 0.05789 0.06158

0.00717 0.01448 0.02389 0.02117 0.02098 0.02701 0.06003 0.06446
0.10508 0.09050 0.08514 0.08671 0.08463 0.07877 0.07581 0.08956
0.07413 0.08587 0.08000 0.06830 0.09902 0.09027 0.04480 0.04899
0.04911 0.05684 0.06559 0.04421 0.04472 0.04649 0.04938 0.04804
0.05491 0.05236 0.06166

0.01280 0.02098 0.01835 0.01861 0.02878 0.05804 0.06090 0.10446
0.09084 0.08406 0.08693 0.08262 0.07596 0.07248 0.08569 0.07263
0.08241 0.07917 0.06630 0.09579 0.08634 0.04434 0.04937 0.04874
0.05473 0.06261 0.04114 0.04390 0.04550 0.05093 0.04646 0.05059
0.04742 0.05739

0.02308 0.02163 0.02122 0.02753 0.06058 0.06345 0.10876 0.09202
0.08339 0.08715 0.08100 0.07547 0.07444 0.08821 0.07744 0.08627
0.08003 0.06517 0.10244 0.09097 0.04617 0.04961 0.05219 0.05289
0.06536 0.04393 0.04632 0.04631 0.05090 0.04549 0.05318 0.05138
0.05887

0.00606 0.00872 0.03108 0.05505 0.05893 0.10519 0.08836 0.08627
0.08866 0.07935 0.07378 0.07305 0.08532 0.06933 0.08539 0.07964
0.06933 0.10070 0.08825 0.04328 0.05459 0.05112 0.05879 0.05724
0.04549 0.04831 0.04418 0.05460 0.04632 0.05236 0.04759 0.05778

0.00940 0.03167 0.05426 0.05719 0.10644 0.09110 0.08523 0.08809
0.08008 0.07640 0.07508 0.08413 0.07073 0.08484 0.07883 0.06689
0.09739 0.08646 0.04331 0.05392 0.05028 0.05680 0.05729 0.04401
0.04773 0.04611 0.05712 0.04874 0.05179 0.04826 0.05778

0.03107 0.05637 0.05872 0.10994 0.09017 0.08836 0.08951 0.08065
0.07728 0.07598 0.08915 0.07258 0.08849 0.08270 0.06959 0.10388
0.09217 0.04437 0.05461 0.05134 0.06008 0.06020 0.04636 0.04984
0.04348 0.05357 0.04905 0.05338 0.05161 0.06050

0.06816 0.07213 0.10999 0.09970 0.09401 0.09739 0.08569 0.07951
0.07851 0.08773 0.07882 0.08920 0.08468 0.07060 0.10622 0.09360
0.04499 0.05085 0.05826 0.06228 0.07341 0.05190 0.05425 0.05005
0.05902 0.05183 0.05866 0.05485 0.06769

0.02634 0.08747 0.07143 0.06399 0.06832 0.06935 0.06390 0.06162
0.06614 0.06610 0.06843 0.06798 0.05506 0.08974 0.08020 0.04122
0.05796 0.03953 0.05043 0.01540 0.03662 0.04036 0.03229 0.03134
0.03256 0.04056 0.03488 0.02008
```

```
0.09290 0.07168 0.07073 0.07016 0.06915 0.06466 0.06309 0.06802
0.06830 0.07350 0.07105 0.06088 0.09600 0.08297 0.04809 0.06085
0.05401 0.05620 0.03001 0.04586 0.05102 0.05048 0.04730 0.04673
0.05254 0.04758 0.03304

0.06118 0.06698 0.06419 0.08939 0.07803 0.07652 0.09342 0.07511
0.09008 0.08702 0.07639 0.10007 0.09591 0.08743 0.10163 0.09404
0.10275 0.08918 0.08837 0.09055 0.08403 0.09685 0.09034 0.08934
0.08844 0.10128

0.05473 0.05415 0.06196 0.06429 0.06290 0.06945 0.06605 0.07793
0.07915 0.05996 0.08705 0.08357 0.07349 0.08369 0.07536 0.08393
0.07858 0.07831 0.08057 0.07539 0.07667 0.07276 0.07431 0.07676
0.08068

0.01234 0.06798 0.06747 0.06547 0.07767 0.06072 0.07449 0.07083
0.05710 0.08979 0.08000 0.07231 0.08039 0.06738 0.08301 0.06752
0.06099 0.06426 0.05866 0.06544 0.06386 0.06680 0.06986 0.07254

0.07024 0.06892 0.06464 0.07910 0.06385 0.07731 0.07376 0.05842
0.09049 0.08187 0.07376 0.08328 0.07493 0.08411 0.07223 0.06732
0.06856 0.06227 0.07137 0.06802 0.07294 0.07523 0.07662

0.03944 0.04000 0.05466 0.04941 0.06034 0.06129 0.04132 0.08201
0.07676 0.06902 0.08184 0.06544 0.07772 0.07330 0.06261 0.07216
0.06130 0.07631 0.06977 0.06241 0.06704 0.07989

0.01314 0.04898 0.04335 0.05186 0.05138 0.03241 0.07535 0.07061
0.06349 0.07267 0.06564 0.07579 0.06723 0.06066 0.06515 0.05757
0.06790 0.06029 0.06104 0.06409 0.07197

0.04589 0.03930 0.04960 0.04962 0.02817 0.07143 0.06520 0.05941
0.06919 0.05936 0.06966 0.06609 0.05739 0.06017 0.05446 0.06497
0.05518 0.05948 0.06454 0.07001

0.03657 0.04976 0.04962 0.03927 0.07989 0.07955 0.06891 0.07788
0.06994 0.07994 0.07002 0.06571 0.07270 0.06296 0.07656 0.06905
0.06111 0.06662 0.07832

0.03035 0.03431 0.03712 0.06957 0.07176 0.06336 0.07247 0.06236
0.06641 0.07087 0.05915 0.07029 0.06940 0.06738 0.06443 0.06027
0.06185 0.06435

0.00442 0.04268 0.08103 0.07635 0.06677 0.07712 0.06892 0.07571
0.07000 0.06105 0.06631 0.06443 0.07504 0.06791 0.06918 0.06831
0.07865

0.04444 0.08083 0.07975 0.07080 0.07657 0.07099 0.07439 0.07435
0.06296 0.07091 0.06673 0.07202 0.06516 0.06940 0.06691 0.07413
```

```
0.06229 0.05518 0.05525 0.06739 0.05674 0.06502 0.06130 0.05459
0.05507 0.05041 0.06126 0.05046 0.05197 0.05141 0.05761

0.04951 0.08618 0.09801 0.08476 0.09667 0.09146 0.08108 0.09227
0.08349 0.08814 0.08574 0.08763 0.08675 0.09536

0.07193 0.08719 0.07874 0.08185 0.07831 0.07434 0.08041 0.07989
0.08340 0.07883 0.07610 0.07200 0.07932

0.04151 0.02729 0.04250 0.04669 0.02834 0.03679 0.02624 0.03660
0.02736 0.03080 0.03461 0.04580

0.03653 0.04336 0.06392 0.04163 0.04638 0.03916 0.04882 0.03761
0.04509 0.04533 0.05793

0.04640 0.04617 0.02471 0.03183 0.01996 0.02780 0.02464 0.02837
0.03275 0.04199

0.05133 0.03680 0.04000 0.03676 0.04705 0.03807 0.04637 0.04324
0.05377

0.03919 0.04220 0.03548 0.03642 0.03846 0.04515 0.03849 0.02353

0.01627 0.02053 0.02717 0.02214 0.01990 0.02515 0.03497

0.01957 0.02820 0.02262 0.02809 0.03189 0.03826

0.02080 0.01529 0.02043 0.02141 0.03507

0.02338 0.02802 0.02713 0.04115

0.02023 0.01550 0.03639

0.02613 0.04471

0.04005
```

Jukes-Cantor (1969) method. $d = -0.75 \ln(1-4/3p)$
where p = dissimilarity
= #different/#compared

Soy
Pea
Strawberry
Spinach
Chickweed
Saururus
Peperomia
Magnolia
Hedycarya
Illicium
Drimys
Sycamore
Sweetgum
Ranunculus
Parsley
Chloranthus
Sagittaria
Colocasia
Echinodorus
Najas
Potamogeton
Pistia
Zea
Tripsacum
Sugarcane
Sorghum
Rice
Barley
Oats
Wheat
Bamboo
Nymphaea
Cabomba
Welwitschia
Gnetum
E.Tweediana
E.Distachya
Pine
Juniper
Cryptomeria
Cycad

Encephalartos
Zamia
ZamiaO
Ginkgo
Equisetum
Psilotum
Ceratophyllum
Sabal
Nelumbo
Hosta
Nuphar
Liriodendron
Asimina
Calycanthus
Piper
Saruma
Trochodendron
Aristolochia
Barclaya

```
0.01125 0.03525 0.04460 0.05699 0.02916 0.04715 0.03224 0.03466
0.03547 0.04249 0.02868 0.02653 0.03592 0.03751 0.02692 0.03631
0.03878 0.05292 0.06831 0.06821 0.05520 0.06190 0.06277 0.06009
0.05965 0.05948 0.05930 0.06076 0.06064 0.06350 0.04294 0.05898
0.10374 0.08408 0.08008 0.08210 0.07642 0.06867 0.06608 0.07750
0.07005 0.07906 0.07428 0.06611 0.10207 0.09272 0.04124 0.04542
0.03410 0.05297 0.04616 0.03469 0.04252 0.02959 0.03493 0.02998
0.03104 0.02960 0.05018

0.04241 0.05258 0.06483 0.03462 0.05406 0.03466 0.04126 0.04097
0.05237 0.03565 0.03417 0.04075 0.04668 0.03771 0.04428 0.05105
0.06240 0.08010 0.07999 0.06705 0.07654 0.07332 0.07556 0.06869
0.07259 0.07046 0.07501 0.07372 0.07495 0.05003 0.06159 0.11605
0.09112 0.09035 0.09218 0.08311 0.07340 0.07302 0.08491 0.07399
0.08890 0.08000 0.07445 0.11218 0.10361 0.04552 0.05496 0.04020
0.06931 0.05406 0.03696 0.04767 0.03540 0.04060 0.03700 0.03494
0.03595 0.05551

0.03983 0.03833 0.02762 0.04544 0.02952 0.03029 0.03003 0.04619
0.02945 0.02765 0.03376 0.04125 0.02387 0.03993 0.04309 0.05330
0.06122 0.06113 0.05395 0.06117 0.05984 0.05577 0.05628 0.05629
0.05235 0.05252 0.05241 0.06054 0.03605 0.05179 0.10171 0.08054
0.07237 0.07690 0.07174 0.07066 0.06968 0.07623 0.06615 0.07873
0.07067 0.05891 0.09313 0.08032 0.03174 0.05104 0.02847 0.05003
0.04147 0.02924 0.03388 0.02648 0.03365 0.02464 0.03218 0.03173
0.04608
```

```
0.04463 0.04422 0.05374 0.04031 0.04942 0.03720 0.05473 0.04465
0.03906 0.04390 0.04078 0.03768 0.05393 0.05289 0.06361 0.07043
0.07032 0.06130 0.06783 0.06780 0.06158 0.05755 0.05781 0.05982
0.06218 0.05965 0.06581 0.05496 0.06743 0.11848 0.09660 0.08925
0.09152 0.08476 0.07474 0.07441 0.09159 0.08109 0.08815 0.08262
0.06889 0.10968 0.09923 0.04390 0.06144 0.04889 0.06583 0.05407
0.04496 0.04666 0.04513 0.04486 0.04061 0.04556 0.04386 0.05386

0.04103 0.06155 0.04672 0.04386 0.05271 0.06452 0.05064 0.04565
0.05168 0.05485 0.04464 0.06241 0.05726 0.06650 0.07197 0.07187
0.06382 0.07442 0.07206 0.07080 0.06600 0.07008 0.06247 0.06453
0.06714 0.07330 0.05075 0.07027 0.11301 0.09470 0.08767 0.09192
0.08891 0.08143 0.07567 0.09120 0.07237 0.08485 0.08031 0.06902
0.10221 0.09211 0.04756 0.06734 0.04558 0.06291 0.05452 0.05132
0.05058 0.04521 0.05488 0.04685 0.04782 0.04345 0.05823

0.03935 0.03283 0.04205 0.02959 0.04423 0.02292 0.03037 0.02913
0.04381 0.02449 0.03805 0.03625 0.04455 0.06214 0.06214 0.04611
0.05553 0.05958 0.05732 0.05584 0.05861 0.04891 0.05231 0.05526
0.05745 0.03969 0.05163 0.08146 0.08062 0.06652 0.07068 0.06356
0.05972 0.05680 0.06310 0.06593 0.06723 0.06648 0.05327 0.08380
0.06578 0.03406 0.04507 0.03297 0.04443 0.03952 0.03275 0.03534
0.02972 0.02667 0.02414 0.03647 0.02663 0.03894

0.03649 0.04483 0.04008 0.05636 0.04062 0.04437 0.04736 0.04906
0.03765 0.05355 0.05131 0.06199 0.07271 0.07262 0.05952 0.06317
0.06197 0.06188 0.05977 0.06341 0.06356 0.06437 0.06310 0.07174
0.05042 0.05916 0.10408 0.08178 0.07557 0.08110 0.07980 0.07586
0.07344 0.08729 0.07281 0.08396 0.08188 0.06526 0.09083 0.07834
0.04229 0.05780 0.04786 0.05952 0.05457 0.03666 0.04204 0.03806
0.03010 0.03496 0.04654 0.03991 0.05154

0.02330 0.01707 0.04352 0.01426 0.02123 0.02760 0.02712 0.01705
0.03500 0.03447 0.04590 0.06074 0.06065 0.04521 0.05410 0.05296
0.04912 0.04672 0.04905 0.05268 0.05128 0.05224 0.05734 0.03753
0.05082 0.09647 0.08365 0.06554 0.07413 0.06711 0.06448 0.06406
0.07150 0.06814 0.06842 0.07277 0.05514 0.09698 0.08549 0.02892
0.04096 0.02818 0.04835 0.04154 0.00988 0.01417 0.01566 0.02420
0.01774 0.02236 0.03017 0.04070

0.03502 0.03647 0.02033 0.02847 0.02882 0.03613 0.02136 0.03707
0.03159 0.04669 0.04907 0.04907 0.04302 0.05793 0.06131 0.05594
0.05331 0.06036 0.05076 0.04992 0.05756 0.06007 0.04529 0.06150
0.09579 0.08226 0.06871 0.07314 0.07657 0.06988 0.06374 0.07241
0.07323 0.06766 0.07255 0.05751 0.09359 0.08044 0.02942 0.04981
0.02818 0.04231 0.04915 0.02086 0.03241 0.03158 0.03439 0.02510
0.02769 0.02681 0.04693

0.05017 0.01738 0.02049 0.02975 0.03300 0.01906 0.04086 0.03284
```

```
0.04835 0.06416 0.06416 0.04327 0.04827 0.05170 0.04689 0.04412
0.04441 0.04582 0.04677 0.04630 0.05481 0.04107 0.04767 0.09238
0.08306 0.05849 0.06880 0.06609 0.06076 0.05854 0.07109 0.06730
0.06953 0.07012 0.05339 0.09725 0.07726 0.03082 0.04568 0.02965
0.04265 0.04028 0.02389 0.02206 0.02021 0.02936 0.02096 0.02249
0.03390 0.03656

0.04299 0.03925 0.04671 0.05237 0.03260 0.05219 0.05281 0.06856
0.06869 0.06860 0.06198 0.08212 0.07785 0.07842 0.07417 0.07618
0.07176 0.07122 0.07458 0.07732 0.05009 0.06142 0.11395 0.09191
0.09566 0.09255 0.09155 0.08781 0.09217 0.09533 0.07583 0.09609
0.09024 0.07902 0.11548 0.10091 0.04444 0.06592 0.04482 0.06627
0.05414 0.03577 0.04670 0.03605 0.04342 0.03745 0.03808 0.03758
0.05742

0.01914 0.02937 0.03150 0.01739 0.03951 0.03107 0.04902 0.05858
0.05853 0.04208 0.06090 0.05719 0.05353 0.05022 0.05390 0.05675
0.05759 0.05511 0.05749 0.03422 0.04753 0.09932 0.07858 0.07018
0.07769 0.07659 0.06516 0.06464 0.07588 0.06221 0.07032 0.06467
0.05870 0.10205 0.08726 0.02918 0.04905 0.02261 0.04665 0.03914
0.01564 0.02164 0.01152 0.02545 0.01886 0.01722 0.02322 0.04120

0.02998 0.03228 0.01722 0.03551 0.03650 0.05171 0.06305 0.06296
0.05033 0.05246 0.05062 0.05185 0.04865 0.04977 0.05357 0.05289
0.05361 0.06108 0.03911 0.05069 0.10246 0.08438 0.07389 0.07671
0.07624 0.06931 0.06690 0.07662 0.06379 0.07438 0.06759 0.05792
0.09624 0.08387 0.03307 0.04183 0.02903 0.04226 0.04264 0.02113
0.02549 0.02168 0.03076 0.01829 0.02556 0.02764 0.04182

0.03900 0.02311 0.04497 0.04189 0.05851 0.06530 0.06521 0.05689
0.06535 0.06373 0.05999 0.05434 0.05977 0.05795 0.06069 0.05961
0.06324 0.04322 0.05848 0.10693 0.08548 0.08110 0.08693 0.07745
0.07420 0.07408 0.08421 0.07022 0.08347 0.07494 0.06649 0.10801
0.09387 0.03768 0.04981 0.03031 0.05604 0.04459 0.02505 0.03351
0.02308 0.03418 0.02659 0.02643 0.02732 0.05044

0.03126 0.04460 0.04768 0.06313 0.07592 0.07581 0.05690 0.06205
0.06312 0.06117 0.05553 0.05824 0.05800 0.05792 0.06059 0.06766
0.05082 0.05949 0.11430 0.08869 0.08292 0.08644 0.08444 0.07799
0.07463 0.08767 0.07949 0.08734 0.08123 0.06781 0.11419 0.09447
0.04036 0.05817 0.04403 0.06121 0.05580 0.03467 0.03893 0.03146
0.04348 0.03579 0.03116 0.03796 0.05171

0.03470 0.03232 0.04774 0.05480 0.05480 0.04154 0.05247 0.05306
0.04956 0.04643 0.04823 0.04854 0.04871 0.05148 0.05522 0.03527
0.04956 0.09971 0.07683 0.06934 0.07417 0.07506 0.06605 0.06369
0.07335 0.06145 0.06921 0.06746 0.05406 0.09210 0.07908 0.02842
0.04024 0.02349 0.04085 0.03848 0.01513 0.02234 0.01304 0.02596
0.01491 0.01781 0.02191 0.03686
```

```
0.04054 0.04479 0.06553 0.06544 0.05189 0.06489 0.06370 0.06283
0.06047 0.06462 0.06401 0.06539 0.06404 0.06906 0.04764 0.05985
0.09996 0.08492 0.08524 0.09268 0.08660 0.07952 0.07420 0.08475
0.06982 0.08566 0.07821 0.06563 0.10424 0.09698 0.04301 0.04883
0.04183 0.05784 0.04987 0.03536 0.03601 0.03365 0.03826 0.03480
0.03672 0.03973 0.05231

0.05113 0.06421 0.06412 0.01914 0.05234 0.05200 0.05171 0.04689
0.04833 0.05349 0.05231 0.05503 0.05123 0.04791 0.05936 0.09979
0.08055 0.07677 0.08263 0.07657 0.07023 0.06708 0.07403 0.06102
0.07200 0.06926 0.05796 0.09306 0.08187 0.03362 0.03173 0.03309
0.04652 0.05106 0.03143 0.03810 0.02881 0.03630 0.02813 0.03570
0.02913 0.05076

0.05675 0.05597 0.05549 0.07320 0.07566 0.07063 0.06943 0.07189
0.07046 0.07115 0.07291 0.07611 0.05224 0.07161 0.11489 0.09253
0.09268 0.09927 0.08789 0.08423 0.07865 0.09113 0.07104 0.08158
0.08064 0.07378 0.10553 0.09952 0.04745 0.05998 0.04658 0.06443
0.05428 0.04642 0.04986 0.04433 0.05342 0.04256 0.05178 0.04859
0.06368

0.00062 0.06399 0.08691 0.09298 0.08802 0.08546 0.09246 0.07791
0.08219 0.08529 0.09249 0.07056 0.07611 0.12070 0.09063 0.10217
0.10310 0.09253 0.09255 0.08803 0.09968 0.07722 0.09749 0.08877
0.08324 0.11266 0.10627 0.05351 0.07876 0.05955 0.07900 0.07231
0.05390 0.06141 0.05115 0.05719 0.05048 0.06249 0.05430 0.07829

0.06390 0.08753 0.09284 0.08789 0.08533 0.09232 0.07781 0.08208
0.08517 0.09236 0.07056 0.07611 0.12070 0.09049 0.10203 0.10303
0.09239 0.09242 0.08791 0.09954 0.07722 0.09736 0.08862 0.08311
0.11250 0.10687 0.05346 0.07865 0.05946 0.07889 0.07231 0.05390
0.06131 0.05105 0.05709 0.05040 0.06241 0.05423 0.07818

0.05261 0.05653 0.05155 0.04949 0.04985 0.05529 0.05397 0.05755
0.05388 0.05673 0.06103 0.10601 0.08581 0.07819 0.08552 0.08226
0.07485 0.07100 0.08333 0.06815 0.07824 0.07950 0.06030 0.09442
0.08081 0.03765 0.03973 0.04194 0.04482 0.05892 0.04322 0.04780
0.03978 0.04908 0.04136 0.04725 0.04419 0.05465

0.00772 0.02098 0.01952 0.02264 0.02840 0.02773 0.03078 0.03476
0.06612 0.07285 0.10859 0.09941 0.08685 0.08999 0.09266 0.07998
0.07922 0.10017 0.07744 0.09069 0.08651 0.07402 0.10569 0.09691
0.04636 0.05800 0.05750 0.06012 0.06888 0.04831 0.05187 0.05199
0.05872 0.05295 0.05990 0.05946 0.06585

0.01653 0.01329 0.01925 0.03046 0.02975 0.02804 0.03341 0.06765
0.07464 0.11291 0.10193 0.09089 0.09408 0.09478 0.08314 0.08052
0.10040 0.07766 0.09048 0.08637 0.07350 0.10588 0.09619 0.04784
```

```
0.05570 0.05499 0.06002 0.07440 0.04750 0.05106 0.05378 0.05820
0.05083 0.05824 0.06024 0.06426

0.00721 0.01462 0.02428 0.02148 0.02128 0.02751 0.06257 0.06740
0.11321 0.09645 0.09037 0.09214 0.08979 0.08322 0.07992 0.09538
0.07806 0.09120 0.08460 0.07161 0.10619 0.09618 0.04619 0.05066
0.05079 0.05911 0.06864 0.04557 0.04610 0.04799 0.05108 0.04965
0.05703 0.05428 0.06434

0.01291 0.02128 0.01858 0.01884 0.02934 0.06040 0.06352 0.11249
0.09683 0.08916 0.09239 0.08754 0.08009 0.07623 0.09100 0.07639
0.08730 0.08367 0.06941 0.10248 0.09172 0.04570 0.05107 0.05040
0.05683 0.06538 0.04232 0.04524 0.04694 0.05274 0.04796 0.05238
0.04899 0.05971

0.02344 0.02195 0.02152 0.02805 0.06317 0.06630 0.11750 0.09817
0.08840 0.09264 0.08572 0.07954 0.07840 0.09385 0.08174 0.09164
0.08463 0.06818 0.11014 0.09698 0.04765 0.05133 0.05409 0.05485
0.06838 0.04526 0.04781 0.04781 0.05271 0.04693 0.05516 0.05322
0.06130

0.00608 0.00878 0.03174 0.05717 0.06137 0.11334 0.09401 0.09165
0.09436 0.08387 0.07766 0.07685 0.09058 0.07275 0.09065 0.08419
0.07275 0.10814 0.09389 0.04458 0.05668 0.05295 0.06122 0.05954
0.04693 0.04994 0.04554 0.05669 0.04781 0.05428 0.04917 0.06012

0.00946 0.03236 0.05632 0.05949 0.11479 0.09713 0.09047 0.09370
0.08469 0.08057 0.07911 0.08924 0.07429 0.09003 0.08329 0.07006
0.10432 0.09186 0.04461 0.05595 0.05204 0.05906 0.05959 0.04536
0.04932 0.04759 0.05941 0.05040 0.05367 0.04989 0.06012

0.03174 0.05861 0.06115 0.11889 0.09607 0.09401 0.09532 0.08532
0.08156 0.08011 0.09491 0.07634 0.09416 0.08763 0.07303 0.11181
0.09834 0.04574 0.05670 0.05318 0.06263 0.06275 0.04786 0.05157
0.04479 0.05558 0.05073 0.05538 0.05347 0.06308

0.07146 0.07584 0.11895 0.10698 0.10045 0.10432 0.09100 0.08404
0.08293 0.09330 0.08328 0.09497 0.08985 0.07414 0.11454 0.09997
0.04639 0.05265 0.06065 0.06502 0.07726 0.05378 0.05631 0.05180
0.06147 0.05371 0.06108 0.05696 0.07094

0.02681 0.09300 0.07506 0.06689 0.07163 0.07277 0.06679 0.06429
0.06924 0.06920 0.07175 0.07126 0.05719 0.09558 0.08483 0.04239
0.06032 0.04061 0.05220 0.01556 0.03755 0.04149 0.03301 0.03201
0.03329 0.04170 0.03572 0.02036

0.09918 0.07534 0.07430 0.07366 0.07255 0.06761 0.06591 0.07131
0.07161 0.07736 0.07464 0.06349 0.10272 0.08792 0.04971 0.06346
0.05605 0.05842 0.03063 0.04732 0.05284 0.05226 0.04885 0.04825
```

```
0.05447 0.04915 0.03379

0.06382 0.07016 0.06710 0.09518 0.08240 0.08071 0.09977 0.07915
0.09597 0.09250 0.08057 0.10740 0.10262 0.09296 0.10920 0.10048
0.11051 0.09494 0.09402 0.09650 0.08913 0.10370 0.09626 0.09512
0.09410 0.10881

0.05683 0.05621 0.06467 0.06721 0.06569 0.07288 0.06914 0.08228
0.08364 0.06249 0.09253 0.08861 0.07735 0.08874 0.07942 0.08901
0.08301 0.08270 0.08523 0.07945 0.08087 0.07653 0.07825 0.08098
0.08536

0.01244 0.07126 0.07069 0.06850 0.08199 0.06332 0.07845 0.07440
0.05939 0.09563 0.08460 0.07603 0.08504 0.07061 0.08797 0.07076
0.06361 0.06718 0.06108 0.06848 0.06675 0.06996 0.07333 0.07629

0.07375 0.07230 0.06760 0.08359 0.06673 0.08159 0.07764 0.06082
0.09643 0.08669 0.07764 0.08827 0.07894 0.08921 0.07595 0.07054
0.07190 0.06501 0.07500 0.07131 0.07674 0.07928 0.08082

0.04051 0.04111 0.05675 0.05111 0.06291 0.06393 0.04250 0.08685
0.08098 0.07240 0.08666 0.06847 0.08204 0.07713 0.06538 0.07587
0.06395 0.08048 0.07323 0.06516 0.07023 0.08447

0.01326 0.05065 0.04465 0.05374 0.05322 0.03313 0.07941 0.07416
0.06634 0.07643 0.06869 0.07990 0.07043 0.06325 0.06816 0.05990
0.07117 0.06285 0.06367 0.06700 0.07567

0.04735 0.04037 0.05131 0.05134 0.02871 0.07506 0.06821 0.06189
0.07259 0.06184 0.07311 0.06919 0.05971 0.06272 0.05654 0.06795
0.05731 0.06198 0.06748 0.07349

0.03749 0.05149 0.05134 0.04034 0.08447 0.08410 0.07229 0.08222
0.07342 0.08453 0.07351 0.06876 0.07647 0.06576 0.08075 0.07243
0.06374 0.06977 0.08272

0.03098 0.03512 0.03808 0.07301 0.07542 0.06620 0.07622 0.06510
0.06954 0.07444 0.06162 0.07381 0.07283 0.07060 0.06737 0.06283
0.06455 0.06727

0.00444 0.04394 0.08575 0.08052 0.06993 0.08138 0.07229 0.07981
0.07349 0.06368 0.06943 0.06737 0.07906 0.07118 0.07258 0.07162
0.08309

0.04582 0.08553 0.08432 0.07436 0.08077 0.07458 0.07834 0.07830
0.06576 0.07449 0.06989 0.07572 0.06817 0.07283 0.07009 0.07805

0.06503 0.05731 0.05739 0.07061 0.05900 0.06802 0.06395 0.05668
0.05719 0.05218 0.06390 0.05224 0.05385 0.05326 0.05995
```

```
0.05122 0.09155 0.10504 0.08994 0.10349 0.09753 0.08581 0.09846
0.08851 0.09377 0.09105 0.09318 0.09219 0.10199

0.07561 0.09268 0.08319 0.08667 0.08271 0.07828 0.08505 0.08447
0.08841 0.08328 0.08025 0.07569 0.08384

0.04270 0.02780 0.04376 0.04820 0.02889 0.03773 0.02671 0.03752
0.02787 0.03145 0.03544 0.04726

0.03745 0.04467 0.06681 0.04283 0.04788 0.04022 0.05048 0.03858
0.04650 0.04676 0.06029

0.04789 0.04765 0.02513 0.03253 0.02023 0.02833 0.02506 0.02892
0.03349 0.04321

0.05317 0.03774 0.04111 0.03770 0.04859 0.03907 0.04787 0.04454
0.05580

0.04026 0.04344 0.03635 0.03733 0.03948 0.04657 0.03952 0.02391

0.01645 0.02081 0.02768 0.02248 0.02016 0.02558 0.03581

0.01983 0.02875 0.02297 0.02863 0.03259 0.03927

0.02109 0.01545 0.02071 0.02172 0.03592

0.02375 0.02856 0.02764 0.04232

0.02051 0.01567 0.03730

0.02660 0.04610

0.04116
```

Distance matrix calculated by Kimura's (1980) two-parameter formula.

$d = -0.5 \ln[(1-2P-Q)(1-2Q)^{0.5}]$
P = #transitions/#compared
Q = #transversions/#compared

Soy
Pea
Strawberry
Spinach
Chickweed
Saururus
Peperomia
Magnolia
Hedycarya
Illicium
Drimys
Sycamore
Sweetgum
Ranunculus
Parsley
Chloranthus
Sagittaria
Colocasia
Echinodorus
Najas
Potamogeton
Pistia
Zea
Tripsacum
Sugarcane
Sorghum
Rice
Barley
Oats
Wheat
Bamboo
Nymphaea
Cabomba
Welwitschia
Gnetum
E.Tweediana
E.Distachya
Pine
Juniper

Cryptomeria
Cycad
Encephalartos
Zamia
ZamiaO
Ginkgo
Equisetum
Psilotum
Ceratophyllum
Sabal
Nelumbo
Hosta
Nuphar
Liriodendron
Asimina
Calycanthus
Piper
Saruma
Trochodendron
Aristolochia
Barclaya

```
0.01127 0.03536 0.04476 0.05728 0.02921 0.04732 0.03236 0.03474
0.03561 0.04267 0.02873 0.02661 0.03601 0.03763 0.02699 0.03641
0.03899 0.05314 0.06862 0.06853 0.05546 0.06233 0.06323 0.06047
0.06005 0.05986 0.05970 0.06115 0.06102 0.06382 0.04314 0.05928
0.10482 0.08475 0.08069 0.08274 0.07708 0.06930 0.06665 0.07829
0.07053 0.07968 0.07485 0.06659 0.10321 0.09335 0.04135 0.04553
0.03420 0.05315 0.04626 0.03480 0.04269 0.02965 0.03501 0.03008
0.03114 0.02966 0.05043

0.04254 0.05280 0.06522 0.03471 0.05428 0.03483 0.04139 0.04116
0.05269 0.03572 0.03430 0.04090 0.04686 0.03784 0.04441 0.05141
0.06273 0.08057 0.08045 0.06744 0.07723 0.07413 0.07624 0.06937
0.07323 0.07103 0.07564 0.07431 0.07551 0.05035 0.06195 0.11728
0.09187 0.09114 0.09301 0.08388 0.07412 0.07368 0.08577 0.07453
0.08973 0.08068 0.07507 0.11360 0.10438 0.04566 0.05513 0.04037
0.06956 0.05423 0.03711 0.04797 0.03552 0.04070 0.03713 0.03506
0.03605 0.05576

0.03994 0.03848 0.02765 0.04559 0.02957 0.03036 0.03013 0.04644
0.02948 0.02772 0.03384 0.04146 0.02392 0.04007 0.04329 0.05347
0.06145 0.06137 0.05419 0.06159 0.06025 0.05608 0.05660 0.05659
0.05261 0.05274 0.05264 0.06089 0.03619 0.05207 0.10253 0.08108
0.07278 0.07732 0.07221 0.07131 0.07026 0.07700 0.06659 0.07925
0.07118 0.05931 0.09396 0.08088 0.03179 0.05116 0.02852 0.05014
0.04155 0.02931 0.03398 0.02653 0.03376 0.02472 0.03226 0.03181
```

```
0.04623

0.04475 0.04428 0.05389 0.04039 0.04954 0.03725 0.05504 0.04472
0.03916 0.04400 0.04099 0.03778 0.05416 0.05315 0.06384 0.07071
0.07061 0.06151 0.06814 0.06815 0.06184 0.05778 0.05805 0.06002
0.06238 0.05987 0.06600 0.05517 0.06773 0.11953 0.09736 0.08979
0.09211 0.08544 0.07549 0.07511 0.09247 0.08163 0.08886 0.08323
0.06937 0.11061 0.09989 0.04395 0.06160 0.04903 0.06604 0.05412
0.04506 0.04675 0.04521 0.04496 0.04073 0.04573 0.04397 0.05401

0.04114 0.06188 0.04687 0.04401 0.05291 0.06511 0.05081 0.04583
0.05190 0.05526 0.04485 0.06280 0.05769 0.06680 0.07234 0.07224
0.06416 0.07495 0.07255 0.07119 0.06639 0.07052 0.06275 0.06481
0.06745 0.07364 0.05101 0.07076 0.11384 0.09537 0.08826 0.09252
0.08961 0.08224 0.07629 0.09218 0.07277 0.08542 0.08088 0.06949
0.10306 0.09278 0.04766 0.06764 0.04572 0.06319 0.05468 0.05146
0.05073 0.04533 0.05521 0.04708 0.04803 0.04361 0.05852

0.03947 0.03286 0.04210 0.02961 0.04439 0.02293 0.03041 0.02914
0.04396 0.02454 0.03815 0.03637 0.04466 0.06231 0.06231 0.04621
0.05580 0.05991 0.05757 0.05612 0.05892 0.04905 0.05246 0.05550
0.05762 0.03984 0.05178 0.08209 0.08123 0.06681 0.07100 0.06386
0.06010 0.05715 0.06356 0.06637 0.06760 0.06710 0.05353 0.08440
0.06607 0.03419 0.04517 0.03305 0.04447 0.03957 0.03278 0.03536
0.02973 0.02673 0.02416 0.03653 0.02665 0.03898

0.03654 0.04490 0.04014 0.05674 0.04071 0.04452 0.04751 0.04925
0.03776 0.05379 0.05153 0.06222 0.07308 0.07298 0.05977 0.06354
0.06237 0.06221 0.06015 0.06373 0.06388 0.06469 0.06342 0.07205
0.05062 0.05951 0.10495 0.08232 0.07597 0.08158 0.08058 0.07666
0.07421 0.08840 0.07339 0.08470 0.08272 0.06574 0.09137 0.07882
0.04235 0.05796 0.04809 0.05971 0.05469 0.03671 0.04215 0.03817
0.03017 0.03508 0.04670 0.04004 0.05170

0.02330 0.01708 0.04375 0.01426 0.02126 0.02764 0.02718 0.01706
0.03511 0.03461 0.04599 0.06098 0.06089 0.04534 0.05436 0.05331
0.04932 0.04695 0.04929 0.05293 0.05147 0.05252 0.05754 0.03765
0.05100 0.09717 0.08427 0.06584 0.07455 0.06749 0.06495 0.06454
0.07201 0.06854 0.06877 0.07322 0.05547 0.09793 0.08603 0.02895
0.04101 0.02825 0.04839 0.04160 0.00988 0.01419 0.01567 0.02422
0.01777 0.02239 0.03024 0.04081

0.03506 0.03659 0.02033 0.02850 0.02884 0.03618 0.02138 0.03715
0.03167 0.04673 0.04918 0.04918 0.04308 0.05811 0.06155 0.05604
0.05344 0.06057 0.05090 0.05001 0.05770 0.06019 0.04539 0.06175
0.09655 0.08274 0.06894 0.07349 0.07693 0.07032 0.06409 0.07282
0.07350 0.06789 0.07287 0.05773 0.09417 0.08083 0.02947 0.04985
0.02821 0.04295 0.04922 0.02086 0.03242 0.03161 0.03445 0.02513
0.02773 0.02686 0.04709
```

```
0.05052 0.01741 0.02051 0.02983 0.03306 0.01911 0.04101 0.03297
0.04846 0.06440 0.06440 0.04340 0.04847 0.05193 0.04701 0.04426
0.04453 0.04597 0.04689 0.04644 0.05494 0.04117 0.04779 0.09319
0.08369 0.05873 0.06915 0.06637 0.06108 0.05888 0.07152 0.06757
0.06986 0.07048 0.05367 0.09798 0.07758 0.03087 0.04576 0.02973
0.04270 0.04030 0.02391 0.02208 0.02022 0.02940 0.02099 0.02255
0.03403 0.03666

0.04318 0.03943 0.04698 0.05271 0.03273 0.05247 0.05323 0.06897
0.06927 0.06918 0.06244 0.08302 0.07874 0.07921 0.07496 0.07702
0.07239 0.07177 0.07523 0.07799 0.05039 0.06181 0.11542 0.09289
0.09666 0.09346 0.09269 0.08903 0.09348 0.09670 0.07657 0.09734
0.09163 0.08001 0.11685 0.10202 0.04463 0.06625 0.04509 0.06657
0.05430 0.03593 0.04697 0.03622 0.04374 0.03770 0.03830 0.03781
0.05779

0.01915 0.02940 0.03155 0.01741 0.03959 0.03117 0.04915 0.05885
0.05880 0.04228 0.06118 0.05752 0.05371 0.05042 0.05408 0.05698
0.05779 0.05530 0.05767 0.03430 0.04771 0.09992 0.07908 0.07056
0.07815 0.07709 0.06556 0.06499 0.07639 0.06253 0.07081 0.06511
0.05904 0.10303 0.08790 0.02921 0.04911 0.02266 0.04671 0.03917
0.01564 0.02165 0.01152 0.02550 0.01889 0.01724 0.02324 0.04131

0.03006 0.03236 0.01724 0.03561 0.03665 0.05185 0.06332 0.06323
0.05052 0.05277 0.05095 0.05214 0.04894 0.05005 0.05389 0.05314
0.05391 0.06137 0.03927 0.05089 0.10323 0.08495 0.07424 0.07705
0.07679 0.06990 0.06740 0.07724 0.06411 0.07483 0.06798 0.05821
0.09702 0.08429 0.03314 0.04191 0.02910 0.04234 0.04272 0.02115
0.02552 0.02171 0.03086 0.01833 0.02563 0.02770 0.04193

0.03916 0.02316 0.04512 0.04206 0.05868 0.06563 0.06554 0.05714
0.06589 0.06435 0.06033 0.05473 0.06023 0.05832 0.06108 0.05996
0.06364 0.04332 0.05867 0.10758 0.08601 0.08171 0.08753 0.07805
0.07486 0.07470 0.08506 0.07062 0.08411 0.07553 0.06698 0.10902
0.09469 0.03774 0.04992 0.03037 0.05618 0.04462 0.02507 0.03357
0.02309 0.03426 0.02667 0.02647 0.02738 0.05064

0.03137 0.04480 0.04799 0.06341 0.07637 0.07626 0.05713 0.06249
0.06361 0.06151 0.05583 0.05864 0.05836 0.05824 0.06091 0.06797
0.05103 0.05969 0.11551 0.08934 0.08356 0.08709 0.08522 0.07887
0.07545 0.08861 0.08011 0.08816 0.08187 0.06839 0.11572 0.09525
0.04042 0.05840 0.04422 0.06147 0.05593 0.03472 0.03905 0.03153
0.04367 0.03596 0.03123 0.03809 0.05197

0.03482 0.03245 0.04785 0.05501 0.05501 0.04175 0.05277 0.05341
0.04977 0.04665 0.04843 0.04877 0.04890 0.05171 0.05539 0.03535
0.04973 0.10052 0.07736 0.06971 0.07462 0.07565 0.06665 0.06421
0.07394 0.06184 0.06966 0.06793 0.05440 0.09282 0.07959 0.02846
```

```
0.04029 0.02353 0.04091 0.03851 0.01514 0.02238 0.01305 0.02604
0.01495 0.01786 0.02196 0.03695

0.04076 0.04485 0.06582 0.06573 0.05208 0.06519 0.06402 0.06315
0.06076 0.06494 0.06436 0.06571 0.06437 0.06938 0.04788 0.06013
0.10053 0.08549 0.08571 0.09321 0.08750 0.08033 0.07486 0.08568
0.07026 0.08631 0.07885 0.06606 0.10529 0.09777 0.04312 0.04890
0.04202 0.05795 0.05003 0.03551 0.03612 0.03378 0.03842 0.03491
0.03682 0.03988 0.05254

0.05136 0.06468 0.06459 0.01915 0.05264 0.05229 0.05196 0.04714
0.04852 0.05382 0.05257 0.05534 0.05141 0.04825 0.05975 0.10061
0.08115 0.07716 0.08309 0.07715 0.07077 0.06758 0.07465 0.06134
0.07248 0.06985 0.05832 0.09394 0.08238 0.03373 0.03177 0.03326
0.04661 0.05128 0.03153 0.03823 0.02893 0.03646 0.02822 0.03591
0.02924 0.05107

0.05691 0.05613 0.05573 0.07356 0.07612 0.07099 0.06975 0.07223
0.07079 0.07145 0.07329 0.07644 0.05246 0.07204 0.11577 0.09321
0.09314 0.09989 0.08856 0.08499 0.07924 0.09207 0.07146 0.08218
0.08129 0.07424 0.10649 0.10020 0.04751 0.06011 0.04673 0.06457
0.05441 0.04650 0.04993 0.04443 0.05355 0.04264 0.05192 0.04870
0.06395

0.00062 0.06443 0.08752 0.09379 0.08861 0.08607 0.09318 0.07829
0.08262 0.08577 0.09310 0.07087 0.07650 0.12175 0.09120 0.10289
0.10380 0.09329 0.09343 0.08877 0.10062 0.07768 0.09833 0.08952
0.08390 0.11361 0.10710 0.05363 0.07914 0.05978 0.07932 0.07248
0.05408 0.06164 0.05139 0.05747 0.05072 0.06284 0.05458 0.07873

0.06433 0.08813 0.09365 0.08848 0.08594 0.09303 0.07818 0.08251
0.08564 0.09296 0.07087 0.07650 0.12175 0.09106 0.10275 0.10373
0.09316 0.09330 0.08865 0.10047 0.07768 0.09820 0.08937 0.08378
0.11344 0.10773 0.05358 0.07903 0.05969 0.07921 0.07248 0.05408
0.06154 0.05129 0.05737 0.05064 0.06275 0.05451 0.07862

0.05272 0.05670 0.05165 0.04961 0.04993 0.05549 0.05412 0.05772
0.05396 0.05714 0.06141 0.10683 0.08645 0.07849 0.08598 0.08283
0.07538 0.07144 0.08405 0.06856 0.07884 0.08023 0.06070 0.09521
0.08128 0.03770 0.03980 0.04206 0.04490 0.05922 0.04331 0.04792
0.03988 0.04930 0.04151 0.04753 0.04436 0.05492

0.00772 0.02100 0.01956 0.02266 0.02853 0.02785 0.03088 0.03480
0.06661 0.07327 0.10958 0.10045 0.08759 0.09080 0.09369 0.08078
0.07998 0.10141 0.07809 0.09157 0.08741 0.07466 0.10668 0.09763
0.04644 0.05818 0.05778 0.06030 0.06923 0.04854 0.05207 0.05229
0.05908 0.05327 0.06035 0.05991 0.06623

0.01655 0.01333 0.01928 0.03061 0.02990 0.02814 0.03347 0.06815
```

```
0.07506 0.11389 0.10289 0.09172 0.09498 0.09585 0.08404 0.08136
0.10170 0.07835 0.09140 0.08722 0.07408 0.10691 0.09689 0.04795
0.05585 0.05533 0.06021 0.07484 0.04775 0.05131 0.05415 0.05857
0.05112 0.05872 0.06078 0.06461

0.00721 0.01463 0.02434 0.02151 0.02132 0.02756 0.06290 0.06763
0.11409 0.09718 0.09104 0.09288 0.09057 0.08397 0.08066 0.09636
0.07860 0.09213 0.08532 0.07209 0.10716 0.09684 0.04626 0.05076
0.05096 0.05925 0.06886 0.04570 0.04622 0.04818 0.05125 0.04986
0.05737 0.05459 0.06464

0.01292 0.02134 0.01862 0.01888 0.02938 0.06078 0.06376 0.11337
0.09765 0.08984 0.09319 0.08838 0.08084 0.07691 0.09198 0.07696
0.08819 0.08448 0.06990 0.10334 0.09234 0.04580 0.05119 0.05065
0.05699 0.06564 0.04248 0.04540 0.04718 0.05301 0.04822 0.05276
0.04931 0.05999

0.02350 0.02198 0.02155 0.02807 0.06353 0.06653 0.11837 0.09904
0.08905 0.09339 0.08649 0.08015 0.07900 0.09469 0.08230 0.09248
0.08532 0.06862 0.11118 0.09759 0.04774 0.05142 0.05432 0.05497
0.06865 0.04539 0.04795 0.04799 0.05293 0.04712 0.05552 0.05354
0.06156

0.00609 0.00878 0.03179 0.05749 0.06159 0.11440 0.09468 0.09232
0.09501 0.08450 0.07823 0.07738 0.09144 0.07315 0.09139 0.08489
0.07325 0.10903 0.09449 0.04467 0.05686 0.05313 0.06143 0.05973
0.04713 0.05014 0.04573 0.05701 0.04804 0.05461 0.04941 0.06039

0.00946 0.03240 0.05657 0.05964 0.11600 0.09790 0.09110 0.09433
0.08533 0.08123 0.07970 0.09004 0.07474 0.09075 0.08396 0.07049
0.10508 0.09240 0.04468 0.05609 0.05221 0.05921 0.05974 0.04549
0.04947 0.04777 0.05974 0.05063 0.05394 0.05010 0.06036

0.03180 0.05888 0.06133 0.11996 0.09679 0.09475 0.09611 0.08590
0.08225 0.08084 0.09578 0.07687 0.09519 0.08838 0.07352 0.11286
0.09903 0.04583 0.05686 0.05339 0.06279 0.06292 0.04804 0.05179
0.04500 0.05581 0.05097 0.05572 0.05377 0.06335

0.07183 0.07607 0.11982 0.10778 0.10109 0.10502 0.09153 0.08453
0.08344 0.09400 0.08384 0.09582 0.09067 0.07456 0.11567 0.10051
0.04648 0.05274 0.06085 0.06511 0.07751 0.05393 0.05647 0.05189
0.06166 0.05386 0.06139 0.05719 0.07121

0.02690 0.09374 0.07562 0.06726 0.07208 0.07328 0.06731 0.06482
0.06988 0.06966 0.07227 0.07177 0.05753 0.09658 0.08542 0.04268
0.06060 0.04077 0.05241 0.01557 0.03765 0.04160 0.03311 0.03210
0.03339 0.04190 0.03583 0.02042

0.10006 0.07594 0.07475 0.07417 0.07295 0.06806 0.06637 0.07195
```

```
0.07219 0.07809 0.07521 0.06384 0.10382 0.08856 0.05002 0.06368
0.05635 0.05867 0.03070 0.04744 0.05298 0.05249 0.04910 0.04844
0.05477 0.04935 0.03389

0.06409 0.07054 0.06750 0.09586 0.08298 0.08123 0.10046 0.07962
0.09660 0.09318 0.08108 0.10807 0.10336 0.09365 0.10989 0.10125
0.11106 0.09545 0.09478 0.09723 0.08972 0.10443 0.09700 0.09581
0.09472 0.10956

0.05702 0.05651 0.06506 0.06767 0.06612 0.07336 0.06949 0.08296
0.08442 0.06295 0.09308 0.08903 0.07798 0.08922 0.07992 0.08953
0.08343 0.08337 0.08590 0.08033 0.08158 0.07734 0.07883 0.08145
0.08580

0.01245 0.07173 0.07116 0.06900 0.08259 0.06366 0.07901 0.07500
0.05973 0.09618 0.08511 0.07642 0.08542 0.07098 0.08840 0.07098
0.06385 0.06752 0.06131 0.06879 0.06713 0.07030 0.07371 0.07663

0.07432 0.07285 0.06812 0.08431 0.06724 0.08230 0.07836 0.06123
0.09703 0.08723 0.07811 0.08874 0.07947 0.08967 0.07624 0.07086
0.07233 0.06530 0.07540 0.07177 0.07716 0.07970 0.08115

0.04070 0.04133 0.05711 0.05130 0.06332 0.06434 0.04265 0.08744
0.08156 0.07271 0.08731 0.06891 0.08262 0.07744 0.06577 0.07628
0.06436 0.08111 0.07389 0.06553 0.07069 0.08505

0.01329 0.05096 0.04481 0.05400 0.05352 0.03324 0.07999 0.07468
0.06674 0.07690 0.06937 0.08032 0.07081 0.06365 0.06862 0.06031
0.07178 0.06337 0.06406 0.06753 0.07618

0.04766 0.04052 0.05161 0.05167 0.02882 0.07560 0.06862 0.06223
0.07303 0.06238 0.07342 0.06958 0.06008 0.06313 0.05696 0.06853
0.05777 0.06237 0.06804 0.07402

0.03755 0.05170 0.05153 0.04049 0.08512 0.08496 0.07277 0.08277
0.07402 0.08512 0.07396 0.06928 0.07707 0.06629 0.08168 0.07303
0.06414 0.07029 0.08339

0.03108 0.03528 0.03817 0.07344 0.07602 0.06662 0.07660 0.06549
0.06982 0.07478 0.06187 0.07414 0.07329 0.07113 0.06781 0.06315
0.06491 0.06758

0.00444 0.04412 0.08639 0.08122 0.07042 0.08189 0.07289 0.08023
0.07382 0.06390 0.06972 0.06775 0.07969 0.07165 0.07305 0.07204
0.08361

0.04605 0.08624 0.08508 0.07491 0.08135 0.07526 0.07884 0.07869
0.06602 0.07485 0.07036 0.07631 0.06862 0.07347 0.07062 0.07856
```

```
0.06536 0.05758 0.05772 0.07092 0.05941 0.06833 0.06421 0.05694
0.05751 0.05246 0.06435 0.05252 0.05424 0.05360 0.06022

0.05141 0.09224 0.10589 0.09075 0.10424 0.09829 0.08658 0.09919
0.08928 0.09452 0.09176 0.09397 0.09287 0.10280

0.07585 0.09313 0.08378 0.08700 0.08303 0.07869 0.08548 0.08496
0.08913 0.08385 0.08080 0.07616 0.08427

0.04275 0.02785 0.04381 0.04840 0.02892 0.03778 0.02676 0.03760
0.02792 0.03155 0.03555 0.04739

0.03751 0.04472 0.06698 0.04287 0.04795 0.04026 0.05059 0.03863
0.04661 0.04684 0.06050

0.04801 0.04776 0.02517 0.03260 0.02026 0.02839 0.02512 0.02904
0.03364 0.04337

0.05331 0.03777 0.04112 0.03776 0.04867 0.03913 0.04798 0.04461
0.05595

0.04033 0.04348 0.03643 0.03739 0.03956 0.04670 0.03957 0.02392

0.01646 0.02083 0.02773 0.02252 0.02021 0.02565 0.03586

0.01985 0.02879 0.02302 0.02873 0.03268 0.03935

0.02113 0.01546 0.02076 0.02178 0.03601

0.02381 0.02867 0.02772 0.04253

0.02061 0.01569 0.03746

0.02674 0.04631

0.04129
```

APPENDIX 3

The polymerase chain reaction and cloning for future studies.

In order to expand the evolutionary study of the flowering plants, the addition of sequences from another molecule may be desirable. A new molecule may be informative at a level different from nuclear rRNA, that is, it may be informative below the subfamily level which appears to be the limit of resolution for the coding region of nuclear rRNA. Nuclear rRNA sequences could then be used for assigning taxa to the proper order or family, and relationships at the lower taxonomic levels could be resolved by sequences from the other molecule. Alternatively, if the second molecule were rRNA from one of the other plant genomes, mitochondrial or chloroplast, then even more interesting questions could be asked. For example, one can ask whether the rDNA of the nucleus evolves at the same rate as the rDNA of the plastids, and whether the patterns of change are similar? This could be tested, in part, by comparing phylogenetic trees inferred from both molecules. The chloroplast rRNA/rDNA is a good molecular yardstick to investigate these questions Preliminary evidence indicates that although the chloroplast is evolving overall more slowly than the nuclear genome, the rDNA of both is evolving at rates no different than two-fold.

The polymerase chain reaction, or PCR, (Saiki *et al.*, 1985) offers a rapid means to selectively amplify particular segments of DNA from a total DNA preparation so as to bypass cloning and screening a library. PCR does require some *a priori* knowledge of the primary sequence of the gene of interest so that primers can be synthesized specifically for the desired gene. Once the desired fragment is amplified, it can be cloned into a bacterial vector for sequencing. It is possible to sequence a double-stranded amplification product directly, but the failure rate is quite high. It is possible with a much higher success rate to sequence a single-stranded amplification product, but single-stranded amplifications are not always possible, so that it sometimes remains impossible to sequence both strands of a gene if desired.

Table 10 is a list of primers useful for PCR and sequencing the chloroplast rDNA.

Below I detail the protocols for PCR amplification from total DNA and my experiences with cloning. The PCR steps are straight forward and have been quite successful. The cloning of the PCR product has, on the other hand, been quite difficult, and my success has been very limited, despite an abundance of expert advice.

Table 10. A list of primers useful for PCR and sequencing within the chloroplast 16S rDNA. All positions are relative to those of tobacco (Tohdoh and Sugiura (1982).

Chloroplast 16S rRNA primers which anneal to the coding strand and to RNA

NAME	LENGTH	PRIMER SEQUENCE	ANNEALS TO
CT16A	18	CTGCTGGCACAGAGTTAG	TOBACCO 453-470
CT16B	18	AGGCGGGATACTTAACGC	TOBACCO 813-830
CTPCR3	18	CACCTTCCAGTACGGCTA	TOBACCO 1472-1455
3SAL3	28	GGAGGTCGACCACCTTCCAGTACGGCTA	TOBACCO 1472-1455
3PST3	28	GGAGCTGCAGCACCTTCCAGTACGGCTA	TOBACCO 1472-1455

Chloroplast 16S rRNA primers which anneal to the non-coding strand

NAME	LENGTH	PRIMER SEQUENCE	IDENTICAL TO
CTPCR5	18	ATGCTTAACACATGCAAG	TOBACCO 50-67
CT16BC	18	GCGTTAAGTATCCCGCCT	TOBACCO 818-835
5SAL5	28	GGAGGTCGACATGCTTAACACATGCAAG	TOBACCO 50-67

PCR Protocol

Note: Use the positive displacement Pipetmen for all additions to prevent contamination

1. If you are doing many reactions simultaneously, it can be faster to make a master mix of enzyme, buffer, dNTPs and water. To make a master mix allow for each tube:

Per Tube	Final Concentration
10 μl 10X Taq buffer	1X
10 μl of 1 mM dNTP mix (1 mM in *each* dNTP)	100 μM
0.5 μl of Taq polymerase (5 Units/ul)	2.5 Units
20 μl ddH$_2$O	

If you do not make a master mix, add water to the tubes first, then the buffer and dNTPs.

2. Aliquot out the master mix into 0.5 ml tubes. Be sure to mix the master mix well before each aliquot is removed because sometimes the enzyme sinks to the bottom and will only be dispensed into the last few tubes. For less than five tubes, I do not bother with a master mix.

3. For a double-standed amplification, add sufficient primer to bring the

final concentration of each to 1 μM.

4. Keep one tube for a negative control. Add enzyme to this tube (if not using a master mix), 50-100 μl of mineral oil and cap before adding DNA to any tubes.

5. Add 1 μg of total DNA to each tube.

6. Bring the total reaction volume to 100 μl by addition of double-distilled sterile H2O. Give the reactants a quick spin.

7. Layer on 50-100 μl of sterile mineral oil to prevent evaporation. Some people do not use mineral oil and claim that it does not affect their yield.

8. Place one drop of mineral oil into each well of the PCR heating block that you will use.

9. Label the reaction tubes *on the tops* as well as the sides because the mineral oil in the wells of the heating block will remove most anything written on the sides of the tubes.

9. Program the machine for the desired number of cycles and desired

temperatures. Most protocols tell you to choose an annealing temperature five degrees below the theoretical melting temperature of the oligo primer. This is calculated by multiplying the number of G's and C's by 4 and multiplying the number of A's and T's by 2, and then adding the two numbers together. This is fine for oligos up to say 20 to 22 nucleotides. For oligos longer than that, just use 49°C, it will work fine. I usually use 25 cycles, and the yield is generally good.

10. When the reactions are completed, the easiest way to remove the mineral oil is to drop the entire reaction mixture onto parafilm and roll it around. You can then lift off the reaction mix and leave the mineral oil behind on the parafilm. Alternatively you can extract once with chloroform.

11. Run 4 μl of each reaction mix on a minigel to confirm that the amplification was successful and that only one product was made.

I have normally produced chloroplast rDNA by PCR from total DNA preparations once, then diluted the reaction mixture to 1 ml. This diluted mixture then can be used for subsequent amplifications. Usually 10-20 μl of the dilute mix is sufficient for the next amplification. This is also quite helpful if you are consistently making more than one product as demonstrated by a minigel. If this happens, run the entire mixture out on a low melting temperature minigel, cut out the band of the desired length, recover the DNA by phenol:chloroform extraction and use this as a source of template for subsequent amplifications.

There are other tricks to get rid of extra amplification products: among them are lowering the concentration of dNTPs to as low as 25 μM, raising the annealing temperature, and lowering the concentration primers, or a combination of these. There are currently several manuals filled with protocols for PCR techniques, and they all have many helpful tips. Like any other laboratory skill, PCR is difficult at first, but with practice one can become fairly adept at the reactions.

Cloning the PCR product proved to be a very difficult task, one with a very low success rate. Blunt-end cloning did not work at all in my hands. An oligonucleotide synthesized for PCR (or any other purpose) normally will not have a 5' phosphate, so that the first step in blunt end cloning must be kinasing the PCR product. An alternative is to ligate linker molecules with restriction sites to the ends of the PCR product (the linkers must be purchased with 5' phosphates or they must be kinased first). A more straight-forward method is to incorporate a restriction site into the sequence of the oligonucleotide near the 5' end. After the PCR reaction is complete, the product must be recovered and digested with the proper restriction endonuclease. I always recovered the PCR product by two rounds of ethanol precipitation with ammonium acetate because it is supposed to keep unincorporated dNTPs and unused primers in solution. The dNTPs and primers can interfere with ligation reactions later. I believe now that the main reason for my poor success at cloning was that I purified the PCR product by this method. It has been recommended to me that the PCR product be gel purified and extracted from the gel by glass milk. It has also been suggested that for cloning that the primer concentration in the PCR reactions be reduced by a factor of ten.

Another problem occurs in quantifying the amount of DNA made in the PCR reaction. This is important for determining the ratio of insert to vector for the ligation steps. It is possible to dilute the PCR product to 250 μl and determine the concentration on a spectrophotometer using low-volume cuvettes. No dilution of the PCR product should be necessary to get a good reading. The PCR product can then be recovered by drying down the sample or ethanol precipitation. It is also possible to estimate the concentration of DNA by running a fraction of the DNA on a minigel along with some standards.

Once the PCR product is digested and quantified, the cloning is the same as any other directed cloning. I used Bluescript KS II (Stratagene) as a vector. Blue-white selection with this vector is not particularly good, and after about 12 hours at 4°C, almost all colonies will turn blue. It is possible that the chloroplast rDNA is lethal to the bacterial cell if it is expressed, as it must be to use blue-white selection. To get around this, I used a lacIQ super-repressor strain of *E. coli* for the transformation and stopped adding Xgal and IPTG to the plates. This meant that all colonies were white, and they all had to be screened for inserts.

I tried both in-gel and out-of-gel ligations; the only successful transformations that I got were out-of-gel ligations with a PCR product with restriction sites in the primers.

The following controls were used for each transformation: uncut vector, cut vector with no insert, and cut vector with ligase. The first control indicates whether the cells are competent. The second indicates the efficiency of the digestion of the vector and the third indicates if the ligase if active. Theoretically, there should be a confluent lawn on the first plate and there should be no surviving colonies on the second plate because cut vector cannot transform *E. coli*, hence all surving colonies are due to transformation by vectors that were uncut or partially cut. The third plate is used for background against which the actual transformations are measured. If the vector was cut with two different enzymes (because there were different restriction sites in the two PCR primers), then none of the vector should be able to religate and transform the bacteria. If the same restriction site is used for both ends of the PCR product, then the vector should be treated with alkaline phosphotase to remove 5' phosphates and prevent reclosing of the vector with the addition of ligase.

When the colonies are picked, a simple mini-prep procedure should be used to test for inserts. The positive clones should then be grown up overnight and an aliquot stored at -70°C for future recovery. The plasmid vector recovered from the mini-prep can then be used for sequencing with

Sequenase (U.S.B.). The protocol for an easy mini-prep is listed below.

1. Grow up 3.0 ml overnight cultures in LB.
2. Spin down 1.5 ml of culture for 10 s in microfuge.
3. Decant supernatant and resuspend in 50μl TE; vortex to resuspend cells.
4. Add 300 μl TENS (1X TE brought to 0.1N NaOH and 0.5% SDS), invert tube and vortex 3-5 seconds until mixture thickens.
5. Place tubes on ice until all are brought to this stage.
6. Add 150 μl KOAc (3M K^+, 5MOAc).
7. Spin 2-3 minutes in microfuge to pellet cellular debris and chromosomal DNA.
8. Transfer supernatant to fresh tube.
9. Extract once with 450 μl phenol:chlorform.
10. Extract once with 450μl chloroform.
11. Add 900 μl ethanol to precipitate DNA. Sit tube on ice for a few minutes or place in -20°C for a little while (this is actually not necessary usually).
12. Spin 15 minutes in microfuge.
13. Resuspend pellet in 50 μl TE.

This protocol can also be used to isolate plasmids for sequencing with the addition of an RNase digestion after step 12.

VITA

Robert Keith Hamby

HOME ADDRESS:	**SCHOOL ADDRESS:**
819 America Street	Department of Biochemistry
Baton Rouge, LA 70802	322 Choppin Hall
(504) 343-7444	Louisiana State University
	Baton Rouge, LA 70803

PERSONAL:

Date of Birth:	November 25, 1956
Place of Birth:	Mobile, AL
Citizenship:	USA

EDUCATION:

Louisiana State University, Baton Rouge, Louisiana - Jan 1985 to Present

Ph.D program in Biochemistry

GPA: 3.9/4.00

Expected Completion Date: December 1990

Thesis Advisor: Dr. Elizabeth A. Zimmer

Virginia Polytechnic Institute, Blacksburg, Virginia

M.S. in Chemical Engineering - Sept. 1984

GPA: 3.5/4.00

Thesis Title: Fundamental Studies of Unpacked and Screen-Packed Fluidized Beds in Axial and Transverse Magnetic Fields

Thesis Advisor: Dr. Y.A. Liu

Auburn University, Auburn, AL

Bachelor of Chemical Engineering, June 1978

GPA: 3.6/4.00

PUBLICATIONS:

Hamby, R.K. (1984). Fundamental studies of unpacked and screen-packed fluidized beds in axial and transverse magnetic fields. 410 Pages. M.S. Thesis, Virginia Polytechnic Institute.

Hamby, R.K., Habbal, M.K., Roos, J.W., and Ristroph, D.L., (1986). Feasibility of storing intermediate acids in two-stage anaerobic digestion of starch biomass. *Trans. Am. Soc. Ag. Eng.* **29**:1714-1719.

Hamby, R.K. and Zimmer, E.A. (1988). Ribosomal RNA sequences for inferring phylogeny within the grass family (*Poaceae*). *Plant Sys. Evol.* **160**:29-37.

Hamby, R.K., Sims, L.E., Issel, L.E., and Zimmer, E.A. (1988). Direct ribosomal RNA sequencing: Optimization of extraction and sequencing methods for work with higher plants. *Plant Mol. Biol. Rep.* **6**:172-195.

Zimmer, E.A., Hamby, R.K., Arnold, M.L., LeBlanc, D.A., and Theriot, E.L. (1989). Ribosomal RNA phylogenies and flowering plant evolution. *In* The Hierarchy of Life, Proceedings of the 1988 Nobel Foundation Symposium (eds B. Fernholm, K. Bremer and H. Jörnvall), Elsevier Press, Amsterdam, pp. 205-214.

Knaak, C., Hamby, R.K., Arnold, M.L., LeBlanc, M.D., Chapman, R.L., and Zimmer, E.A. (1989). Ribosomal DNA variation and its use in plant biosystematics. *In* Proc. 4th Int. Symposium on Plant Biosystematics, Academic Press, New York, in press.

Hamby, R.K., and Liu, Y.A., (1990). Studies in magnetochemical engineering: Part VI. An experimental study of screen-packed fluidized beds in axial and transverse magnetic fields. *Powder Tech.*, in press.

Hamby, R.K. and Zimmer, E.A. (1991). Ribosomal RNA as a phylogenetic tool in plant systematics. *In* Plant Molecular Systematics (eds P. Soltis, D. Soltis and J. Doyle), Chapman and Hall, New York, in press.

Liu, Y.A., Hamby, R.K, and Colberg, R.D. Fundamental and practical developments of magnetofluidized beds: A review. *Powder Technology*, in press.

Penny, D., Hendy, M.D., Zimmer, E.A., and Hamby, R.K. Trees from sequences: Panaceae or Pandora's box? *Australian Syst. Bot.*, **3**, in press.

ABSTRACTS

Hamby, R.K. and Liu, Y.A. (1985). Fundamental studies of packed magnetofluidized beds for particle - gas and particle-particle-particle separations. Digest of Technical Papers, 1985 IEEE International Magnetics Conference, Publications No. 85CH2180-8, p. AD-8, Institute of Electrical and Electronic Engineers, Inc. New York, New York.

Sims, L.E., Hamby, R.K., and Zimmer, E.A., (1986). Ribosomal RNA structure and evolution in higher plants. *Plant Phys.* **80**:236.

Hamby, R.K., Sims, L.E., and Zimmer, E.A., (1986). Ribosomal RNA sequence evolution in the Poaceae. International Symposium on Grass Systematics and Evolution Abstracts, p. 35.

Hamby, R.K., Issel, L.E., Zimmer, E.A., (1986). Direct sequencing of chloroplast ribosomal RNA. Am. Soc. for Microbiology and La. Biochem. Soc. 1986 Joint Meeting abstract No. CD-4.

Hamby, R.K., Issel, L.E., and Zimmer, E.A. (1987). Nuclear and organellar evolution in higher plants - ribosomal RNA as a marker molecule. *Genetics* **116**:s20.

Hamby, R.K, and Zimmer, E.A. (1988). A molecular perspective on monocot origins. *Genome* **30**s1:393.

Issel, L.E., Hamby, R.K., and Zimmer, E.A. (1990) Further development of a ribosomal RNA phylogeny for the grasses. *Proc. IV Int. Congress Syst. Evol. Biol.*, in press.

PRESENTATIONS:

1986	Presented poster at ASPP National Meeting, LSU-Baton Rouge.
1986	Presented poster at AIBS International Symposium for Grass Systematics.
1986	Presented talk at American Society for Microbiology and Louisiana Biochemistry Society Joint Meeting: "Direct Sequencing of Chloroplast Ribosomal RNA."
1987	Presented poster at Genetics Society of America national meeting: "Nuclear and Organellar Evolution in Higher Plants - Ribosomal RNA as a Marker Molecule."
1988	Presented a poster at the XVIth International Congress of Genetics in Toronto, Canada: "Molecular Perspectives on Monocot Origins."
1990	Presented a talk entitled "Ribosomal RNA and the early evolution of flowering plants" at the Ninth Willi Hennig Society Meeting, Canberra, Australia

HONORS AND AWARDS:

1984-1988	LSU Alumni Federation Fellowship
1987	Graduate Student Travel Grant by Genetics Society of America to attend national meeting.
1988	Young Investigator Travel Grant to attend XVIth International Congress of Genetics in Toronto.
1988	Fellowship to attend UCLA International School of Molecular Evolution.

WORK EXPERIENCE:

August 1978 -	Dow Chemical USA, Louisiana Division, P.O. Box 150, Plaquemine,
March 1982	LA 70764

Research Engineer in R&D. Designed and built a minireactor system in support of Chlorinated Methanes

Unit in order to gather data for catalyst research project. 8/78 - 11/79

Production Engineer in Naphtha Cracker. Assistant to, and later, engineer in charge of utilities section of plant. Assisted in eight-month long start-up of the operation, and trouble shooting of steam systems, cooling tower, water polishers, air compressor and nitrogen and fuel gas systems. Extensive experience in chemical treatment of boilers and cooling towers. 11/79 - 3/81

Production engineer in charge of daily operation of Louisiana Division Water Treatment Facility. Oversaw shift personnel, scheduled maintenance and directed physical improvements to facilities. Also conducted field test for corporate research program in reverse osmosis. Experienced in clarification, ion exchange, water polishing, deaeration and reverse osmosis. 3/81 - 3/82.

TEACHING EXPERIENCE:

Jun - Aug 1983,

Jun - Aug 1984	Instructor in Chemical Engineering Department, Virginia Polytechnic Institute. Designed experiments, and instructed undergraduates in theory and operation of industrial-scale distillation columns, fluidized beds, and other chemical processing equipment.
Sep - Dec 1985	Instructor in Chemical Engineering Department, Virginia Polytechnic Institute. Taught a course in process control by minicomputers. Designed laboratory experiments to introduce undergraduate to organization and operation of TTL-based computers.

RESEARCH EXPERIENCE:

1977 - 1978	Undergraduate research in the removal of sulfurous particles from pulverized coal by magnetic separation.
1978 - 1979	Proprietary catalyst evaluation for Dow Chemical USA.
1982 - 1984	Masters thesis research on the magnetic stabilization of fluidized beds. Designed and built solenoid to create an invariant magnetic field oriented axially to the bed, this axial system conformed to constraints imposed by an

	existing transverse orientation. Investigated use of various packing materials and non-magnetic materials in the fluidized bed.
1985	LSU Chemical Engineering Department. Performed feasibility studies on storage of biomass by-product. Learned wet lab analyses for CO_2, TOD, as well as HPLC and GC techniques for acid composition determinations.
1985-present	LSU Biochemistry Department. Using molecular biological techniques including RNA/DNA isolation, plasmid isolation, reverse transcription, polymerase chain reaction and electrophoresis to investigate phylogenetic relationship among higher plants. I have extensive computer experience. Marker molecules used are nuclear and chloroplast ribosomal RNAs.

DOCTORAL EXAMINATION AND DISSERTATION REPORT

Candidate: Robert Keith Hamby

Major Field: Biochemistry

Title of Dissertation: Ribosomal RNA and the Early Evolution of Flowering Plants

Approved:

Elizabeth A. Zimmer

Major Professor and Chairman

H Keim

Dean of the Graduate School

EXAMINING COMMITTEE:

Martin Hjortso

Kathleen Morden

Billie J. Bartlett

Walter A. Deutsch

Russell L. Chapman

Date of Examination:

8/22/90