NOTE TO USERS

University of Alberta

Analysis of multiple levels of vascular-plant phylogeny using a nuclear RNA polymerase subunit gene (RPB2) *and a large plastid data set: Studies of the tracheophytes, cycads and the lilies and relatives*

by

Jessie Zgurrski

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the

requirements for the degree of *Master of Science*

in
Systematics and Evolution

Department of *Biological Sciences*

Edmonton, Alberta
Fall, 2004

Your file Votre référence
ISBN: 0-612-95888-4
Our file Notre référence
ISBN: 0-612-95888-4

Acknowledgements

Many people made this project possible by providing plant samples, or by generously sharing their time and knowledge with me. I would like to thank my supervisor, Sean Graham, for giving me the opportunity to conduct this research and for his guidance and advice. I would also like to thank Ruth Stockey for co-supervising my project and, Curtis Strobeck and Bruce Rannala for being part of my exam committee.

I am very grateful to Hardeep Rai and Marc McPherson for their help with laboratory procedures and data analysis. I would also like to thank Pat Murray and Lisa Ostafichuck for their useful advice and assistance regarding DNA sequencing, Benjamin Hall for letting me visit his lab, and Jie Luo for help with RNA work. Many thanks to my husband, Quentin, for laboratory assistance, support and encouragement, as well as my parents, Phillip and Yvonne, for their support and encouragement.

I would also like to acknowledge those who kindly helped me obtain plant samples: Javier Francisco-Ortega (Fairchild Tropical Gardens), Peter Weston (Royal Botanical Gardens, Sydney), Gar Rothwell (Ohio University), John Bain (University of Lethbridge), Steven Williams (University of Alberta Greenhouses), and Doug Ewing (University of Washington Greenhouses). This study was supported by NSERC and Alberta Ingenuity grants to JMZ.

Table of Contents

List of Tables

List of Figures

~Chapter One~

Phylogeny of Vascular Plants: An Overview.

Introduction

Our understanding of evolutionary relationships among and within the major lineages of vascular plants has altered dramatically over the past two decades, primarily because of the availability of technology (hardware and software) for producing and analyzing new types of data – mainly DNA sequences (Savolainen and Chase, 2003; Felsenstein, 2004). The use of DNA sequence data for inferring vascular-plant phylogenetic relationships became prevalent during the early 1980s, with the production of sequences for the chloroplast gene *rbc*L, which codes for the large subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase (RUBISCO). Universal primers that can be used to amplify and sequence the gene for green plants were developed and made available to plant systematists at no charge (Zurawski et al., 1981). Systematists sequenced the gene for hundreds of plant species, in what came to be a large co-ordinated effort (Chase et al., 1993). A number of other genes from all three plant genomes have since been characterized and sequenced, and as a result, we now have a much better understanding of relationships among conifer families (e.g., Chaw et al., 1997; Stefanovic et al., 1998; Gugerli et al., 2001; Quinn et al., 2002), cycad genera (e.g., Rai et al., 2003; Hill et al., 2003; Bogler and Francisco-Ortega, in press), pteridophytes (e.g., Hasebe et al.,1994, 1995; Pryer

et al., 1995, 2001; Wolf, 1997), and especially the flowering plants (e.g., Soltis et al., 1999; 2000; Qiu et al., 1999, 2000; Graham and Olmstead, 2000; Savolainen et al., 2000), and within them, the monocots (e.g., Chase et al., 2000; McPherson, 2003).

However, there are still a substantial number of unanswered questions relating to vascular-plant phylogeny, particularly with regards to the deep relationships among major lineages (e.g., Doyle, 1998). For example, the relationships among the five extant seed plant groups remain controversial, as several recent papers have arrived at well-supported, but strongly conflicting conclusions (Bowe et al., 2000; Chaw et al., 2000; Rydin et al., 2002; Soltis et al., 2002; Rai et al., 2003). In addition, several free-sporing taxa, such as *Equisetum*, remain difficult to place.

Overview of Deep Vascular- Plant Phylogeny

The vascular plants (tracheophytes) are a diverse group that date back at least to the upper Silurian (Stewart and Rothwell, 1993). They are united by the presence of annular or helical thickenings on their water conducting cells (tracheids) (Kenrick and Crane, 1997). In terms of extant taxa, the tracheophyte clade consists of six major lineages of seedless plants, and a clade of five extant seed-bearing lineages (the flowering plants and four groups of gymnosperms: the conifers, cycads, *Ginkgo* and Gnetales). The extant lineages of seedless plants are Equisetales, Lycopodiales, Marattiales, Ophioglossales, Polypodiales, and Psilotales. Five of these orders (Equisetales, Marattiales, Ophioglossales, Polypodiales, and Psilotales) appear to comprise a clade (Pryer et al., 2001), and they are sometimes referred to collectively as the "moniliforms" or

"monilophytes" (Kenrick and Crane, 1997). The clade comprising moniliforms and seed plants (i.e., vascular plants minus lycopods) is often referred to informally as the euphyllophytes (Kenrick and Crane, 1997). I outline here recent taxonomic and phylogenetic concepts of each major group. It should be noted that within the seed plants (spermatophytes), extant gymnosperms (conifers, cycads, *Ginkgo* and Gnetales) and angiosperms (flowering plants) represent only a small proportion of the total diversity of major seed-plant taxa. Many distinct lineages, including multiple groups of seed ferns (seed plants with a fern-like vegetative morphology; pteridosperms), are now extinct. This overview focuses on extant taxa, as only these are amenable to study using molecular systematics. For the sake of convenience I have divided my overview into two groups, the pteridophytes and the seed plants, although only the latter appear to be monophyletic. Nonmonophyletic taxa are referred to from here on using quotation marks (e.g., "pteridophytes"). Taxa are introduced here at the highest Linnaean rank at which they are widely recognized (Ginkgoales and Ophioglossales, for example), but in the rest of my thesis I generally refer to taxa at lower ranks, when the higher ranks do not contain additional extant taxa (thus: *Ginkgo* and Ophioglossaceae).

The "pteridophytes" (ferns and allies)

I. Equisetales

There is only one extant family in Equisetales (Equisetaceae), which contains the single genus, *Equisetum* L., and 15 species (Hauke, 1990). *Equisetum* (horsetails) occurs on every continent except Australia and Antarctica,

and tends to occur in sunny, damp habitats. Equisetales were a far more diverse group during the Carboniferous, and contained herbaceous and arborescent, and homosporous and heterosporous members (Stewart and Rothwell, 1993). Equisetaceae are likely descendents of Calamitaceae, a family of large trees that existed from the Carboniferous to the Permian (Stewart and Rothwell, 1993). Both taxa have stem nodes with whorls of leaves or branches, a similar internal stem anatomy (with a large empty pith cavity), spores with elaters and a rhizomatous habit. Unlike Calamitaceae, Equisetaceae lack secondary tissue and have bractless strobili.

Molecular analyses have not placed *Equisetum* in a clear and consistent position among the deep branches of vascular-plant phylogeny. An analysis of the mitochondrial 18S rDNA locus placed it as the sister taxon of leptosporangiate ferns (Duff and Nickrent, 1999), and an analysis of morphology, three chloroplast genes and one nuclear gene place it within the moniliforms, as the sister group of the eusporangiate fern family, Marattiaceae (Pryer et al., 2001). However, analyses based on morphological and cellular characters (Parenti, 1980) or male gametogenesis characters (Garbary et al., 1993) place Equisetales as the sister group of all ferns. An analysis based on a limited number of extinct and extant taxa placed it as the sister group of lignophytes, a group composed of the progymnosperms and the seed plants (Rothwell and Serbet, 1994). Clearly, more investigation is needed concerning the placement of this taxon.

II. "Filicopsida"

There are over 12,000 species of extant "true" ferns, making it the second largest group of vascular plants and the largest extant group of spore-forming plants. The ferns ("Filicopsida") have often been divided into two groups: the eusporangiate and leptosporangiate ferns. For the sake of convenience, I follow current usage and use "fern" in its conventional sense. However, the "fern" concept may need to be substantially revised in future textbooks, because of recent phylogenetic findings (Pryer et al., 2001) that indicate that class "Filicopsida" may be non-monophyletic. Characteristics of modern ferns (ignoring Equisetaceae and Psilotaceae) include a free-sporing habit, the presence of large, complex leaves derived from modified branching systems, mesarch steles, and sporangia located at the tips or margins of pinnules or the abaxial surface of the leaves. The eusporangiate ferns (~375 extant species; Kramer and Green, 1990) have massive, thick-walled sporangia that arise from numerous initials and produce a large number of spores. The leptosporangiate ferns (~12,000 extant species) typically have thin-walled sporangia that arise from a single initial cell, and produce a relatively small number of spores. A phylogenetic analysis of Rothwell (1999) found that ferns are a polyphyletic group composed of three main clades: (1) A group composed exclusively of extinct taxa including Cladoxylales and zygopteraleans; (2) A group composed of extant and extinct eusporangiate ferns, including Ophioglossales and Marattiales, and; (3) A group composed of extinct and extant leptosporangiate ferns. Morphological analyses support the monophyly of a clade consisting of Marattiales and Ophioglossales, and their status as the sister group of leptosporangiate ferns (Rothwell, 1999). This is partly consistent with molecular

analyses if the recent molecular findings on the position of Equisetaceae and Psilotaceae are ignored (Pryer et al., 2001)

II a). Eusporangiate Ferns: Ophioglossales and Marattiales.

The two extant orders of eusporangiate ferns each consist of a single family: Ophioglossales (Ophioglossaceae) and Marattiales (Marattiaceae). There are also several extinct groups of eusporangiate Palaeozoic ferns (Stewart and Rothwell, 1993). Marattiaceae are composed of four extant genera (*Angiopteris* Hoffman, *Christensia* Maxon, *Danaea* Smith and *Marattia* Swartz), all of which are confined to the tropics or subtropics (Camus, 1990). The family possesses a number of characteristics unique among modern ferns, including mucilage canals in the roots, leaves and the stem, large starchy stipules and multi- and unicellular root hairs (Camus, 1990). The group has an extensive fossil record, and by the upper Carboniferous it was very widespread (Stewart and Rothwell, 1993). The family reached its peak diversity during the Mesozoic, before the Cretaceous, and modern genera in the order appeared during that period (Tidwell and Ash, 1994).

Ophioglossaceae are made up of three genera (*Botrychium* Sw., *Helminthostachys* Kaulfuss, and *Ophioglossum* L.) and have a worldwide distribution. The family has a very sparse fossil record, although Rothwell and Stockey (1989) described the fossil plant *Botrychium wightonii* Rothwell & Stockey from the Paleocene, which is very similar to extant *Botrychium virginianum* (L.) Swartz. A few authors have suggested that Ophioglossales are not true ferns, but are instead related to extinct progymnosperms, because members of the order possess characteristics that are unusual for ferns, including upright stems, a eustele-like vascular organization, the presence of a cork cambium and large, circular bordered pits (Wagner, 1990; Bierhorst, 1971).

However, the phylogenetic analysis of Rothwell (1999) demonstrated that Ophioglossaceae are indeed nested within the ferns, broadly construed.

II (b). Leptosporangiate Ferns (Polypodiales)

The origin of modern leptosporangiate ferns remains obscure, but it has been suggested that they are descendents of the "trimerophyte complex" (Stewart and Rothwell, 1993; Rothwell and Serbet, 1994). The oldest ones date back to the early Devonian, and by the end of the Carboniferous, there were at least five families, all now extinct (Stewart and Rothwell, 1993). Classifications vary on the number of modern families recognized. Kramer and Green (1990) recognize 31 families, and I follow their treatment here.

Osmundaceae, a widespread family of three genera (*Leptopteris* C. Presl, *Osmunda* L., and *Todea* Willdenow in Bernhardi), are well-supported by molecular evidence as being the sister group of all remaining extant leptosporangiate families (Hasebe et al., 1994, 1995; Manhart, 1995; Pryer et al., 1995, 2000; Raubeson and Stein, 1995; Wolf, 1997). This is congruent with morphological observations suggesting that Osmundaceae are intermediate between eu- and leptosporangiate ferns, as they have large thin-walled sporangia with massive stalks that produce a large number of spores, and which develop from multiple initial cells. The fossil record also suggests that Osmundaceae are an ancient group. The family has the most extensive fossil record of any group of ferns, with over 150 extinct species (Tidwell and Ash, 1994). Its fossil record begins in the late Permian, making it the oldest known extant fern family (Tidwell and Ash, 1994), and fossils virtually identical to modern *Osmunda* species have been found that date to the Triassic (Phipps et al., 1998) and Cretaceous (Serbet and Rothwell, 1999).

Morphological studies have identified nine other families that occupy basal positions among extant "Filicales" (Smith, 1995). These are: Cyatheaceae, Dicksoniaceae, Dipteridaceae, Gleicheniaceae, Hymenophyllaceae, Loxsomataceae, Matoniaceae, Plagiogyriaceae and Schizaeaceae. Molecular analyses of leptosporangiate ferns (Hasebe et al., 1995; Pryer et al., 1995; Wolf et al., 1997; Pryer et al., 2001) have largely confirmed these results, although support for this part of the backbone of leptosporangiate fern phylogeny is poor (Hasebe et al. 1995). Clades emerging from basal nodes of the leptosporangiate ferns include Hymenophyllaceae (the "filmy ferns") and the "gleichenioid" ferns (Gleicheniaceae and Matoniaceae; Pryer et al., 2001). The fossil record suggests that Gleicheniaceae are an ancient group, as the family appears to have a Palaeozoic origin (Tidwell and Ash, 1994). Dipteridaceae and Cheiropleuriaceae may belong in a clade with the gleichenioid ferns or they may represent a distinct basal clade (see Hasebe et al., 1995). Molecular results also indicate that Schizaeaceae represent an additional near-basal lineage in the leptosporangiate ferns (Hasebe et al., 1995; Pryer et al, 2001).

The heterosporous water ferns and tree ferns represent two major clades that emerged along the mid-point of the backbone of leptosporangiate fern phylogeny ("mid" from the perspective of current taxonomy). The water fern families Azollaceae, Marsileaceae and Salviniaceae are unique among leptosporangiate ferns because they are both heterosporous and aquatic (or amphibious). These three families were previously placed in two orders, as it was thought that they evolved from separate homosporous, terrestrial ancestors (Bierhorst, 1971). However, molecular evidence suggests that they form a clade (Hasebe et al. 1995; Pryer et al., 2001) that is the sister group of the tree ferns and the remaining leptosporangiate ferns (Hasebe et al. 1995; Raubeson and Stein, 1995; Pryer, 1999; Pryer et al. 2001). The fossil taxon *Hydropteris*

pinnata is morphologically intermediate between Salviniaceae and Marsileaceae, and in cladistic analyses it groups within a clade of heterosporous water ferns, as the sister group of Salviniaceae and Azollaceae (Rothwell and Stockey, 1994).

The tree ferns, which are comprised of the families Cyatheaceae, Dicksoniaceae, and Metaxyaceae all belong in a clade that also contains Plagiogyriaceae and Loxsomataceae (Hasebe et al., 1995; Smith, 1995; Pryer et al., 2001). Hasebe et al. (1995) also confirm that Plagiogyriaceae are not a close relative of Osmundaceae, as was previously thought (Mickel, 1974). The tree ferns (broadly construed to include Plagiogyriaceae and Loxsomataceae) have been found to be the sister group of a large clade that includes all remaining leptosporangiate taxa (Pryer et al., 2001). Members of Dennstaedtiaceae, Monachosoraceae, Pteridaceae and Vittariaceae emerge from the basal nodes in the polypodiaceous clade (Hasebe et al., 1995). However, substantial re-circumscription of Dennstaedtiaceae and Pteridaceae may be necessary, as members of these families are apparently dispersed among several clades in the polypodiaceous ferns (Hasebe et al., 1995; Wolf, 1995). The remainder of the polypodiaceous clade includes Aspleniaceae, Blechnaceae, Davalliaceae, Dryopteridaceae, Grammitidaceae, Lomariopsidaceae, Nephrolepidaceae, Oleandraceae, Polypodiaceae and Thelypteridaceae (Hasebe et al., 1995). A number of these families appear to be non-monophyletic as currently circumscribed, including Dryopteridaceae and Polypodiaceae (Hasebe et al., 1995), and the affinities of some leptosporangiate ferns, such as Hymenophyllopsidaceae, remain obscure (Hasebe et al., 1995).

III. Lycopodiales

There are three extant lycopod families (Isoetaceae, Lycopodiaceae, and Selaginellaceae) all of which are small herbs. Lycopodiaceae, with approximately 380 species, were traditionally divided into two genera, *Lycopodium* L. and *Phyloglossum* Kunze, although a recent treatment of Wagner and Beitel (1992), recognizes seven genera. Selaginellaceae contains one large genus, *Selaginella* Pal. Beauv., with about 750 species. Isoetaceae also contains only one genus, *Isoetes* L., with about 130 species. A second genus, *Stylites* Amstutz, was described in 1957 (Amstutz, 1957), but this taxon is currently recognized at the subgeneric level (Jermy, 1990). All lycopod families are cosmopolitan, although Lycopodiaceae are absent from arid areas. Lycopods are characterized by the presence of microphylls, and adaxial, reniform sporangia. Overall, they do not constitute a very large proportion of the earth's terrestrial flora, although this was not the case during the Carboniferous, when the group was far more diverse, and large, arborescent lycopods dominated forests.

Lycopods have the longest fossil record of any extant vascular plant group. The oldest representative of the group is the upper-Silurian or lower-Devonian aged *Baragwanathia longifolia* Lang & Cookson., which is remarkably similar to modern day *Huperzia* Bernhardi (Garratt, 1984). Lycopods have generally been regarded as descendents of zosterophyllophytes (Banks, 1968; Kenrick and Crane, 1997; Gensel, 1992), a paraphyletic Silurian/ Devonian assemblage (Kotyk et al., 2002) which consisted of short, herbaceous plants with lateral, reniform sporangia, and exarch protosteles. Both lycopods and zosterophyllophytes possess exarch protosteles and reniform sporangia, but zosterophyllophytes differ from lycopods in lacking microphylls. During the late

Silurian, the zosterophyllophyte/ lycopod lineage was diverse and well-established (Kotyk et al., 2002). By the Devonian, lycopods comprised two main lineages: those that were homosporous and lacked ligules (represented today by Lycopodiaceae) and those that were heterosporous and possessed ligules (represented today by Selaginellaceae and Isoetaceae). One fossil taxon (*Leclerqia* Tsuneki) is intermediate between these two groups as it is homosporous and ligulate (Grierson and Bonamo, 1979). Further diversification of lycopods occurred during the Carboniferous, with the appearance of large, arborescent heterosporous lycopods in the order Lepidodendrales, which were extinct by the Permian.

Extant lycopods are likely the sister group of all remaining extant vascular plants, based on analyses of a variety of data types. For example, a major chloroplast genome inversion found only in bryophytes (non-vascular plants) and lycopods is lacking in all other vascular plants (Raubeson and Jansen, 1992). Additional support for their position as the sister group of all other vascular plants comes from morphological (Kenrick and Crane, 1997) and molecular (Duff and Nickrent, 1999; Pryer et al., 2001) evidence. These studies also indicate that Isoetaceae and Selaginellaceae form a clade that is the sister group of Lycopodiaceae.

IV. Psilotales

Psilotales are a small and enigmatic group that contains only one extant family, Psilotaceae (whisk ferns), with two genera: *Psilotum* Sw. and *Tmesipteris* Sw. The actual number of species in this group is unclear. Several species of *Psilotum* have been described, but only two of them are usually recognized

(Kramer, 1990). There are ten species of *Tmesipteris*, but many of them are very poorly known (Kramer, 1990). *Psilotum* has a widespread distribution, but *Tmesipteris* is found only in southeastern Asia and Australasia. As this family contains leafless, rootless plants that are similar to Devonian rhyniophytes, it has been implicated as being a potential close relative, a conclusion supported by a limited number of studies (Parenti, 1980; Bremer et al., 1987). However, no rhyniophyte fossils have yet been discovered that date after the mid-Devonian, and there are no known ancient Psilotaceae fossils (Gensel, 1977; Stewart and Rothwell, 1993), and the family has no other obvious relatives among other fossil plants (Gensel, 1977). Bierhorst (1971, 1977) argues that Psilotaceae are related to *Stromatopteris* Mett., a fern genus from the family Gleicheniaceae. He based this theory on his observations of embryo and gametophyte characters in the two taxa. However, Wagner (1977) argues that the two taxa are too distinct to be considered close relatives, and that Psilotales are not a close relative of any extant group. A morphological analysis by Rothwell (1999), which included fossils, placed the family as the sister group of euphyllophytes. This result is not congruent with molecular studies, which indicate that the family is the sister group of Ophioglossaceae, a family of eusporangiate ferns (Manhart, 1995; Wolf, 1997; Pryer et al., 2001). Psilotaceae could be viewed as a group of highly modified eusporangiate ferns under the latter phylogenetic placement.

The seed plants

Relationships among the extant seed plants: An overview

Some recent molecular studies (Chaw et al., 1997, 2000; Bowe et al., 2000; Schmidt and Schneider-Poetsch, 2002) support a clade consisting of conifers, cycads, *Ginkgo* and Gnetales; the gymnosperms ("naked" seed-bearing plants, where the seed is not protected by a carpel). They depict the angiosperms as the sister-group of all remaining extant seed plants. This result is not congruent with the fossil record, as it would suggest that the line leading to the angiosperms (flowering plants, where seeds are protected by a carpel) lead a long, unrecorded existence independent from the other extant seed-plants. Recognizable angiosperms appeared only relatively recently in the fossil record, about 130 million years ago (Crane et al., 1995). This result, if correct, would imply that the synapomorphies that we use to recognize angiosperms (such as the carpel) arose only recently in the stem-lineage leading to them, a line that by this view split off from the seed plants near their origin, and that evades collection or correct interpretation in the fossil record. Many angiosperm synapomorphies, such as a reduced megagametophyte, double fertilization leading to a triploid endosperm and the presence of companion cells in the phloem (Judd et al., 2002), would not be well-preserved in fossils.

Systematists disagree as to whether or not *Ginkgo* is the sister group of the cycads. Rai et al. (2003) found a moderately-supported sister-group relationship between these two using a large plastid data set, and noted that they share a highly reduced rate of evolution in the chloroplast genome, and an elevated transition-transversion ratio. Chaw et al. (1997) also found support for a

cycad-*Ginkgo* relationship. Other molecular studies, however, place *Ginkgo* as the sister group of a Gnetales-conifer clade (Bowe et al., 2000; Chaw et al., 2000; Soltis et al., 2002), and some morphological analyses have placed it as the sister group of conifers (Crane, 1985; Doyle and Donoghue, 1986, 1992). The phylogenetic position of cycads among the other extant groups of seed plants is thus also an open question. Several morphological studies suggest that cycads are the sister taxon of all remaining extant seed plant groups (Crane, 1985; Doyle, 1996; Doyle and Donoghue, 1986, 1992; Loconte and Stevenson, 1990), and a number of molecular studies (Bowe et al., 2000; Chaw et al., 2000; Magellon and Sanderson, 2002) also support this relationship. However, Rai et al. (2003) and Rydin et al. (2002) found Gnetales as the sister group of all remaining extant seed plants, and in both cases this result was well supported by bootstrap analyses of very large data sets. Both sets of workers noted that this result could be misleading and a possible consequence of long-branch attraction (Felsenstein, 1978; Hendy and Penny, 1989). Sanderson et al. (2000) quantified the propensity for erroneous placements of the different seed plants using several plastid genes, and found substantial opportunity for artifactual tree inference. This suggests that all current phylogenetic results concerning seed-plant relationships should be treated with great caution.

The placement of conifers among the other extant gymnosperm groups also remains unclear. Some morphological studies place them as the sister group of Ginkgoales (Parenti, 1980; Crane, 1985; Doyle and Donoghue, 1986), but others do not (Loconte and Stevenson, 1990). Many molecular studies find the conifers to be related to Gnetales among extant vascular-plant groups (Goremykin et al., 1996; Chaw et al., 1997; Chaw et al., 2000; Bowe et al., 2000). The position of Gnetales among the deep branches of vascular-plant phylogeny is particularly problematic. A number of morphological cladistic

analyses have supported a sister-group relationship between angiosperms and Gnetales (Crane, 1985; Doyle and Donoghue, 1986, 1992; Doyle, 1996; Loconte and Stevenson, 1990). Potential synapomorphies for an angiosperm-Gnetales clade include: a tunica in the apical meristem, lignin chemistry (Maüle reaction), a shift to granular exine in the pollen, and reduction of the megaspore wall (Doyle, 1998). Angiosperms and Gnetales may belong in the same clade as the extinct Bennettitales and *Pentoxylon,* and these groups have together been referred to as the "anthophytes" based on their flower-like reproductive structures (Crane, 1985; Donoghue and Doyle, 1986, 1992; Loconte and Stevenson, 1990; Doyle, 1996). Molecular data have yet to provide any evidence for a sister-group relationship between angiosperms and Gnetales. However, the simulation studies of Sanderson et al. (2000) suggest that if this relationship were true, it would be particularly difficult to infer using DNA sequences.

Most early molecular studies suggesting a close relationship of Gnetales to conifers (Goremykin et al., 1996; Chaw et al., 1997; Bowe and dePamphilis, 1997) were based on a single gene. More recent studies using DNA sequence data from all three genomes (Bowe et al., 2000; Chaw et al., 2000) find Gnetales to be embedded within them, as the sister-group of Pinaceae, the so-called gnepine hypothesis. In contrast, Rai et al. (2003) and Rydin et al. (2003) found Gnetales to be the sister group of all other seed plants. These conflicting studies are often very well-supported by bootstrap analyses (Felsenstein, 1985), and, as already noted, these results could possibly be a result of systematic error, such as long-branch attraction (Sanderson et al., 2000).

I. Angiosperms

Very substantial progress has been made in flowering-plant phylogeny, making it one of the best characterized, phylogenetically, of all major groups of organisms (Savolainen and Chase 2003, APG II 2003). This large clade is comprised of ~260,000 species, divided into ~40 well circumscribed and well supported orders (Judd et al., 2002). Nonetheless, substantial phylogenetic work remains to be done concerning relationships among and within each order. I contributed to ongoing studies of flowering-plant phylogeny by addressing higher-order relationships within a group of flowering plants in the monocots, the order Liliales (see Chapter 4). Although this is one of the smaller orders of flowering plants, it includes more species (~1,300, Judd et al., 2002) than all living gymnosperms combined. Current concepts of Liliales are discussed in more detail in Chapter 4.

II. Coniferales

With approximately 650 species grouped into ~70 genera (Kramer and Green, 1990; http://schwarzbach.biology.kent.edu.conifer/), the conifers are by far the largest extant gymnosperm group. The extant conifers are trees or shrubs, which possess pycnoxylic wood surrounded a very small pith and cortex. Most have simple, needle-like, scale-like or strap-shaped leaves, which generally have 1-2 veins. They can be either mono- or dioecious, and all (except Taxaceae) have compound ovulate cones, with seed-bearing cone scales derived from fertile short shoots, simple pollen cones (compound in some extinct members; Hernandez-Castillo et al., 2001), resin canals in wood, leaves, roots and/or the

seed coat (except in some Taxaceae) and tiered proembryos. Conifers are found in both hemispheres and are most common in cooler areas. Kramer and Green (1990) recognize nine families of conifers: Araucariaceae, Cephalotaxaceae, Cupressaceae, Phyllocladaceae, Pinaceae, Podocarpaceae, Sciadopityaceae, Taxaceae and Taxodiaceae. However, molecular (Stefanovic et al., 1998) and morphological (Eckenwalder, 1976) evidence supports grouping Cupressaceae and Taxodiaceae into one family, Cupressaceae *s.l.* Opinions differs as to whether the distinctive genus *Phyllocladus* Rich. ex Mirbel should be included in Podocarpaceae, or recognized in its own family (e.g., Farjon, 1998), Phyllocladaceae. Recent molecular work indicates that *Phyllocladus* is either nested in Podocarpaceae (Conran et al. 2000; Kelch 2002), or is the sister group of all other Podocarpaceae (Sinclair et al., 2002). In the former case a combination in Podocarpaceae would be needed, in the latter it would be a matter of taste whether to combine it, or recognize it at the family level. A similar level of uncertainty exists concerning whether Cephalotaxaceae (1-2 genera, depending on whether *Amentotaxus* Pilg. is placed here or in Taxaceae) should be recognized as its own family, or combined within Taxaceae (Quinn et al. 2002).

The extant conifers have sometimes been divided into two orders with distinct phylogenetic origins: Taxales and Coniferales. Taxales were defined by Florin (1951) to contain a single family, Taxaceae, which consist of four-five genera and ~20 species, that are found primarily in the Northern Hemisphere (Page, 1990). This family has sometimes been regarded as a separate lineage from other conifers because its female "cone" contains only a single terminal, erect ovule, and there are no resin canals in the leaves (Page, 1990). Miller (1999) suggests that Taxaceae may have descended from a different group of fossil conifers (Utrechtiaceae) than the remaining families, which may be

descendents of Majonicaceae. Taxaceae place as the sister group of all other conifers in the cladistic analysis of Miller (1999), which included fossil conifers. However, many other morphological and molecular studies have demonstrated that Taxaceae are well nested within the conifer clade (e.g., Hart, 1987; Chaw et al., 1997; Stefanovic, 1998; Quinn et al., 2002). Molecular analyses have consistently inferred Pinaceae (a primarily Northern Hemisphere family of 12 genera and ~200 species), as the sister group of all remaining modern conifer families (Chaw et al., 1997; Stefanovic et al., 1998; Gugerli et al., 2001; Quinn et al., 2002; Schmidt and Schneider-Poetsch, 2002), as does a morphological cladistic study (Hart, 1987). However, Pinaceae have a young fossil record (early Cretaceous; Miller, 1999), which is incongruent with their inferred position as the sister group of all other conifers, because most of the other extant families have fossil records dating back to the Triassic (Stewart and Rothwell, 1993). This suggests that the stem lineage leading to Pinaceae arose long before either the crown Pinaceae (i.e., all extant taxa and descendants of their most recent common ancestor), or before the various features by which we recognize this taxon, such as the presence of a seed wing that develops from the cone scale, the presence of two inverted ovules on the adaxial face of each scale, and bract/ scale complexes that are free for most of their length from the subtending bracts (Thieret, 1993; Judd et al., 2002). Many recent molecular studies have also supported a sister-group relationship between Pinaceae and Gnetales (see below). If correct, this would indicate that some of the coniferous features of Pinaceae, such as resin canals and tiered proembryos, arose in parallel with other conifers, or that they were lost (or transformed beyond recognition) in the stem lineage leading to modern Gnetales (Donoghue and Doyle, 2000). However, either scenario seems unlikely, so studies that indicate that Pinaceae and Gnetales are related should be regarded with some skepticism.

Molecular evidence demonstrates that the families Podocarpaceae and Araucariaceae are sister taxa, a clade which is in turn the sister group of all remaining conifers except Pinaceae (Chaw et al., 1997; Stefanovic et al., 1998; Schmidt and Schneider-Poetsch, 2002). Hart (1987) did not find this result, although he used only a small number of morphological characters. Sciadopityaceae are a monotypic family (comprised of *Sciadopitys verticillata* Sieb. & Zucc.) and are the sister-taxon of Taxaceae, Cephalotaxaceae and Cupressaceae (Chaw et al., 1997; Stefanovic et al., 1998; Schmidt and Schneider-Poetsch, 2002). Several recent studies have clarified relationships within each conifer family (see Quinn et al., 2002).

III. Cycadales

The cycads are an order of seed plants consisting of ~300 long-lived, dioecious trees and shrubs (Hill et al., 2003). The order is defined by a number of synapomorphies, including coralloid roots, multilacunar nodes, girdling leaf traces, omega-shaped bundle patterns in the petioles, double vasculature of the integuments, and the presence of cycasins (Loconte and Stevenson, 1990). During the Mesozoic, cycads were widespread across both hemispheres (Stewart and Rothwell, 1993), but today they have a much narrower distribution in the Old and New World tropics, with centres of diversity in Mexico, South Africa, and northeast Australia. A widely used taxonomic scheme by Stevenson (1992) recognizes eleven genera and three families of cycads: Cycadaceae (one genus, *Cycas* L.), Stangeriaceae (two genera: *Stangeria* T. Moore and *Bowenia* Hook ex Hook. f.) and Zamiaceae (eight genera: *Ceratozamia* Brongn., *Chigua* D. W.

Stevenson, *Dioon* Lindl., *Encephalartos* Lehm., *Lepidozamia* Regel, *Microcycas* (Miq.) DC., *Macrozamia* Miq., and *Zamia* L.).

The fossil record of cycads extends back into the lower Permian (Mamay, 1969; Gao and Thomas, 1989), and many Tertiary cycad fossils can be assigned to modern genera, or even species. For example, fossils comparable to *Lepidozamia hopei* von Regel have been found that date to the Oligocene (Johnson, 1959). Stewart and Rothwell (1993) suggest that the cycads evolved during the Permian from medullosan pteridosperms (seed ferns). A close relationship between cycads and medullosan pteridosperms has also been noted by Crane (1985), who found a sister-group relationship between these two groups in his morphological cladistic analysis. However, the cladistic analysis by Rothwell and Serbet (1994) found medullosans to be quite distantly related to cycads, and Doyle and Donoghue (1992) suggest cycads may related to *Peltaspermum.*

Within cycads, both molecular (Treutlein and Wink, 2002; Hill et al., 2003; Rai et al., 2003; Bogler and Francisco-Ortega, in press) and morphological evidence (Stevenson, 1990) indicate that *Cycas* is the sister-group of the remaining taxa. This genus has been regarded as distinctive from other cycads, because its megasporophylls are leaf-like and indeterminate. *Dioon* is likely the sister group of the remaining cycads (Rai et al., 2003; Bogler and Francisco-Ortega, in press). Morphological evidence suggests that *Stangeria* and *Bowenia* form a clade that is the sister group of all cycads except *Cycas* (Stevenson, 1990), but molecular evidence indicates that the former two genera are not closely related to each other (Rai et al., 2003; Hill et al., 2003). A close relationship between *Ceratozamia, Zamia* and *Microcycas* is supported by morphological and molecular evidence, as is a close relationship between *Encephalartos, Lepidozamia*, and *Macrozamia* (Crane, 1988; Caputo et al., 1991;

Hill et al., 2003). However, these relationships are only moderately supported by bootstrap analyses in Hill et al. (2003).

IV. Ginkgoales

Ginkgo biloba L. is the sole remaining member of its order, Ginkgoales, which were at their peak diversity during the Mesozoic and the early Tertiary. At that time, they were a prominent part of the flora of the Northern Hemisphere (Stewart and Rothwell, 1993). *Ginkgo biloba* is a tree that combines pycnoxylic conifer-like wood with many cycad-like reproductive traits. These include its dioecy, monocolpate pollen grains, large motile sperm, haustorial pollen tubes, and a long period of free nuclear divisions during early embryogeny (Stewart and Rothwell, 1993).

V. Gnetales

Gnetales are a small and enigmatic order that are comprised of three very divergent families, each consisting of a single genus. These are: Ephedraceae (*Ephedra* L.), Gnetaceae (*Gnetum* L.) and Welwitschiaceae (*Welwitschia* J. D. Hooker). The genus *Ephedra* contains 35-45 species of mostly dioecious plants with scale- or needlelike leaves distributed throughout the arid regions of Eurasia, northern Africa and the Americas. The majority of *Ephedra* species are shrubs, although one species is a tree, and a few are climbers. There are approximately 30 species of *Gnetum*, which are tropical dioecious trees, shrubs or lianas distributed throughout west-central Africa, southern Central America,

northern South America, and south Asia. Species of *Gnetum* have broad, dicot-like leaves with reticulate venation. Welwitschiaceae contain one species, *Welwitschia mirabilis* J. D. Hooker, a long-lived, woody, dioecious plant with two very large, strap-like leaves. It is found only in the Namib desert of Namibia and southwestern Angola.

Although the members of this order are morphologically and ecologically distinct from each other, they are very well supported as a clade by both morphological (Crane, 1985; Doyle and Donoghue, 1986) and molecular data (Goremykin et al., 1996; Chaw et al., 1997; Chaw et al., 2000; Bowe et al., 2000). Traits found in all members of the group include: the presence of vessels in the secondary xylem, the presence of compound strobili, the presence of envelopes around the ovules and antherophores, and the presence of a micropylar projection of the integument which produces a pollination droplet (Kubitzki, 1990). The latter two traits are synapomorphies of the group (Judd et al., 2002). Within Gnetales, *Gnetum* and *Welwitschia* are more closely related to each other than either is to *Ephedra* (Hasebe et al., 1992; Crane, 1985; Doyle and Donoghue, 1986; Carlquist, 1996; Price, 1996). Morphological synapomorphies of the *Gnetum-Welwitschia* clade include: the lack of archegonia, a reduced number of cell divisions in the development of the microgametophyte, and a lack of a free-nuclear stage in the development of the embryo (Price, 1996).

The fossil record of Gnetales is quite sparse, particularly compared to that of other extant gymnosperm groups. A possible contributing factor to this is the difficulty of positively identifying Gnetales fossils (Crane, 1996). For example, the leaves of *Gnetum* are difficult to distinguish from dicot leaves, and the small leaves of *Ephedra* could be mistaken for conifer leaves (Crane, 1996). However, ribbed "ephedroid" pollen that characterizes many Gnetales taxa first appears in the fossil record during the Triassic and became quite diverse and common by

the early Cretaceous, before again becoming uncommon (Crane, 1996). Only a handful of Gnetales megafossils are known, most of which date from the Triassic to the Cretaceous (Crane, 1996; Rydin et al., 2003).

Objectives of the Thesis

A major goal of this thesis is to contribute to a clearer and more robust understanding of the relationships among the major groups of vascular plants, by sequencing a nuclear RNA polymerase II subunit gene, *RPB*2, for representatives of almost all of them (Chapter 2). Aside from 18S or 26S rDNA loci, few nuclear genes are used for inferring phylogenies at a deep level in vascular plants, in part because they are often present as multiple copies (Lawton-Rauh, 2003) potentially leading to misinference of gene orthology, due, for example, to undetected gene duplication and extinction events. *RPB*2, however, is single or double-copy in all vascular plants examined to date (Denton et al., 1998; Oxelman and Bremer, 2000; Oxelman et al., 2004), and since it is slowly evolving it is suitable for inferring phylogenies at a deep level (Denton et al., 1998). I also use a simulation approach, similar to the ones used by Huelsenbeck (1997) and Sanderson et al. (2000) to examine whether some of the major relationships found here were a possible artifact of long-branch attraction (Felsenstein, 1978; Hendy and Penny, 1989) or other sources of systematic error (bias). This is a concern in deep vascular-plant phylogenies, because there exist relatively few major extant lineages, each separated by a long time-depth, and some groups have very high rates of molecular evolution compared to others. The resulting long branches may often appear to be more closely related than they actually are (Felsenstein, 1978; Hendy and Penny, 1989).

I also aimed to resolve relationships within two major groups of vascular plants: the cycads (Chapter 3), and Liliales, an order of monocots (Chapter 4). Table 1.1 lists genera and families included in the studies presented in Chapters 2–4. Although good progress has been made in cycad phylogeny (see above), areas of substantial uncertainty still remain. Here I build on a recent study of cycad phylogeny that sampled a ~13.5 kb data set from the plastid genome (Rai et al., 2003), by adding representatives for three of the four remaining genera that are widely accepted. These new plastid data are combined with my new data from *RPB2* (Chapter 2), and with published data from the nuclear 26S rDNA locus.

I also collected a substantial new plastid data set (~17.0 kb per taxon) for the monocot order Liliales. In the pre-molecular era (e.g., Cronquist, 1988) this order served as a large dumping ground for many petaloid monocot taxa that have since been re-assigned to multiple, redefined monocot orders (e.g., Chase et al. 1995, 2000; APG II 2003). Liliales are a major branch of monocot phylogeny (e.g., Graham et al, submitted). My goal is to understand better this taxon's circumscription from a phylogenetic perspective, its higher-order relationships, and its placement deep in monocot phylogeny.

References

Amstutz, E. 1957. *Stylites*, a new genus of Isoetaceae. *Annals of the Missouri Botanical Garden* **44**: 121–213.

Angiosperm Phylogeny Group II. 2003. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG II. *Botanical Journal of the Linnean Society* **141**: 399–436.

Banks, H. P. 1968. The early history of land plants. In: *Evolution and Environment.* (E. T. Drake, ed.) Yale University Press, New Haven.

Bierhorst, D. W. 1971. *Morphology of Vascular Plants.* Macmillan, New York.

Bierhorst, D. W. 1977. The systematic position of *Psilotum* and *Tmesipteris*. *Brittonia* **29**: 3–13.

Bogler, D. J., and Francisco-Ortega, J. In Press. Molecular systematic studies in cycads: Evidence from *trn*L intron and ITS2 sequences. *Memoirs of the New York Botanical Garden.*

Bowe, L. M., Coat, G., and dePamphilis, C. W. 2000. Phylogeny of seed plant based on all three genomic compartments: Extant gymnosperms are monophyletic and Gnetales' closest relatives are conifers. *Proceedings of the National Academy of Sciences USA* **97:** 4092–4097.

Bremer, K., Humphries, C. J., Mishler, B. D., and Churchill, S. P. 1987. On cladistic relationships of green plants. *Taxon* **36**: 339–349.

Camus, 1990. Marattiaceae. Pp. 174–180. in: *The Families and Genera of Vascular Plants.* (K. U. Kramer and P. S. Green, eds. K. Kubitzki, series ed). Springer-Verlag, Berlin.

Caputo, P., Stevenson, D. W., and Wurtzel, E. T. 1991. A phylogenetic analysis of American Zamiaceae (Cycadales) using chloroplast DNA restriction fragment length polymorphisms. *Brittonia* **43**: 135–145.

Carlquist, S. 1996. Wood, bark and stem anatomy of Gnetales: A summary. *International Journal of Plant Sciences* **157**: S58–S76.

Chase, M. W., Soltis, D. E., Olmstead, R. G., Morgan, D., Les, D. H., Mishler, B., Duvall. M. R., Price, R. A., Hills, H. G., Qiu, Y-L., Kron, K. A., Rettig, J. H., Conti, E. Palmer, J. D., Manhart, J. R., Systma, K. J., Michaels, H. J., Kress, W. J., Karol, K. G., Clark, W. D., Hedren, M., Gaut, B. S., Jansen, R. K., Kim, K-J., Wimpee, C. F., Smith, J. F., Soltis, P. S., Swenson, S. M., Williams, S. E., Gadek, P. A., Quinn, C. J., Eguiarte, L., Golenburg, E., Learn, G. H., Graham, S. W., Barrett, S. C. H., Dayanandan, S., and Albert, V. A. 1993. Phylogenetics of seed plants: an analysis of nucleotide sequences from the chloroplast gene *rbc*L. *Annals of the Missouri Botanical Garden* **80**: 528–580.

Chase, M. W., Stevenson, D. W., Wilkin, P., and Rudall, P. J. 1995. Monocot systematics: a combined analysis. Pp. 685–730 in: *Monocotyledons: Systematics and Evolution.* (P. J. Rudall, P. J. Cribb, D. F. Cutler, & C. J. Humphries. eds.) Royal Botanic Gardens, Kew.

Chase, M. W., Soltis, D. E., Soltis, P. S., Rudall, P. J., Fay, M. F., Hahn, W. H., Sullivan, S., Joseph, J., Molvray, M., Kores, P. J., Givnish, T. J., Sytsma, J., and Pires, J. C. 2000. Higher-level systematics of the monocotyledons: an assessment of current knowledge and a new classification. in: *Monocots: Systematics and Evolution.* (K. L. Wilson and D. A. Morrison. eds.) Royal Botanic Gardens, Kew.

Chaw, S-M., Parkinson, C. L., Cheng, Y., Vincent, T. M., and Palmer, J. D. 2000. Seed plant phylogeny inferred from all three plant genomes: Monophyly of extant gymnosperms and origin of Gnetales from conifers. *Proceedings of the National Academy of Sciences USA.* **97:** 4086–4091.

Chaw, S., Zharkikh, A., Sung, H., Lau, T., and Li, W. 1997. Molecular phylogeny of extant gymnosperms and seed plant evolution: Analysis of 18S rRNA sequences. *Molecular Biology and Evolution* **14**: 56–68.

Conran, J. G., Wood, G. A., Martin, P. G., Dowd, J. M., Quinn, C. J., Gadek, P. A., and Price, R. A. 2000. Generic relationships within and between the gymnosperm families Podocarpaceae and Phyllocladaceae based on an analysis of the chloroplast gene *rbc*L. *Australian Journal of Botany* **48:** 715–724.

Crane, P. R. 1985. Phylogenetic relationships in seed plants. *Cladistics* **1**: 329–348.

Crane, P. R. 1988. Major clades and relationships in the "higher" gymnosperms. Pp. 218–272. in: Origin and Evolution of Gymnosperms. (C. B. Beck, ed.) Columbia University Press, New York.

Crane, P. R. 1996. The fossil history of the Gnetales. *International Journal of Plant Sciences* **157**: S50–S57.

Crane, P. R., Friis, E. M., and Pederson, K. R. 1995. The origin and early diversification of angiosperms. *Nature* **374**: 27–33.

Cronquist, A. 1988. *The Evolution and Classification of Flowering Plants, 2nd edition.* The New York Botanical Garden, Bronx, New York.

Denton, A. L., McConaughy, B. L., and Hall, B. D. 1998. Usefulness of RNA polymerase II coding sequences for estimation of green plant phylogeny. *Molecular Biology and Evolution* **15**: 1082–1085.

Donoghue, M. J., and Doyle, A. D. 2000. Seed plant phylogeny: Demise of the anthophye hypothesis*?* *Current Biology* **10**: R106–R109.

Doyle, J. A. 1996. Seed plant phylogeny and the relationships of Gnetales. *International Journal of Plant Sciences* **157**: S3–S9.

Doyle, J. A., 1998. Phylogeny of vascular plants. *Annual Review of Ecology and Systematics* **29**: 567–599.

Doyle, J. A., and Donoghue, M. J. 1986. Seed plant phylogeny and the origin of angiosperms: an experimental cladistic approach. *Botanical Reviews* **52**: 321–431.

Doyle, J. A., and Donoghue, M. J. 1992. Fossils and seed plant phylogeny reanalyzed. *Brittonia* **44:** 89–106.

Duff, R. J. and Nickrent, D. L. 1999. Phylogenetic relationships of land plants using mitochondrial small-subunit rDNA sequences. *American Journal of Botany* **86**: 372–386.

Eckenwalder, J. E. 1976. Re-evaluation of Cupressaceae and Taxodiaceae: A proposed merger. *Madrono* **23**: 237–300.

Farjon, A. 1998. *World Checklist and Bibliography of Conifers.* Richmond, U.K.: Royal Botanical Gardens at Kew.

Felsenstein, J. 1978. Cases in which parsimony or compatibility methods will be positively misleading. *Systematic Zoology* **27:** 401–410.

Felsenstein, J. 1985. Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* **39**: 783–791.

Felsenstein, J. 2004. *Inferring Phylogenies.* Sinauer Associates, Inc. Sunderland, MS.

Florin, R. 1951. Evolution of cordaites and conifers. *Acta Horti Bergiani* **15**: 285–388.

Gao, Z., and Thomas, B. A. 1989 A review of fossil cycad megasporophylls, with new evidence of *Crossozamia* Pomel and its associated leaves from the Lower Permian of Taiyuan, China. *Review of Palaeobotany and Palynology.* **60**: 205–223.

Garbary, D. J., Renzaglia, K. S., and Duckett, J. G. 1993. The phylogeny of land plants – A cladistic analysis based on male gametogenesis. *Plant Systematics and Evolution* **188**: 237–269.

Garratt, M. J., Tims, J. D., Rickards, R. B., Chambers, R. C., and Douglas, J. G. 1984. The appearance of *Baragwanathia* (Lycophytina) in the Silurian. *Botanical Journal of the Linnean Society* **89**: 355–355.

Gensel, P. G. 1977. Morphologic and taxonomic relationships of the Psilotaceae relative to evolutionary lines in early land vascular plants. *Brittonia* **29**: 14–29.

Gensel, P.G. 1992. Phylogenetic relationships of the zosterophylls and lycopsids: Evidence from morphology, paleoecology, and cladistic methods of inference. *Annals of the Missouri Botanical Garden* **79**: 450–473.

Goremykin, V., Bobrova, B., Pahnke, J., Troitsky, A., Antonov, A., and Martin, W. 1996. Noncoding sequences for the slowly evolving chloroplast inverted repeat do not support gnetalean affinities of angiosperms. *Molecular Biology and Evolution* **13**: 383–396.

Graham, S. W., and Olmstead, R. G. 2000. Utility of 17 plastid genes for inferring the phylogeny of the basal angiosperms. *American Journal of Botany* **87**: 1712–1730.

Graham, S. W., Zgurski, J. M., McPherson, M. A., Cherniawsky, D. M., Saarela, J. M., Horne, E. S. C., Smith, S. Y., Wong, W. A., O'Brien, H. E., Biron, V. L., Pires, J. C., Olmstead, R. G., Chase, M. W., and Rai, H. S. Submitted. Robust inference of monocot deep phylogeny using an expanded multigene plastid data set.

Grierson, J. D., and Bonamo, P. M. 1979. *Leclercqia complexa:* earliest ligulate lycopod (Middle Devonian). *American Journal of Botany* **66**: 474–476.

Gurgerli, F., Sperisen, C., Buchler, U., Brunner, I., Brodbeck, S., Palmer, J. D., and Qiu, Y. L. 2001. The evolutionary split of Pinaceae from other conifers: Evidence from an intron loss and a multigene phylogeny. *Molecular Phylogenetics and Evolution* **21**: 167–175.

Hart, J. A. 1987. A cladistic analysis of conifers: Preliminary results. *Journal of the Arnold Arboretum* **68**: 269–307.

Hasebe, M., Ito, M., Kofugi, R., Iwatsuki, K., and Ueda, K. 1992. Phylogenetic relationships in Gnetophyta deduced from *rbc*L gene sequences. *Botanical Magazine of Tokyo* **105**: 385–391.

Hasebe, M., Omori, T., Nakazawa, M., Sano, T., Kato, M., and Iwatsuki, K. 1994. *rbc*L gene sequences provide evidence for the evolutionary lineages of leptosporangiate ferns. *Proceedings of the National Academy of Sciences USA*. **91**: 5730–5734.

Hasebe, M., Wolf, P. G., Pryer, K. M., Ueda, K., Ito, M., Sano, R., Gastony, G. J., Yokoyama, J., Manhart, J. R., Murakami, N., Crane, E. H., Haufler, C. H., and Hauk, W. D. 1995. Fern phylogeny based on *rbc*L nucleotide sequences. *American Fern Journal* **85**: 134–181.

Hauke, R. L. 1990. Equisetaceae. Pp. 46–48. In: Kramer, K. U., and Green, P. S., eds. *The Families and Genera of Vascular Plants, Vol. 1: Pteridophytes and Gymnosperms*. (series ed. K. Kubitzki) Springer-Verlag, Berlin.

Hendy, M. D., and Penny, D. 1989. A framework for the quantitative study of evolutionary trees. *Systematic Zoology* **38**: 297–309.

Hernandez-Castillo, G. R., Rothwell, G. W., and Mapes, G. 2001. Compound pollen cone in a Paleozoic conifer. *American Journal of Botany* **88**: 1139-1142.

Hill, K. D., Chase, M. W., Stevenson, D. W., Hills, H. G., and Schutzman, B. 2003. The families and genera of cycads: A molecular phylogenetic analysis of cycadophyta based on nuclear and plastid DNA sequences. *International Journal of Plant Sciences* **164**: 933–948.

Huelsenbeck, J. P. 1997. Is the Felsenstein zone a fly trap? *Systematic Biology* **46**: 69-74.

Jermy, A. C. 1990. Isoetaceae. Pp. 26–31 In: Kramer, K. U., and Green, P. S., eds. 1990. *The Families and Genera of Vascular Plants, Vol. 1: Pteridophytes and Gymnosperms*. (series ed. K. Kubitzki). Springer-Verlag, Berlin.

Johnson, L. A. S. 1959. The families of cycads and the Zamiaceae of Australia. *Proceedings of the Linnean Society of New South Wales* **84**: 64–117.

Judd, W. S., Campbell, C. S., Kellogg, E. A., Stevens, P. F., and Donoghue, M. J. 2002. *Plant Systematics: A Phylogenetic Approach, second edition.* Sinauer Associates, Sunderland, MA.

Kelch, D. G. 2002. Phylogenetic assessment of the monotypic genera *Sundacarpus* and *Manoao* (Coniferales: Podocarpaceae) utilizing evidence from 18S rDNA sequences. *Australian Systematic Botany* **15:** 29–35.

Kenrick, P., and Crane, P. R. 1997. *The Origin and Early Diversification of Land Plants: A Cladistic Study.* Smithsonian Institution Press, Washington.

Kramer, K. U. 1990. Psilotaceae. Pp. 22–25 In: Kramer, K. U., and Green, P. S., eds. *The Families and Genera of Vascular Plants, Vol. 1: Pteridophytes and Gymnosperms.* (series ed. K. Kubitzki). Springer-Verlag, Berlin.

Kramer, K. U., and Green, P. S., eds. 1990. *Pteridophytes and Gymnosperms.* In*: The Families and Genera of Vascular Plants, Vol. 1.* (series ed. K. Kubitzki). Springer-Verlag, Berlin.

Kotyk, M. E., Basinger, J. F., Gensel, P. G., and De Freitas, T. A. 2002. Morphologically complex plant macrofossils from the late Silurian of arctic Canada. *American Journal of Botany* **89**: 1004–1013.

Kubitzki, K. 1990. General traits of the Gnetales. Pp. 378–379. in: *The Families and Genera of Vascular Plants.* (K. U. Kramer and P. S. Green, eds. K. Kubitzki, series ed). Springer-Verlag, Berlin.

Lawton-Rauh, A. 2003. Evolutionary dynamics of duplicated genes in plants. *MolecularPhylogenetics and Evolution* **29:** 396–409.

Loconte, H., and Stevenson, D. W. 1990. Cladistics of the Spermatophyta. *Brittonia* **42**: 197–211.

Magallon, S., and Sanderson, M. J. 2002. Relationships among seed plants inferred from highly conserved genes: Sorting conflicting phylogenetic signals among ancient lineages. *American Journal of Botany* **89**: 1991–2006.

Mamay, S. H. 1969. Cycads: fossil evidence of late Paleozoic origin. *Science* **164**: 295–296.

Manhart, J. R. 1995. Chloroplast 16S rDNA sequences and phylogenetic relationships of fern allies and ferns. *American Fern Journal* **85**: 182–192.

McPherson, 2003. *Phylogenetics of Asparagales (Liliopsida) as inferred from a large plastid DNA sequence data set.* M.Sc. Thesis. University of Alberta, Edmonton, Alberta, Canada.

Mickel, J. T. 1974. Phyletic lines in modern ferns. *Annals of the Missouri Botanical Garden* **61**: 474–482.

Miller, C. N. Jr. 1999. Implications of fossil conifers for the phylogenetic relationships of living families. *The Botanical Review* **65**: 239–262.

Oxelman, B., and Bremer, B. 2000. Discovery of paralogous nuclear gene sequences coding for the second-largest subunit of RNA polymerase II (*RPB*II) and their phylogenetic utility in Gentianales of the asterids. *Molecular Biology and Evolution* **17**: 1131–1145.

Oxelman B., Yoshikawa, N., McConaughy, B. L., Luo, J., Denton, A. L., and Hall, B. D. 2004. *RPB*2 gene phylogeny in flowering plants, with particular emphasis on asterids. *Molecular Phylogenetics and Evolution* **32**: 462-479.

Page, C. N. 1990. Taxaceae. Pp. 348–353. In: Kramer, K. U., and Green, P. S., eds. *The Families and Genera of Vascular Plants, Vol. 1: Pteridophytes and Gymnosperms.* (series ed. K. Kubitzki). Springer-Verlag, Berlin

Parenti, L. R. 1980. A phylogenetic analysis of the land plants. *Biological Journal of the Linnean Society* **13:** 225–242.

Phipps, C. J., Taylor, T. N., Taylor, E. L., Cuneo, R., Boucher, L. D., and Yao, X. 1998. *Osmunda* (Osmundaceae) from the Triassic of Antarctica: An example of evolutionary stasis. *American Journal of Botany* **85**: 888–895.

Price, R. A. 1996. Systematics of Gnetales: A review of morphological and molecular evidence. *International Journal of Plant Science* **157**: S40–S49.

Pryer, K.M., Smith, A. R. and Skog, J. E. 1995. Phylogenetic relationships of extant ferns based on evidence from morphology and *rbc*L sequences. *American Fern Journal* **85**: 205–282.

Pryer, K. M. 1999. Phylogeny of marsileaceous ferns and relationships of the fossil *Hydropteris pinnata* reconsidered. *International Journal of Plant Science* **160**: 931–954.

Pryer, K. M., Schneider, H., Smith, A. R., Cranfill, R., Wolf, P. G., Hunt, J. S., and Sipes, S. D. 2001. Horsetails and ferns are a monophyletic group and the closest living relatives to seed plants. *Nature* **409**: 618–622.

Qiu, Y. L., Lee, J., Bernasconi-Quadroni, F., Soltis, D. E., Soltis, P. S., Zanis, M., Zimmer, E. A., Chen, Z., Savolainen, V., and Chase, M. W. 1999. The earliest angiosperms: evidence from mitochondrial, plastid and nuclear genomes. *Nature* **402**: 404–407.

Qiu, Y. L., Lee, J., Bernasconi-Quadroni, F., Soltis, D. E., Soltis, P. S., Zanis, M., Zimmer, E. A., Chen, Z., Savolainen, V., Chase, M. W. 2000. Phylogeny of basal angiosperms: Analyses of five genes from three genomes. *International Journal of Plant Sciences* **161**: S3–S27.

Quinn, C. J., Price, R. A., and Gadek, P. A. 2002. Familial concepts and relationships in the conifers based on *rbc*L and *mat*K sequence comparisons. *Kew Bulletin* **57**: 513–531.

Rai, H. S., O'Brien, H. E., Reeves, P. A., Olmstead, R. G., and Graham, S. W. 2003. Inference of higher-order relationships in the cycads from a large chloroplast data set. *Molecular Phylogenetics and Evolution* **29**: 350–359.

Raubeson, L. A., and Jansen, R. K. 1992. Chloroplast DNA evidence on the most ancient split in vascular land plants. *Science* **85**: 1697–1699.

Raubeson, L. A. and Stein, D. B. 1995. Insights into fern evolution from mapping chloroplast genomes. *American Fern Journal* **85**: 193–120.

Rothwell, G. W. 1999. Fossils and ferns in the resolution of land plant phylogeny. *Botanical Review* **65**: 188–218.

Rothwell, G.W. and R. Serbet. 1994. Lignophyte phylogeny and the evolution of spermatophytes: A numerical cladistic analysis. *Systematic Botany* **19**: 443–482.

Rothwell, G. W., and Stockey, R. A. 1989. Fossil Ophioglossales in Paleocene of western North America. *American Journal of Botany* **76**: 637–644.

Rothwell, G. W., and Stockey, R. A. 1994. The role of *Hydropteris pinnata* gen. et. sp. nov. in reconstructing the cladistics of heterosporous ferns. *American Journal of Botany* **81**: 479–492.

Rydin, C., Kallersjo, M., Friis, E. M. 2002. Seed plant relationships and the systematic position of Gnetales based on nuclear and chloroplast DNA: conflicting data, rooting problems and the monophyly of conifers. *International Journal of Plant Science* **163**: 197–214.

Rydin, C., Mohr, B., and Friis, E. M. 2003. *Cratonia* cotyledon gen. et sp. nov.: A unique Cretaceous seedling related to *Welwitschia*. *Proceedings of the Royal Society of London - Biological Sciences* **270**: S29–S32.

Sanderson, M. J., Wojciechowski, M. F., Hu, J. M., Sher Khan, T., and Brady, S. G. 2000. Error, bias and long-branch attraction in data for two chloroplast photosystem genes in seed plants. *Molecular Biology and Evolution* **17**: 782–797.

Savolainen, V., and Chase, M. W. 2003. A decade of progress in plant molecular phylogenetics. *Trends in Genetics* **19**: 717–724.

Savolainen, V., Fay, M. F., Albach, D. C., Backlund, A., Van der Bank, M., Cameron, K. M., Johnson, S. A., Lledo, M. D., Pintaud, J. C., Powell, M., Sheahan, M. C., Soltis, D. E., Soltis, P. S., Weston, P., Whitten, W. M., Wurdack, K. J., Chase, M. W. 2000. Phylogeny of the eudicots: A nearly complete familial analysis based on *rbc*L gene sequences. *Kew Bulletin* **55**: 257–309.

Schmidt, M. and Schneider-Poetsch, H. A. W. 2002. The evolution of gymnosperms redrawn by phytochrome genes: The Gnetatae appear at the base of the gymnosperms. *Journal of Molecular Evolution* **54**: 715–724.

Sinclair, W. T., Mill, R. R., Gargner, M. F., Woltz, P., Jaffré, T., Preston, J., Hollingsworth, M. L., Ponge, A., and Möller, M. 2002. Evolutionary relationships of the New Caledonian heterotrophic conifer, *Parasitaxus usta* (Podocarpaceae), inferred from chloroplast *trn*L-F intron/spacer and nuclear rDNA ITS2 sequences. *Plant Systematics and Evolution* **233**: 79–104.

Smith, A. R. 1995. Non-molecular phylogenetic hypotheses for ferns. *American Fern Journal* **85**: 104–122.

Serbet, R., and Rothwell, G. W. 1999. *Osmunda cinnamomea* (Osmundaceae) in the upper Cretaceous of western North America: Additional evidence for exceptional species longevity among filicalean ferns. *International Journal of Plant Science* **160**: 425–433.

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

Soltis, D. E., Soltis, P. S., Chase, M. W., Mort, M. E., Albach, D. C., Zanis, M., Savolainen, V., Hahn, W. H., Hoot, S. B., Fay, M.F., Axtell, M., Swensen, S. M., Prince, L. M., Kress, W. J., Nixon, K. C., and Farris, J. S. 2000. Angiosperm phylogeny inferred from 18S rDNA, *rbc*L, and *atp*B sequences. *Botanical Journal of the Linnean Society* **133**: 381–461.

Soltis, D. E., Soltis, P. S., and Zanis, M. J. 2002. Phylogeny of seed plants based on evidence from eight genes. *American Journal of Botany* **89**: 1670–1681.

Stefanovic, S. M.. Jager, J., Deutsch, J. Broutin, J., and Masselot, M. 1998. Phylogenetic relationships of conifers inferred from partial 28S rRNA gene sequences. *American Journal of Botany* **85**: 688–697.

Stevenson, D. W. 1992. A formal classification of the extant cycads. *Brittonia* **44**: 220–223.

Stevenson, D. W. 1990. Morphology and systematics of the Cycadales. *Memoirs of the New York Botanical Garden* **57**: 8–55.

Stewart, W. N., and Rothwell, G. W. 1993. *Paleobotany and the Evolution of Plants, second edition.* Cambridge University Press, Cambridge, United Kingdom.

Thieret, J. W. 1993. Pinaceae. in: *Flora of North America North of Mexico, vol 2.* Flora of North America Editorial Committee, eds. New York, NY.

Tidwell, W. D., and Ash, S. R. 1994. A review of selected Triassic to early Cretaceous ferns. *Journal of Plant Research* **107**: 417–442.

Treutlein, J., and Wink, M. 2002. Molecular phylogeny of cycads inferred from *rbc*L sequences. *Naturwissenschaften* **89**: 221–225.

Wagner, W. H. Jr. 1977. Systematic implications of the Psilotaceae. *Brittonia* **29**: 54–63.

Wagner, W. H. Jr. 1990. Ophioglossaceae. Pp. 193–197. In: Kramer, K. U., and Green, P. S., eds. 1990. *The Families and Genera of Vascular Plants, Vol. 1: Pteridophytes and Gymnosperms.* (series ed. K. Kubitzki). Springer-Verlag, Berlin.

Wagner, W. H., and Beitel, J. M. 1992. Generic classification of modern North American Lycopodiaceae. *Annals of the Missouri Botanical Garden* **79**: 676–686.

Wolf, P. G. 1995. Phylogenetic analyses of *rbcL* and nuclear ribosomal RNA gene sequences in Dennstaedtiaceae. *American Fern Journal* **85:** 306–327.

Wolf, P. G. 1997. Evaluation of *atpB* nucleotide sequences for phylogenetic studies of ferns and other pteridophytes. *American Journal of Botany* **84**: 1429–1440.

Zurawski, G., Perrot, B., Bottomley, W., Whitfeld, P. R. 1981. The structure of the gene for the large subunit of ribulose 1,5 bisphosphate carboxylase from spinach chloroplast DNA. *Nucleic Acids Research* **9**: 3251–3270.

Table 1.1. An overview of vascular-plant classification, listing plant families and genera included in this study. Angiosperm classification follows APG II (2003), cycad classification follows Stevenson (1992) and classification for all other groups follows Kramer and Green (1990), except all Taxodiaceae are placed in Cupressaceae *s. l.* as recommended by Eckenwalder (1976) and Stefanovic et al. (1998). Taxales are not recognized and are placed in Coniferales (Stefanovic et al., 1998). Taxa marked with an asterisk belong to the order Liliales, and are the main focus of Chapter 4 of this thesis. Most non-Liliales monocots included in Chapter 4 are not included here.

Tracheophyta (vascular plants)

Lycopodiales – lycopods: club mosses, quillworts and relatives

Isoetaceae	*Isoetes*
Lycopodiaceae	*Lycopodium*
Selaginellaceae	*Selaginella*

Euphyllophytina – moniliforms & seed plants

Moniliformopses – moniliforms

Equisetales – horsetails

Equisetaceae	*Equisetum*

Marattiales, Ophioglossales – eusporangiate ferns

Ophioglossaceae	*Botrychium, Ophioglossum*

Polypodiales – leptosporangiate ferns

Blechnaceae	*Stenochlaena*
Dicksoniaceae	*Dicksonia*
Dryopteridaceae	*Dryopteris, Polystichum*
Marsileaceae	*Marsilea*
Pteridaceae	*Adiantum, Cryptogramma*
Schizaeaceae	*Anemia*

Psilotales – whisk ferns

Psilotaceae	*Psilotum*

Seed plants

Gymnospermopsida – gymnosperms

Coniferales – conifers

Araucariaceae	*Agathis, Araucaria*
Cephalotaxaceae	*Cephalotaxus*
Cupressaceae	*Cunninghamia, Juniperus, Metasequoia, Sequoia, Sequoiadendron, Thuja*
Pinaceae	*Cedrus, Larix, Pinus, Pseudotsuga*
Podocarpaceae	*Dacrydium, Podocarpus, Saxegothaea*
Sciadopityaceae	*Sciadopitys*
Taxaceae	*Taxus, Torreya*

Cycadales – cycads

Cycadaceae	*Cycas*
Stangeriaceae	*Bowenia, Stangeria*
Zamiaceae	*Ceratozamia, Dioon, Encephalartos, Lepidozamia, Microcycas, Macrozamia, Zamia.*

Ginkgoales

Ginkgoaceae	*Ginkgo*

Gnetales

Ephedraceae	*Ephedra*
Gnetaceae	*Gnetum*
Welwitschiaceae	*Welwtischia*

Angiosperms – flowering plants

Amborellaceae	*Amborella*
Aristolochiaceae	*Aristolochia, Asarum*
Austrobaileyaceae	*Austrobaileya*
Ceratophyllaceae	*Ceratophyllum*
Chloranthaceae	*Chloranthus*
Illicaceae	*Illicium*
Magnoliaceae	*Magnolia*
Nymphaeaceae	*Cabomba, Nymphaea*
Piperaceae	*Peperomia*

Angiosperms – flowering plants, contd.

Trimeniaceae	*Trimenia*
Winteraceae	*Drimys*

Monocots

Acoraceae	*Acorus*
Alstroemeriaceae*	*Alstroemeria*
Campynemataceae*	*Campynema*
Colchicaceae*	*Petermannia, Tripladenia, Wurmbea*
Dioscoreaceae	*Dioscorea*
Liliaceae*	*Calochortus, Lilium, Medeola, Prosartes, Tricyrtis*
Luzuriagaceae*	*Luzuriaga*
Melanthiaceae*	*Anticlea, Trillium*
Musaceae	*Musa*
Philesiaceae*	*Philesia*
Poaceae	*Hordeum*
Rhipogonaceae*	*Rhipogonum*
Smilacaceae*	*Smilax*

Eudicots

Amaranthaceae	*Spinacea*
Brassicaceae	*Arabidopsis*
Caryophyllaceae	*Dianthus*
Plantanaceae	*Plantanus*
Ranunculaceae	*Hydrastis*
Solanaceae	*Solanum*
Vitaceae	*Vitis*

~Chapter Two~

Inference of Deep Vascular-Plant Phylogeny Using an RNA Polymerase Subunit Gene.

Introduction

One of the most challenging problems faced by plant systematists relates to the resolution of phylogenetic relationships among extant vascular plant groups, which are represented by five major groups of seed plants and various seedless (free-sporing) vascular plants, the latter also referred to as "pteridophytes" or the "ferns and fern allies." The extant pteridophytes are often divided into six orders: lycopods (Lycopodiales), horsetails (Equisetales), whisk ferns (Psilotales), eusporangiate ferns (Ophioglossales and Marattiales) and leptosporangiate ferns (Polypodiales); see Table 1.1 in Chapter 1. The extant seed plants comprise the angiosperms (flowering plants) and four groups of gymnosperms: Cycadales, Coniferales, Gnetales, and *Ginkgo biloba*.

A large number of molecular studies completed recently have aimed to resolve phylogenetic relationships among these lineages, and many have come to similar conclusions about some deep vascular-plant relationships. For example, the majority of molecular studies (Raubeson and Jansen, 1992a; Duff and Nickrent, 1999; Pryer et al., 2001; Rydin et al., 2002) have pointed to the lycopods as being the sister group of the remaining vascular plants (= the euphyllophytes). This result is congruent with morphological cladistic studies (Kenrick and Crane, 1997). Additionally, molecular studies have largely pointed to a single answer regarding the phylogenetic position of the bigeneric Psilotaceae

(Psilotales). This family is very well-supported as being the sister-taxon of Ophioglossaceae, a family of eusporangiate ferns (Manhart, 1995; Wolf, 1997; Pryer et al., 2001; Rydin et al., 2002). However, morphological studies suggest an alternative placement for Psilotaceae. Bierhorst (1971, 1977) argues that the family is closely related to *Stromatopteris moniliformis* Mett., a leptosporangiate fern in the family Gleicheniaceae. He based his argument mainly on observations of gametophyte and embryo morphologies. In contrast, Wagner (1977) argued that Psilotaceae are different from all fern groups, and evolved as a lineage distinct from them within the vascular plants. Cladistic analyses based on morphological characters instead place Psilotaceae as the sister group of all remaining euphyllophytes (Stevenson and Loconte, 1996; Rothwell, 1999).

The ferns are by far the dominant group of seedless vascular plant, with approximately 260 genera containing ~12,000 species (e.g., Judd et al., 2002). They are typically divided into two groups (based on sporangial development): the eusporangiate and leptosporangiate ferns. The former have thick-walled sporangia that arise from multiple initial cells, a condition that is also found in all other extant seedless plants (lycopods, Equisetaceae and Psilotaceae). In contrast, leptosporangiate ferns typically have sporangia that arise from a single initial cell, and have walls that are a single cell thick. In addition, leptosporangiate ferns may have an annulus, but eusporangiate ferns never do, when only extant taxa are considered. According to molecular analyses, these two groups, along with Psilotaceae and possibly Equisetaceae, form a clade that is the sister group of seed plants (Pryer et al., 2001; Rydin et al., 2002). This clade has been called the "moniliforms" by some authors (e.g., Kenrick and Crane, 1997). Within this group, the leptosporangiate ferns are well supported as a clade (Manhart, 1995; Duff and Nickrent, 1999; Rothwell, 1999; Pryer et al., 2001), Psilotaceae are the

sister group of Ophioglossaceae, and Equisetaceae appear to be the sister group of Marattiaceae, the second family of eusporangiate ferns. However, the latter result receives only weak support from bootstrap analysis (Pryer et al., 2001; Rydin et al., 2002) and deserves further investigation.

The inference of phylogenetic relationships among the five extant seed plant groups is also problematic. In particular, the position of Gnetales, a diverse and enigmatic group composed of only three extant genera (*Ephedra, Gnetum* and *Welwitschia*), remains controversial. Analyses of morphological data (Crane, 1985; Doyle and Donoghue, 1986, 1992; Loconte and Steveson, 1990; Nixon et al., 1994; Rothwell and Serbet, 1994; Doyle, 1996) have placed them in a clade with the angiosperms and several extinct gymnosperm groups (Bennetitiales and *Pentoxylon*). The notion that these groups belong to the same clade is often referred to as the "anthophyte" hypothesis, as it implies that they evolved from a common ancestor with flower-like reproductive organs. However, molecular studies have yet to provide any strong support for the anthophyte hypothesis. These studies often find Gnetales to be more closely related to conifers than to angiosperms, and find the extant gymnosperms to be monophyletic (Goremykin et al., 1996; Chaw et al., 1997; Winter et al., 1999; Nickerson and Drouin, 2004). More recent studies that consider multiple genes from all three genomes suggest that Gnetales are nested within conifers, as the sister group of Pinaceae (Bowe et al., 2000; Chaw et al., 2000), the so-called "gnepine" hypothesis. Studies using even larger numbers of taxa or more sequence data provide support for a "Gnetales basal" topology, which has Gnetales as the sister group of all remaining seed plants, including angiosperms (Magallon and Sanderson, 2002; Rydin et al., 2002; Rai et al., 2003).

Molecular studies of vascular-plant relationship thus often differ dramatically, depending upon the tree-building methods used to generate the trees, the gene(s) used, and the ingroup and outgroup taxa included. The same basic data set can also provide support for conflicting relationships, depending upon which data partitions are considered. For example, strongly conflicting seed-plant relationships have been inferred using first and second vs. third codon positions (Sanderson et al., 2000; Magallon and Sanderson, 2002; Soltis et al., 2002), or using transitions vs. transversions (Rydin et al., 2002). This may be because nucleotide substitutions in the first two codon positions are predominantly nonsynonymous, whereas those in the third position are predominantly synonymous (e.g., Sanderson et al., 2000), and the latter evolve substantially faster than the former. Transitions are also observed to occur substantially more frequently than transversions in plastid data sets (e.g., Rai et al., 2003).

The inference of relationships among vascular plants has been problematic for a variety of reasons. One potential complication is that parameters in the substitution model may vary substantially among taxa (e.g., Rai et al., 2003). Whether this contributes to misinference of vascular-plant phylogeny is not known, although Rai et al. (2003) showed that it is unlikely to influence phylogenetic inference within cycads, at least. Another major problem may be that there is a mixture of very long and short branches deep in vascular-plant phylogeny. At least some deep, short internodes may represent rapid early radiations (in the seed plants, for example; Donoghue and Doyle, 2000). Many of the long branches in vascular-plant phylogeny are likely a simple function of a deep time-depth, coupled with a relatively low survival rate among all of the major lineages known from the fossil record. For example, the "crown" seed-

plant clade (the extant lineages plus all descendants of their most recent common ancestor) are ~325 Myr old, but only five major lineages are still alive, a small sample of the total seed-plant diversity that once existed (e.g., Stewart and Rothwell, 1993; Kenrick and Crane, 1997). Additionally, the basic substitution rate can differ substantially among major vascular-plant lineages, with Gnetales and ferns having a relatively high substitution rate in plastid genes (Sanderson et al., 2000), and cycads and *Ginkgo* having a substantially lower rate (Rai et al., 2003) relative to all other groups. This rate heterogeneity also contributes to branch-length differences among vascular-plant lineages. Felsenstein (1978) and Hendy and Penny (1989) demonstrated that maximum parsimony may favour an incorrect tree if a phylogeny is composed of a mix of long and short branches. This phenomenon is often referred to as "long-branch attraction" (Hendy and Penny, 1989), because long branches can appear to be more closely related than they actually are. Long-branch attraction may also result in mis-rooted trees, if fast-evolving ingroup taxa are erroneously attracted to highly divergent outgroup taxa (e.g., Graham et al., 2002). Both Rydin et al. (2002) and Rai et al. (2003) suggested that the "Gnetales basal" topology recovered in their studies could possibly be the result of long-branch attraction between Gnetales and free-sporing outgroups, taxa with relatively high rates of evolution.

Long-branch attraction can be difficult to detect or quantify for real data sets, because the correct tree is not known in advance. However, Sanderson et al. (2000) used a simulation approach to survey a data set containing two plastid photosystem genes for 15 seed plants and seven free-sporing plants for potential long-branch attraction problems. Trees were constructed that conformed to three hypotheses of seed-plant relationships, and various molecular-evolution rate parameters were estimated for each tree. This information was used to simulate

multiple data sets for each hypothesis, and maximum-parsimony (MP) trees were then inferred for each simulated data set. Often, the MP trees recovered did not match the tree used to simulate the sequences, particularly if the tree used to simulate the data sets corresponded to the anthophyte hypothesis. The latter implies that, if the anthophyte hypothesis were correct, their molecular data would not recover it reliably (i.e., high type I error). Depending on the details of the model trees used for simulations, one hypothesis ("Gnetales basal in seed plants") often had a high probability of being recovered when incorrect (i.e., a high type II error rate for that hypothesis).

There are several possible solutions to potential long-branch attraction. One of the most effective may be to add more taxa to phylogenies, to effectively break up those branches that may interfere with accurate phylogenetic reconstruction (e.g., Hillis, 1996; Zwickl and Hillis, 2002). However, this strategy can also introduce new problematic long branches (e.g., Rannala et al., 1998; Poe and Swofford, 1999), and because many important vascular-plant lineages are extinct (and cannot be included in molecular phylogenies), there often exist no extant taxa that could be used to break up long branches. Using methods of phylogenetic inference that may be less prone to long-branch attraction, such as maximum likelihood is another option, although likelihood can still be prone to the problem, if an incorrect model of evolution is assumed (Huelsenbeck, 1995; Chang, 1996; Sullivan and Swofford, 1997). Another potential strategy is to use conservatively evolving characters, such as nucleotide data from slowly evolving genes or amino-acid translations of DNA sequences. Such characters are predicted to be less prone to long-branch effects (Felsenstein, 1983). However, because sequences for such genes will provide only a smaller number of variable and informative characters, it is more difficult to obtain a large

enough number of them to minimize the effects of sampling error on phylogenetic inference (e.g., Graham and Olmstead, 2000). We therefore need to develop more genes for study of deep vascular-plant phylogeny, particularly from the nuclear genome, where relatively few have been used at this level of analysis (Donoghue and Doyle, 2000).

To better address outstanding issues in vascular-plant phylogeny, I designed new gymnosperm- and pteridophyte-specific primers for a portion of the slowly-evolving nuclear gene, *RPB2*, and sequenced it for vascular plants representing all major taxa, except one group of eusporangiate ferns (Marattiaceae). I also use a simulation approach to examine the resulting data for potential long-branch attraction problems. *RPB2* codes for the second largest subunit of RNA polymerase II, which catalyzes mRNA synthesis in the nuclei of eukaryotic cells. Nuclear RNA polymerase genes are generally highly conserved across highly divergent organisms and have been used to address questions at very deep levels of phylogenetic history, including eukaryote origins, and the deep relationships among major prokaryote lineages (Iwabe et al., 1991; Sidow and Thomas, 1994). More importantly, *RPB2* has been demonstrated to have potential for estimating vascular-plant phylogeny at a deep level (Denton et al., 1998).

RPB2 is present in either one or two copies in vascular plants. In the angiosperms it is present as two copies in some eudicots (Oxelman and Bremer, 2000; Oxelman et al., 2004), but it is single copy in *Arabidopsis thaliana* (Larkin and Guilfoyle, 1993), and several non-eudicots, as well as *Ginkgo biloba, Cycas revoluta, Sequoia sempivirens, Gnetum gnemon, Selaginella densa* and *Marchantia polymorpha* (Denton et al., 1998; Oxelman et al., 2004). Since the gene is either single- or low-copy in all vascular plants examined to date, errors in

orthology/paralogy assessment should be relatively unlikely. However, even when nuclear genes are present as multiple copies, additional insights can potentially be derived from orthology/paralogy assessment (e.g., see Mathews and Donoghue, 1999; Oxelman et al., 2004). Nonetheless, the high degree of conservation and low-copy number of *RPB2* make it particularly appropriate for investigations of vascular-plant deep phylogeny.

Materials and Methods

Taxon sampling.

Sequences of exons 11-24 (see Denton et al., 1998) were obtained for 65 vascular plants. Exon 11 was not examined for some taxa because I used a primer located downstream of it as one of the amplification primers in those taxa (Table 2.1). Eighteen of these sequences were from previous studies and are available on GenBank. Taxa were chosen to exemplify total vascular-plant diversity, as well as the morphological and taxonomic diversity of each group. The species included are: three lycopods, *Equisetum hymenale*, *Psilotum nudum*, two eusporangiate ferns, eight leptosporangiate ferns, eight cycads, *Ginkgo biloba*, 17 conifers, three Gnetales, and 20 angiosperms (Table 2.1). Repeated attempts to obtain RNA from *Angiopteris* (Marattiaceae), and to amplify *RPB2* from *Osmunda* (Osmundaceae, a basal leptosporangiate family; Hasebe et al. 1995, Pryer et al. 2001) cDNA were unsuccessful. All extant lycopod, cycad, conifer, and Gnetales families are represented, and the sampled angiosperms were chosen to represent a broad sampling of the basal nodes of angiosperm phylogeny. One bryophyte (the liverwort *Marchantia*) was used as an outgroup.

Primer design.

I used several new *RPB2* sequences derived using the primers of Denton et al. (1998), and several available on GenBank (Table 2.1), to design well-spaced primers that take account of observed variation in the pteridophytes and seed plants, respectively. Oligo v. 6 (Molecular Biology Insites, Inc., Cascade, CO) was used to assess the T_M and the potential for hairpin formation in each primer. A primer was discarded or modified if its T_M was lower than 60 C or if it formed obvious hairpin regions. All primers were designed to be longer than 20 nucleotides to increase binding specificity. Figure 2.1 is a primer map showing the locations and sequences of the primers developed for this study.

RNA extraction and cDNA amplification.

Since *RPB2* contains numerous large introns that are very rapidly evolving and difficult to align across distantly related taxa (Denton et al., 1998), a reverse-transcription PCR (RT-PCR) approach was used to obtain exonic sequences. Total RNA was isolated from either fresh plant material or tissue preserved in RNAlater (Ambion Inc., Austin, TX), using the RNeasy plant mini kit (Qiagen Inc., Valencia, CA). Manufacturer instructions were followed for RNA isolation, except that I used 1 ml of lysis buffer adjusted to 2% sarcosyl and 0.1% ß-mercaptoethanol, v/v, and incubated samples in lysis buffer for ten minutes at 65 C before processing.

RT-PCR was performed using "Ready-To-Go" RT-PCR beads (Amersham Biosciences, Piscataway, NJ) with 20 pmol of each amplification primer and 0.5 ug of $pd(T)^{12-18}$ added to each tube, in addition to the manufacturer recommended amounts of RNA extract and water. The $pd(T)^{12-18}$ was used as a primer for cDNA synthesis. Generally, two sets of amplification reactions were required to obtain entire sequences for each taxon. For gymnosperms, ConC6F and ConC8R were used to amplify the 5'-end of the gene and ConC7F and ConC11R were used to amplify the 3'-end of the gene. In non-seed plants, PB6F and PB8R, and PB7F and PB11R were used to amplify these respective portions of the gene. Alternative primers situated close to these were used if amplifications with the above primers failed to produce any product. Each tube was heated at 45 C for 45 minutes to allow for cDNA synthesis. This step was followed by a ten minute incubation at 95 C to denature the reverse transcriptase. The following thermal cycler profile was then used to amplify *RPB2*: 95 C for 45 seconds, 45 C for 45 seconds, and 72 C for two minutes. 35 cycles of this were completed, which was followed by a final extension of 72 C for 20 minutes.

For a few taxa, I used nested PCR reactions to obtain sufficient PCR product for cloning. To do this, the RT-PCR steps outlined above were performed, and 1 to 2 μL of the RT-PCR product were re-amplified using primers internal to the ones used in the first amplification. I used the same PCR cycle described above for the second amplification, but with no reverse transcription step. One unit of *Pfu* polymerase was used instead of *Taq* polymerase in the second amplification because it has a better proofreading ability, which should minimize the number of additional errors introduced during the extra rounds of amplification. However, because *Pfu* polymerase produces blunt-ended products, and the cloning kit used requires that the PCR products have "A" overhangs, the

final amplification products were then incubated at 72 C with one unit of *Taq* polymerase for 25 minutes before they were cloned.

Cloning, sequencing, data compilation and alignment.

RT-PCR products were cloned using the TOPO TA Cloning Kit (Invitrogen, Inc., Carlsbad, CA), following manufacturer instructions. A minimum of 10 positive clones were screened for each RT-PCR product, and at least six that were the correct size were sequenced on at least one strand. For each taxon, at least one set of clones was derived from a separate round of RT-PCR reactions produced on a separate day to control for errors. All sequences were thus sequenced with multiple redundancy, and with minor exceptions most taxa were completely sequenced at least once for forward and reverse strands. Rarely, taxa were sequenced multiple times in one direction. No evidence was found of multiple *RPB2* loci in any of the taxa examined here; any interspecific variation present was consistent with *Taq* polymerase error or allelic variation, coded here as "N." It is not straightforward to distinguish between the latter two sources of variation, but the total amount observed was very small in the context of the variation observed among taxa (see Results), and hence is more likely to represent allelic variation or *Taq* error than locus duplication followed by copy divergence. It would be very unlikely for duplications to be limited solely to terminal taxa. There were 0 to 7 (mean of ~3.4) variable sites across positive clones sequenced per taxon.

Sequencing products were generated using either the "Big Dye" Terminator v 3.1 cycle sequencing kit (Applied Biosystems, Foster City, CA), or the "DYEnamic ET" terminator cycle sequencing kit (Amersham Biosciences,

Piscataway, NJ), following manufacturer instructions. Sequencing reactions were cleaned using Sephadex G50 columns, dried on a vacuum centrifuge and run on an ABI Prism 377 automated DNA sequencer. Sequences were compiled and base-called using Sequencher 4.1 (Gene Codes Corporation; Ann Arbor, MI). The final sequences for each taxon (see Table 2.1) were exported and aligned manually using Se-Al version 1.0 (Rambaut, 1998) using criteria laid out in Graham et al. (2000).

Phylogenetic analysis.

Maximum parsimony analysis.— I conducted all maximum-parsimony (MP) tree searches using PAUP* version 4.0b10 (Swofford, 2002). All characters and character-state changes were equally weighted. I performed heuristic searches using tree-bisection-reconnection (TBR) branch swapping, with 100 random addition replicates and no tree number limits. One MP search included all taxa, and a second excluded four angiosperms (*Arabidopsis, Aristolochia, Hordeum,* and *Peperomia*) that apparently have a destabilizing effect on inference of relationships among the other angiosperms included in the study (see Results). A parsimony-based bootstrap analysis (Felsenstein, 1985) was performed for each of the above described taxon sets using the same search conditions, except that one random addition replicate was used for each bootstrap replicate.

I also performed a series of constrained searches using maximum parsimony, to infer the best trees consistent with particular major hypotheses of seed-plant and fern relationships. The hypotheses (Figs. 2.2, 2.3) are described in more detail below, in the sections on the Shimodairo-Hasegawa tests and the simulation studies. I constructed constraint trees in MacClade v. 4.03 (Maddison

and Maddison, 2001), constraining only those branches consistent with each major hypothesis (asterisks in Figs. 2.2, 2.3). These constraints were loaded into PAUP* as unrooted trees and MP tree searches were performed. Search conditions for the constrained searches were the same as for the unconstrained searches.

Maximum-likelihood analysis.— I performed a maximum-likelihood analysis using PAUP version 4.0b10, using the Hasegawa-Kishino-Yano 85 (HKY 85; Hasegawa et al., 1985) model with a gamma-distribution of rates (Yang, 1994) to take account of rate variation (the "HKY + Γ" model). This model was not the optimal one (Table 2.2, and see next section) but was used so the analysis would be completed in a reasonable time frame (using this model, the analysis took six days on a 0.4 GHz G3 processor). The transition/ tranversion ratio and continuous gamma shape parameter (4 rate categories) were estimated from one of the shortest trees obtained from the MP analyses described above. The starting tree was obtained using neighbour-joining.

Bayesian analysis.— I performed Bayesian analyses using the program MrBayes version 3.0 (Ronquist and Huelsenbeck, 2003). The analyses were run using four chains for 1.5 million generations, with the default temperature (0.2). I used the general time reversible (GTR; Lanave et al., 1984; Tavaré, 1986; Barry and Hartigan, 1987; Rodríguez et al., 1990) model of sequence evolution with a gamma-distribution of rates (Yang, 1994) and with the proportion of invariable sites considered. This model (GTR + Γ + I) was chosen based on the result of a likelihood-ratio test (Huelsenbeck and Crandall, 1997; see Table 2.2). I discarded the first 200 trees (representing 20,000 generations), as the chains had stabilized

after this point. Clade posterior probabilities were estimated from a majority-rule consensus of the sampled trees.

Analyses were performed on the entire data set, on first- and second-codon positions combined, and on third codon-positions only. One additional analysis was performed using all of the data, but excluding *Arabidopsis, Aristolochia, Peperomia* and *Hordeum*. Another analysis was performed using all data and taxa, but allowing separate ("unlinked") estimates of substitution model parameter values for each of two data partitions: one consisting of the first and second codon positions and one consisting of third codon positions only.

Shimodaira-Hasegawa tests.— I used the maximum-likelihood based Shimodaira-Hasegawa test ("SH" test; Shimodaira and Hasegawa, 1999; Goldman et al., 2000), as implemented in PAUP* version 4.0b10, to compare the phylogenetic position of Gnetales in the Bayesian tree (Fig. 2.2a) to four alternative gymnosperm hypotheses found here or in other published studies. The alternative topologies (Fig. 2.2) are: (a) Gnetales as the sister group of Pinaceae (the "gnepine" hypothesis), found in the Bayesian analysis here (see Results) and also supported by the molecular analyses of Bowe et al. (2000) and Chaw et al. (2000); (b) Gnetales as the sister group of all other gymnosperms ("Gnetales basal in gymnosperms"), found by Schmidt and Schneider-Poetsch. (2002) and the MP tree found here (see Results); (c) Gnetales as the sister group of the angiosperms (the "anthophyte" hypothesis), found by Crane (1985), Doyle and Donoghue (1986, 1992), Loconte and Stevenson (1990) and Doyle (1996); (d) Gnetales as the sister group of all other seed plants ("Gnetales basal in seed plants"), found by Rydin et al. (2002), and Rai et al. (2003), and; (e) Gnetales as the sister group of

conifers, with conifer monophyly constrained (the "gnetifer" hypothesis), found by Chaw et al. (1997).

The SH test was also used to compare the position of Psilotaceae found in the optimal Bayesian tree to three alternative topologies. The four hypotheses (Fig. 2.3), are referred to as hypotheses "a"–"d" for convenience. They are: (a) Psilotaceae as the sister group of Ophioglossaceae; (b) Psilotaceae as the sister group of all remaining euphyllophytes; (c) Psilotaceae as the sister group of Ophioglossaceae plus leptosporangiate ferns, and; (d) Psilotaceae as the sister group of the leptosporangiate ferns only. Hypothesis "a" was recovered by the MP, ML and Bayesian analyses presented here, and is supported by other molecular studies (Manhart, 1995; Wolf, 1997; Pryer et al., 2001; Rydin et al., 2002). Morphological studies differ in their placement of Psilotaceae, although they typically support its placement within the euphyllophyte clade (Renzaglia et al., 2000). The latter clade corresponds to all vascular plants except the lycopods (note that the euphyllophytes were not constrained in the other hypotheses). Stevenson and Loconte (1996) and Rothwell (1999) suggest that Psilotaceae are basal within the euphyllophyte clade (hypothesis "b"). I also examined the possibility that ferns in the classical sense (leptosporangiate and eusporangiate ferns) are monophyletic, with Psilotaceae their sister group (hypothesis "c"). Finally, I examined the possibility that Psilotaceae are the sister group of the leptosporangiate ferns (hypothesis "d"). I did not constrain monophyly of the leptosporangiates in the latter case, potentially allowing Psilotaceae to nest within them, consistent with Bierhorst (1971, 1977), who argued that Psilotaceae are related to a family of leptosporangiate ferns. I performed the SH tests using RELL (resampled estimated log-likelihood) estimates of the test distribution

(Shimodaira and Hasegawa, 1999; Goldman et al., 2000) and used the GTR + G + Γ model of sequence evolution.

Simulation-based tests of error and bias.— To determine if the placement of Gnetales or Psilotaceae might be an artifact of a high error rate or underlying biases in the data, five sets of one hundred alternate data sets were simulated using the program Seq-Gen v. 1.2.5 (Rambaut and Grassly, 1997). The trees used to simulate the sequences are the same as those used in the SH test described above. Branch lengths and sequence evolution parameters for each of these trees were estimated via maximum-likelihood as implemented in PAUP* v. 4.0b10, using the the "HKY + Γ" model of sequence evolution. Three groups of simulated data sets were generated for each hypothesis of relationship: one where each data set was 1,773 bp long, one where each was three times the size of the original data set (5,713 bp long), and one where each was nearly six times the size of the original data set (10 kb long). Increasingly larger data sets were simulated to determine if any apparent bias persists as more data are added, a condition known as statistical inconsistency when data sets approach infinite size.

The shortest MP tree(s) were found for each simulated data set. All characters and character-state changes were equally weighted and heuristic searches were conducted using tree-bisection-reconnection (TBR) branch swapping, with 10 random addition replicates and no tree limits. The optimal likelihood tree was also found for the 1,773 bp simulations, using the program PHYML (Guindon and Gascuel, 2003). Because of time constraints, the HKY + Γ model of sequence evolution was used for tree searching. All details of relationship were ignored for scoring the trees inferred using the simulated data

sets, apart from the hypothesis under consideration (i.e., the constrained branches in Figs. 2.2, 2.3).

Results

The final aligned data matrix is 1,773 base pairs (bp) long, and contains 774 sites that are variable and parsimony-informative. The translated portion of *RPB2* is well-conserved across vascular plants, and, with several exceptions, was fairly straightforward to align by eye. However, exons 15, 16 and 22 contained multiple indels. In exon 15, a six bp indel (inferred to be a deletion) is common to all conifers except Pinaceae, a three bp indel (inferred to be an insertion) is common to all eudicots, and two 15 bp indels (one of these is 18 bp long in *Stangeria*), inferred to be insertions, are common to all cycads excluding *Cycas*. In exon 22, a three bp indel (inferred to be a deletion) and a nine bp indel (inferred to be an insertion) are common to all Gnetales, and a six bp indel (inferred to be a deletion) occurs in *Trimenia* that overlaps a three bp indel (inferred to be a deletion) that occurs in *Illicium*. A three bp indel (inferred to be a deletion) occurs in *Selaginella* (exon 16). Finally, an unusual 15 bp frame-shifted duplication and a neighbouring 12 bp (15 bp in *Microcycas*) indel (inferred to be a deletion), which are both common to all cycads, occur in exon 16. The latter two indels are illustrated in Fig. 2.4.

Maximum-parsimony inferences.— One of the 26 most-parsimonious trees obtained from the MP analysis of all taxa and data is shown in Fig. 2.5. Estimates of branch support (bootstrap values, BV) are shown beside branches. In the MP analysis, the lycopods do not form a clade. Two lycopods (*Lycopodium* and

Isoetes) are the successive sister taxa of the remaining vascular plant clades, but the remaining lycopod, *Selaginella,* is depicted as the sister taxon of the seed plants. However, the non-monophyly of the lycopods is not well-supported by bootstrap analysis here. The moniliforms form a clade, also with low bootstrap support. There is, however, a very robustly supported (97% BV) sister-group relationship between *Psilotum* and Ophioglossaceae. Of the leptosporangiate ferns examined here, *Anemia* (Schizaeaceae), *Marsilea* (Marsileacae), and *Dicksonia* (Dicksoniaceae) are successive sister taxa to the others (but of these relationships, only *Dicksonia* placement is well supported here). *Adiantum* (Pteridaceae) and *Cryptogramma* (Pteridaceae) form a clade that is the sister group of a clade composed of *Stenochlaena* (Blechnaceae), *Polystichum* (Dryopteridaceae) and *Dryopteris* (Dryopteridaceae). These relationships (including those found among the latter three species; Fig. 2.5) are inferred with robust bootstrap support.

Gymnosperm monophyly is moderately well supported here (85% BV). Within the gymnosperms, Gnetales are weakly supported as the sister taxon of the remaining taxa. A moderately-supported clade (80% BV) composed of *Ginkgo* and cycads is the sister group of the conifers, which are weakly supported as monophyletic (54% BV). Within the conifers, Pinaceae are the sister group of all remaining families. Araucariaceae and Podocarpaceae comprise a weakly supported clade (57% BV) that is the sister group of the remaining taxa. Sciadopityaceae are the sister group of a clade composed of Taxaceae, Cephalotaxaceae, and Cupressaceae. Taxaceae, represented here by *Taxus* and *Torreya,* do not form a monophyletic group, although their lack of monophyly is only weakly supported. Most relationships within Cupressaceae are robustly supported here (Fig. 2.5).

Within cycads, *Cycas, Dioon* and *Stangeria* are successive sister taxa to the remainder of the group. The placements of *Cycas* and *Dioon* are moderately to robustly supported, that of *Stangeria* only weakly so. There is a strongly supported sister-taxon relationship between *Zamia* and *Microcycas* (100% BV). *Ginkgo* and cycads form a clade with moderate (80% BV) support.

Relationships within angiosperms generally are not well-supported and are somewhat incongruent with a well-accepted classification of the angiosperms (APG II, 2003). For example, neither the monocots nor the eudicots are monophyletic in this analysis (Fig. 2.5). However, the removal of four taxa (*Arabidopsis, Aristolochia, Hordeum* and *Peperomia*) results in 10 most-parsimonious trees with relationships that generally correspond more closely to accepted and well-supported angiosperm clades (Fig. 2.5, inset). Relationships among angiosperms are generally poorly supported by maximum-parsimony bootstrap analysis.

Bayesian analysis.— Clade posterior probability (PP) estimates (expressed as percentages) did not appear to be substantially affected by whether substitution model parameters were forced to be homogeneous across all data, or permitted to differ between two codon partition positions (positions 1 + 2 vs. 3; Fig. 2.6). However, parameter estimates for this data set are often quite different across partitions. Table 2.3 shows parameter estimates for the GTR + Γ + I model of sequence evolution for all codon positions, codon positions 1 and 2 only and codon position 3 only. These estimates were made from an ML tree constructed using all data (see Fig. 2.8). In general, support values for the Bayesian analysis were more robust than in the parsimony analysis (Fig. 2.5), and inferred relationships in the angiosperms bore a closer resemblance to commonly accepted

relationships than they did with the parsimony results. As is the case in the MP tree, the lycopods are non-monophyletic, although *Isoetes* and *Lycopodium* form a well-supported clade, and *Selaginella* emerges from a poorly supported position near the base of the vascular plants. Psilotaceae and Ophioglossaceae form a robustly supported clade, and both are robustly supported as the sister group of Equisetaceae. The moniliforms are non-monophyletic, although support values for branches disrupting their monophyly are relatively low (71% or less). Ignoring *Selaginella*, relationships within the leptosporangiate ferns are completely congruent with those recovered in the MP analysis and are robustly supported here.

There is robust support for a gymnosperm clade (100% PP) and a clade composed of *Ginkgo* and cycads (100% PP), findings that are congruent with the MP results. The position of Gnetales, however, is quite different. Here, they are robustly supported (99% PP) as being nested within conifers, as the sister group of Pinaceae. Otherwise, conifer relationships are similar to those found in the MP analyses.

Within angiosperms, *Amborella* (Amborellaceae), *Nymphaea* (Nymphaeaceae) and a clade composed of *Trimenia* (Trimeniaceae) and *Illicium* (Illicaceae) are successive sister taxa to the remaining angiosperms. The position of *Amborella* is supported with 100% PP. Relationships among the other angiosperms are generally poorly supported, but are generally less at odds with accepted higher-order groupings of angiosperms than the parsimony analysis. The two angiosperm relationships that are well supported in the parsimony analysis (*Spinacea-Dianthus*, *Trimenia-Illicium*) are also robustly supported here, and a few other other relationships also find robust support (*Dioscorea-Musa*, *Arabidopsis-Solanum*, *Platanus-Vitis*). Again, neither the monocots nor the

eudicots are depicted as monophyletic, and the monophyly of both is robustly contradicted. The two rosids included here, *Arabidopsis* and *Vitis*, also do not form a monophyletic group. An analysis that excluded *Arabidopsis, Aristolochia, Hordeum* and *Peperomia* from consideration depicts the monocots and eudicots as monophyletic (tree not shown).

Within leptosporangiate ferns, *Anemia* (Schizaeaceae) is the sister taxon of the remaining taxa, and *Marsilea* (Marsileaceae) and *Dicksonia* (Dicksoniaceae) are the successive sister taxa of the remaining taxa. *Adiantum* and *Cryptogramma* (Pteridaceae) form a clade that is the sister taxon to one composed of *Stenochlaena* (Blechnaceae) and *Dryopteris* and *Polystichum* (Dryopteridaceae). These results are congruent with the ones found by the MP analysis.

The major vascular-plant relationships seen in Bayesian analysis of codon positions 1 + 2 or codon position three are summarized in Fig. 2.7. In the former case, lycopods are non-monophyletic, and moniliforms are recovered as a clade that is the sister group of seed plants. Ophioglossaceae and Psilotaceae are robustly supported (100% PP) as sister taxa. The position of *Equisetum* is only weakly supported, and the gymnosperms are again depicted as monophyletic. This analysis also recovers a clade (100% PP) with Gnetales nested in conifers (moderately well supported as the sister group of Pinaceae; 80% PP), but has the cycads as the sister group of all remaining gymosperms, with moderate support (87% PP).

Relationships inferred by the Bayesian analysis using only third codon position data are very similar to those seen with all of the data combined (cf. Figs. 2.6, 2.7b), and any incongruence between the two is not well-supported. There is some fairly substantial incongruence between the trees obtained from the Bayesian analyses of the two codon partitions considered here (Fig. 2.7a, b). In

the tree obtained from using only codon positions 1 and 2, *Equisetum* is the sister group of the remaining moniliforms with low support (65% PP) and *Ginkgo* is the sister group of a clade composed of conifers and Gnetales. In contrast, the analysis that considered only codon position 3 places *Equisetum* as the sister group of a clade composed of Ophioglossaceae and Psilotaceae (100% PP) and places *Ginkgo* as the sister group of the cycads (100% PP). A few minor conflicts between the two analyses (regarding relationships within major taxa, not shown here) were not robustly supported.

Maximum-likelihood analysis.— The tree obtained from the ML analysis of the *RPB2* data set is shown in Fig. 2.8. Relationships among and within seed plant groups are similar to the ones found in the Bayesian analysis of all data (Fig. 2.6). However, in the ML tree, Podocarpaceae and Araucariaceae do not form a clade, as they do in the Bayesian and parsimony analyses. In the angiosperms, monocots and eudicots are inferred to be monophyletic, and *Platanus* and *Vitis* do not form a clade (although *Vitis* is not found to be the sister group of *Arabidopsis*, the only other rosid in the analysis). The latter findings conflict with the corresponding Bayesian results.

Relationships among free-sporing plants between the ML and Bayesian trees are also different. In the ML tree, *Selaginella* is the sister group of the seed plants. Additionally, the eusporangiate ferns plus *Psilotum* are the sister group of the leptosporangiate ferns in the ML tree.

Shimodaira-Hasegawa tests.— In the SH test focussing on the position of Gnetales, the tree representing the "gnepine" hypothesis was the optimal tree of those examined (as it is in the Bayesian and ML analysis, Figs. 2.6–2.8). Trees

rejected by the SH test, with a critical value of 0.05, are: (1) The "Gnetales basal in gymnosperms" tree (Fig. 2.2b; $P = 0.0083$); (2) The "anthophyte" tree (Fig. 2.2c; $P = 0.0011$), and; (3) the "Gnetales basal in seed plants" tree (Fig. 2.2d; $P = 0.029$). The "gnetifer" tree (Fig. 2.2e) was not rejected by the test ($P = 0.657$). In the SH test focussing on the postition of Psilotaceae, tree "a" (Fig. 2.3a) was best, and all others were strongly rejected (Figs. 2.3b–d; $P < 0.0005$).

Simulation results.—The diagonal elements of Tables 2.4 and 2.5 represent estimates of the probabilities of (1 minus) the type 1 error for each major hypothesis (the probablility of rejecting the null hypothesis when it is true; in this case the null hypothesis is the tree used in simulations, to the left of each row). Type 2 errors are estimated here by reference to the non-diagonal elements of each column. This estimate is the probability of reconstructing the null hypothesis (the one heading that column) given that some alternative hypothesis (in this case a tree used for simulation) is correct (see Sanderson et al., 2000). We did not test all possible alternative hypotheses, so these estimates are conditional on the ones we did examine.

For the five main seed-plant hyptheses considered here, type 1 error rates are generally very low, except for the anthophyte hypothesis (Table 2.4). Using maximum parsimony to reconstruct trees, this hypothesis has a high type 1 error rate (up to 57%, for 1,773 bp). The type I error rate for the anthophyte hypothesis is not substantially reduced when a large number of nucleotides (~5–10 kb) are considered, and so this hypothesis would be difficult to infer with parsimony using the current data, if correct. However the predicted type 1 error rate of this hypothesis is reduced to a moderate level (20%) when maximum likelihood is used to infer trees, even for data sets of the size actually used. The "gnetifer"

hypothesis also has a moderately large type I error when the data set is as large as what I used, although the error is less for likelihood than parsimony (11% vs. 29%), and decreases for parsimony to zero when larger data sets are used (Table 2.4).

Type 2 error rates are also generally low for the seed-plant relationships examined here, apart from the "Gnetales basal in seed plants" hypothesis, which would often be inferred if the anthophyte hypothesis were correct (even with 10 kb of data), unless maximum likelihood is used to reconstruct trees (then there is only 7% error). Moderately small type 2 error rates (up to ~20%) are also observed for the "gnepine" hypothesis and one hypothesis of gymnosperm monophyly (with Gnetales basal; Fig. 2.2b), when either likelihood or parsimony are used to reconstruct trees, using small to moderately large data sets (1.77–5.32kb).

There is substantial type 1 error for all hypotheses concerning Psilotaceae placement considered here (Table 2.5), except for the one observed (hypothesis "a"; Fig. 2.3). For the three other hypotheses considered here, most would not have a high chance of being inferred, if correct, even if substantial amounts of data were employed (type 1 error rates of 84–63% across the three hypotheses for all three data set sizes, using maximum parsimony tree inference). The type 2 error rates of all hypotheses are also substantial for most combinations of simulation tree and data set size (Table 2.5), except for hypothesis "b" (Fig. 2.3), where they are uniformly low. Using ML tree inference only slightly decreases the chance of making a type 1 error. There appears to be a moderate bias towards inference of the hypothesis oberved with the real data ("Ophioglossaceae plus Psilotaceae"), regardless of the trees actually used in simulation.

Discussion

None of analyses presented here indicate that extant lycopods are monophyletic, because *Selaginella* is always observed to be isolated from *Lycopodium* and *Isoetes*. However, both morphological and molecular studies support the group's monophyly (Kenrick and Crane, 1997; Maden et al., 1997; Nickrent et al., 2000; Renzaglia et al., 2000; Pryer et al., 2001; Dombrovska and Qiu, 2004). The non-monophyly of lycopods in analyses of *RPB2* is not well-supported by parsimony bootstrap values or Bayesian analysis, so the unusual placement of Selaginellaceae could simply be dismissed as sampling error. However, the extremely long branch leading to *Selaginella* (Fig. 2.8) may also be a contributing factor. Manhart's (1995) analysis of plastid 16S rDNA sequences indicate that *Selaginella* is unrelated to other lycopods, but it was also subtended by a long branch in their analysis. Korall and Kenrick (2002) and Korall and Kenrick (2004) also report that Selaginellaceae have high levels of sequence divergence relative to other vascular-plant groups in plastid *rbc*L and nuclear 26S rDNA sequences, respectively. Korall and Kenrick (2004) argue that this elevated rate of sequence divergence is the result of a high substitution rate in *Selaginella*, rather than its long evolutionary history, because other related taxa that also have ancient origins (such as Lycopodiaceae) do not possess such a high degree of sequence divergence.

Ignoring *Selaginella*, the moniliforms and seed plants (together, the euphyllophytes) form a clade in all unconstrained analyses (Figs. 2.5–2.8), although the only Bayesian analysis to support moniliform monophyly was the one that considered the first two codon positions (Fig. 2.7a). However, moniliform monophyly was not robustly contradicted by the other Bayesian

analyses (Figs. 2.6, 2.7b). Thus, no strong conclusions can be made here regarding the monophyly of the moniliforms. Most recent molecular (Pryer et al., 2001; Rydin et al., 2002; Dombrovska and Qiu, 2004) and morphological (Renzaglia et al., 2000) studies indicate that the ferns (eu- and leptosporangiate ones), *Psilotum*, and *Equisetum* form a clade. There is little consensus here regarding the exact position of *Equisetum*, or the relative positions of the eusporangiate and leptosporangiate ferns. *Equisetum* is well-supported by the main Bayesian analyses and the Bayesian analysis of codon position 3 as the sister group of a clade composed of Ophioglossaceae and Psilotaceae (Figs. 2.6, 2.7b), but this relationship is not observed in the MP or ML analyses or the Bayesian analysis of the first two codon positions, which all place *Equisetum* as the sister group of all other moniliforms (none of the latter analyses with strong support; see Figs. 2.5, 2.7a).

Although only a small sampling of leptosporangiate ferns were included here, relationships among them were identical in the parsimony, likelihood and Bayesian analyses. Of the taxa included here, *Anemia* (Schizaeaceae), *Marsilea* (Marsileaceae) and *Dicksonia* (Dicksoniaceae) were successive sister taxa to the remaining leptosporangiate ferns. This result is congruent with other molecular studies on fern phylogeny (Hasebe et al., 1995; Pryer et al., 2001). In all analyses, Dryopteridaceae and Blechnaceae are depicted as a clade that is the sister group of Pteridaceae. This is also congruent with other studies (Hasebe et al., 1995; Pryer et al., 2001).

Despite the lack of resolution among major pteridophyte groups, all analyses here were consistent in strongly supporting a sister-group relationship between Psilotaceae and Ophioglossaceae. The family is composed of two genera, *Psilotum* and *Tmesipteris*, but the latter was not included in this study

because of difficulties in getting hold of fresh material. When the two genera were described in the 1700s, they were presumed to be lycopods, and *Psilotum nudum* (L.) Beauv. and *Tmesipteris tannensis* (Spreng.) Bernh. were both placed in the genus *Lycopodium* as *L. nudum* L. and *L. tannensis* Spreng. In the early 1900s, this view changed and they were generally regarded as relatives of newly discovered early Devonian rhyniophytes (Gensel, 1977). Bierhorst (1977) argues that this was based largely on superficial similarities, such as the lack of leaves and roots, and the presence of dichotomous branching. Bierhorst (1971, 1977) asserts that arguments supporting a close relationship between rhyniophytes and Psilotaceae tended to focus exclusively on *Psilotum nudum* and generally ignored *Tmesipteris*. He instead argued that Psilotaceae belong within leptosporangiate ferns and are closely related to *Stromatopteris*, because the two share many morphological similarities. For instance, both taxa have bean-shaped monolete spores with similar perispore patterns, cylindrical gametophytes that are similar to their subterannean sporophyte axes, septate rhizoids on the gametophytes, and antheridia with lateral, opercular cells. Wagner (1977), however, argued that Psilotaceae should not be considered a leptosporangiate fern, because sporangial morphology is quite different in each case. Psilotaceae, for instance, are eusporangiate and have fused sporangia (synangia) and gleicheniaceous ferns are leptosporangiate and do not have fused sporangia. Additionally, the gametophytic characters shared between Psilotaceae and Gleicheniaceae could easily have arisen independently in the two taxa, as has happened in other unrelated taxa. They also could be plesiomorphic character states that have been retained by the two taxa and thus do not provide information on whether the two taxa are related.

Other recent studies have indicated that Psilotaceae belong with the euphyllophytes and are not related to early Devonian land plants or lycopods. For

instance, Psilotaceae lack a chloroplast inversion found only in bryophytes and lycopods (Raubeson and Jansen, 1992a), and possess an intron in the mitochondrial gene *nad*2 that is common to all other moniliforms (Dombrovska and Qiu, 2004). A morphological cladistic study suggests that Psilotaceae are the sister group of all remaining euphyllophytes (Rothwell, 1999). Most molecular analyses, however, suggest that they are the sister group of Ophioglossaceae (Malek et al., 1996; Nickrent et al., 2000; Manhart, 1995; Wolf, 1997; Pryer et al., 2001).

The Shimodaira-Hasegawa (SH) test (Shimodaira and Hasegawa, 1999) performed here indicate that the topology that depicted Psilotaceae as the sister group of Ophioglossaceae is optimal. The sensitivity of this test to the high levels of error demonstrated here (Table 2.5) is not clear. Felsenstein (2004) notes that one limitation of this test is that all proposed trees are initially assumed to be equally probable, so including too many trees can dilute the power of the test. Since most recent data suggest that Psilotaceae are a part of the euphyllophyte clade, I included four trees in the analysis that were congruent with this and did not consider topologies that placed Psilotaceae as the sister to all remaining vascular-plants

A finding that Psilotaceae and Ophioglossaceae are sister taxa would contradict Bierhorst's (1971, 1977) theory that Psilotaceae are related to leptosporangiate ferns. However, since a relationship between Ophioglossaceae and Psilotaceae has not yet been proposed in any morphological study, I decided to examine the possibility that this is a result of systematic error (such as long-branch attraction). Psilotaceae and members of Ophioglossaceae are apparently not on particularly long branches here when compared to other moniliforms (Fig. 2.8), but Hendy and Penny (1989) have shown that parsimony can be inconsistent

in some cases when rates of change are even, if very short internal branches separate other, longer branches. I used a simulation approach to examine the potential here for systematic error regarding the relationship between Psilotaceae and Ophioglossaceae.

If a tree depicting Ophioglossaceae and Psilotaceae as sister taxa (Fig. 2.3; hypothesis "a") is used to simulate data sets, then the correct tree is always recovered, regardless of the size of data set or the tree building method used (Table 2.5). However, very high type 1 error rates were seen when any other topology was used to simulate data sets, with hypothesis "b" having the highest amount. However, the inferred trees for simulations based on topology "b" rarely matched any of the other hypotheses examined here (Fig. 2.3) and were extremely variable. They often placed Psilotaceae as the sister group of the seed plants or *Equisetum*. This suggests that tree "b" may be incorrect because, although there was a high type 1 error rate for it, the trees that were inferred were not recovered by the data here, or by any other other molecular or morphological studies.

Topologies "c" or "d", derived from analyses that constrained Psilotaceae as the sister group of all ferns or leptosporangiate ferns only, repectively, also resulted in very high type 1 error rates. In both cases, topologies consistent with hypothesis "a" were frequently recovered. When parsimony was used to reconstruct trees, increasing the size of the data set to ~six times the size of the original one resulted in an increased probability of mistakenly inferring topologies consistent with hypothesis "a" and an increased type 1 error rate for hypothesis "d." As adding more data apparently increases the chance of inferring an incorrect tree, this may indicate that the *RPB2* data are statistically inconsistent concerning this relationship. However, it may also be true that the "large" data sets considered here do not provide strong insights into what would be obtained

with infinitely large data sets, the size needed to infer inconsistency. In addition, the pattern for another hypothesis,"c", was not linear with regards to data size, as the largest type 1 error rate for this hypothesis occurred with an intermediate-sized data set. Using maximum-likelihood for hypotheses "b", "c" and "d" resulted in a slightly higher probability of inferring the true tree, but the improvement was not dramatic.

These results suggest that if Psilotaceae are related to leptosporangiate ferns, or are unrelated to ferns but still belong within the euphyllophyte clade, molecular data sets such as this one may not recover this result correctly, and may instead appear to support a close relationship between Psilotaceae with Ophioglossaceae. It should be noted that these simulation results do not give much indication as to which hypothesis is true, and a sister group relationship between Ophioglossaceae and Psilotaceae was strongly supported by the SH tests presented here. Many other molecular studies have supported a sister-group relationship between Psilotaceae and Ophioglossaceae (Malek et al., 1996; Nickrent et al., 2000; Manhart, 1995; Wolf, 1997; Pryer et al., 2001), and the probability that in each case the result was due to a type-2 error is likely quite low. If Psilotaceae and Ophioglossaceae are sister taxa, it is possible that they diverged an extremely long time ago and that sufficient time has elapsed to leave the two groups very distinctive, morphologically. Unfortunately, the fossil record for Psilotaceae is non-existent (Stewart and Rothwell, 1993), and for Ophioglossaceae it is very sparse, although Rothwell and Stockey (1989) described the Palaeocene-age fossil *Botrychium wightonii*. Nonetheless, the results presented here indicate that molecular studies that support a sister-group relationship between Psilotaceae and Ophioglossaceae should be treated with caution. Additional studies such as this one using plastid or mitochondrial genes

would be useful in determining if the bias noted here is unique to *RPB2* or is also a potential problem with other genes, and the extent to which adding data (nucleotides) might help (or hinder) with maximum likelihood analysis or Bayesian inference.

I also used the SH test and a simulation approach to examine the phylogenetic position of Gnetales within the seed plants. The maximum parsimony analysis (Fig. 2.5) presented here supported a sister-group relationship between Gnetales and all remaining gymnosperms, with low bootstrap support. However, all Bayesian analyses (Figs. 2.6, 2.7) strongly supported the placement of Gnetales within conifers, as the sister group of Pinaceae, a relationship that was also seen on the best likelihood tree. Thus, there was no conflict in the Bayesian analyses of the first two vs. third codon positions regarding the position of Gnetales (Fig. 2.7), as there often is in phylogenetic studies using plastid genes (e.g., Sanderson et al., 2000; Magallon and Sanderson, 2002; and see Chapter 3). The placement of Gnetales within conifers, found recently in several multigene analyses (Bowe et al., 2000; Chaw et al., 2000), is incongruent with the majority of morphological studies, which support the anthophyte hypothesis (Crane, 1985; Doyle and Donoghue, 1986, 1992; Nixon et al., 1994; Rothwell and Serbet, 1994), and with other recent multigene molecular studies that place Gnetales as the sister group to all remaining extant seed plants (Rydin et al., 2002; Rai et al., 2003).

Although the Bayesian and maximum-likelihood analyses here find the "gnepine" hypothesis with high support, the Shimodaira-Hasegawa test did not reject the gnetifer hypothesis. My *RPB2* data thus indicate that Gnetales are related to conifers, but do not strongly differentiate between Gnetales being nested within conifers, or as the sister group of the entire conifer clade. A close

relationship between Gnetales and conifers in general has been proposed based on morphological characters (see Donoghue and Doyle , 2000). For example, the short shoots of Palaeozoic conifers have been compared to Gnetales cones (Hernandez-Castillo et al., 2001). However, a sister-group relationship between Pinaceae and Gnetales is quite difficult to reconcile with morphological data. It would require that a number of characters common to conifers, such as tiered proembryos, either evolved separately in Pinaceae and other conifers, or alternatively, that they were lost in the lineage leading to modern Gnetales (Donoghue and Doyle, 2000). Additionally, conifers have lost most or all of one copy of the large inverted repeat (IR) region in the chloroplast genome, but Gnetales have apparently not (Raubeson and Jansen, 1992b; see also Wakasugi et al., 1994). If Gnetales are really nested in conifers, they either regained an IR in a position comparable to other seed plants or it was lost independently in Pinaceae and other conifers.

My results provide further evidence that all phylogenetic studies concerned with the position of Gnetales should be interpreted carefully. Sanderson et al. (2000) have demonstrated that results of molecular analyses of seed-plant data can be misleading, at least for the two chloroplast genes they examined. In their simulation study, a topology that placed Gnetales as the sister group of all seed plants was often recovered with a moderate probability, even if this was not the topology used to simulate data sets. Additionally, the anthophyte tree often had a very low probability of being recovered, even if it was the "true" tree. The results presented here for *RPB2* are comparable to those obtained by Sanderson et al. (2000). When I used "gnepine", "Gnetales basal in gymnosperms", and "Gnetales basal in seed plants" trees to simulate data, type 1 error rates were low. The type 1 error rate for the "gnetifer" tree was moderate

when the data set was 1,773 bp long, and it decreased to zero when the data set was increased to 10 kb. However, when the anthophyte tree was used to simulate data sets, the type 1 error rate was high for maximum parsimony. Even with data sets 10 kb long, there was only a 0.52 probability of recovering the correct tree. It is noteworthy that a topology with Gnetales basal in seed plants has been recovered with high support by Rydin et al. (2002), Rai et al. (2003) and here using plastid data (Chapter 3). The *RPB2* results and those of Sanderson et al. (2000) indicate that at least some molecular data may be biased towards this hypothesis using maximum parsimony if some other hypothesis is instead correct. Nonetheless, my *RPB2* data actually found the angiosperms to be the sister group of the remaining seed plants using maximum parsimony (Fig. 2.5).

The use of maximum likelihood instead of maximum parsimony tended to increase the probability of recovering the correct tree and decrease that of recovering the incorrect tree, at least across the hypotheses examined here, for a data set as large as the one I actually used. This supports the idea that maximum likelihood is less prone than parsimony to sources of systematic error (bias) such as long-branch attraction. This may be because it can potentially correct for multiple substitutions ("misinformative" characters) along long branches, as has been demonstrated in other simulation studies (Huelsenbeck, 1995; 1998). However, the simulation tests involving Psilotaceae (Table 2.5) did not indicate that maximum likelihood was much more likely than maximum parsimony to recover the true tree, so ML analysis of *RPB2* is demonstrably error-prone in this case, at least.

Ignoring the problem of Gnetales placement, the relationships inferred among conifers here are largely congruent with other molecular studies. Pinaceae are inferred to be the sister group of all remaining conifers (e.g., Fig. 2.5), a result

that is well-supported by a number of other molecular studies (Chaw et al., 1995; Chaw et al., 1997; Stefanovic et al., 1998; Gugerli et al., 2001; Magallon and Sanderson, 2002; Quinn et al., 2002; Schmidt and Schneider-Poetsch, 2002), and a morphological study (Hart, 1987). Additionally, exon 15 contains a 6 bp indel (likely a deletion) that is common to all conifers except Pinaceae, which also supports a clade consisting of all conifers except Pinaceae. Araucariaceae and Podocarpaceae are weakly supported as sister taxa in the MP (Fig. 2.5) and Bayesian (Fig. 2.6) analyses presented here. Sciadopityaceae are inferred to be the sister group of a clade composed of Taxaceae, Cephalotaxaceae, and Cupressaceae. All of these results are in agreement with most other molecular phylogenetic studies involving conifers (Chaw et al., 1997; Stefanovic et al., 1998; Gugerli et al., 2001; Schmidt and Schneider-Poetsch, 2002; Quinn et al., 2002). Taxaceae, here represented by *Taxus* and *Torreya*, are paraphyletic with respect to Cephalotaxaceae (Figs. 2.5, 2.6, 2.8): *Cephalotaxus* is weakly supported as the sister group of *Taxus*, and the two are the sister group of *Torreya*, also weakly supported. Other studies indicate that Cephalotaxaceae and Taxaceae are closely related, but also could not resolve the exact relationships between them (e.g., Chaw et al., 1997; Stefanovic et al., 1998), or conclusively demonstrate whether or not Taxaceae is monophyletic. In their phylogenetic study using plastid *mat*K and nuclear ITS sequences, Cheng et al. (2000) found that Taxaceae [including *Amenotaxus*, sometimes grouped with Cephalotaxaceae; Page (1990a)] and Cephalotaxaceae were monophyletic sister-taxa, and argued that the two families are distinctive and should remain as separate families. However, Quinn et al. (2002) suggested that Cephalotaxaceae should be merged with Taxaceae, based on their phylogenetic analysis of two plastid regions (*rbc*L and *mat*K). The branch separating members of Taxaceae and Cephalotaxaceae

here is very short with respect to the *RPB2* data (Fig. 2.8), which may contribute to the lack of resolution on these issues. A larger amount of sequence data may therefore be needed to confirm the relationships between these taxa. If it is eventually found that Cephalotaxaceae and Taxaceae are monophyletic sister groups, as Cheng et al. (2000) found, then the issue as to whether or not to recognize Cephalotaxaceae would be largely a matter of taste.

Four taxa included here (*Cunninghamia, Metasequoia, Sequoia* and *Sequoiadendron*) were traditionally included in Taxodiaceae (see Page, 1990b). My results indicate that members of Taxodiaceae are nested in Cupressaceae *s.l.* The recognition of a broadly circumscribed Cupressaceae that includes taxa from Taxodiaceae is congruent with other molecular studies (Brunsfeld, 1994; Stefanovic, 1998; Quinn et al., 2002), and supports the conclusions made by Eckenwalder (1976), based on morphology.

All analyses presented here that use all of the data indicate that cycads and *Ginkgo* are sister taxa (Figs. 2.5, 2.6, 2.8). Molecular analyses by Goremykin et al. (1996), Chaw et al. (1997, 2000) and Rai et al. (2003) also provide support for this relationship. However, here is some conflict between codon positions with regards to the relationship between these taxa. Codon positions 1 and 2 instead place *Ginkgo* as the sister group of a clade consisting of conifers and Gnetales with moderate support, and codon position 3 places *Ginkgo* as the sister group of cycads. Within cycads, *Cycas* is the sister group to the remaining genera in all analyses. Additionally, the presence of two large indels in exon 15 in all cycads excluding *Cycas* provides further support for this relationship, as do a number of other morphological (Stevenson, 1990) and molecular (Treutlein and Wink, 2002; Hill et al., 2003; Rai et al., 2003; Bogler and Francisco-Ortega, in press) studies.

Relationships among cycads and the relationship between cycads and *Ginkgo* will be discussed further in the next chapter.

The gymnosperms are consistently monophyletic, which is congruent with a number of other molecular studies (Goremykin et al., 1996; Chaw et al., 1997, 2000; Bowe et al., 2000; Nickerson and Drouin, 2004); angiosperms are supported here, by all analyses, as the sister group of this gymnosperm clade. Within angiosperms, maximum parsimony, maximum likelihood and Baysian analysis find *Amborella* (Amborellaceae) to be the sister-group of all remaining angiosperms, congruent with most recent analyses (Mathews and Donoghue, 1999; Parkinson et al., 1999, Qiu et al., 1999; Soltis et al., 1999, 2000; Barkman et al., 2000; Graham and Olmstead, 2000; Graham et al., 2000, Zanis et al., 2002, Hilu et al., 2003). In a discordant study, Goremykin et al. (2003) placed *Amborella* in a more nested position in angiosperm phylogeny. However, this result is likely an artifact of the extremely low taxon density in that study, coupled with their choice of exemplar taxa (Soltis and Soltis, 2004). For example, the monocots were represented in their study by three grasses (Poaceae).

The next deepest splits in the angiosperm tree observed in the model-based analyses (Figs. 2.6, 2.8) are congruent with other recent results. My main Bayesian analysis depicts *Nymphaea* (Nymphaeacceae) and *Illicium-Trimenia* (exemplars of Illiciaceae and Trimeniaceae in Austrobaileyales) as the next successive sister groups of the remaining angiosperms, also consistent with most recent studies. However, only the *Illicium-Trimenia* clade is well supported here (Figs. 2.5, 2.6), and the likelihood analysis depicts *Peperomia* (Piperaceae) as the sister group of Nymphaeaceae, another anomalous result that might be a consequence of long-branch attraction. Most other relationships within the angiosperms are relatively poorly supported by Bayesian or parsimony bootstrap

analysis (Figs. 2.5, 2.6). Disturbingly, the eudicots and monocots, two widely recognized angiosperm and otherwise well-supported clades (e.g., APG II 2003) were not found to be monophyletic in most analyses here. There are a number of possible explanations for the general lack of support in the angiosperms. It could be a result of sampling error, as only one gene was used. The angiosperm taxa included here represent most of the major extant lineages, but are only a small fraction of all extant taxa (~260,000 species; Judd et al. 2002). The branches subtending each terminal taxon are very long relative to the deep angiosperm internodes (Fig. 2.8), and so it is quite possible that denser taxon sampling might substantially improve the accuracy of inferred angiosperm relationships using *RPB2* (Hillis, 1998).

Substantial branch-length heterogeneity may also be a contributing factor. Removing four taxa that include some of the longest terminal branches (*Arabidopsis, Aristolochia, Hordeum,* and *Peperomia*) from the data set results in a topology that depicts both eudicots and monocots as monophyletic in the maximum parsimony and Bayesian analyses. Finally, it is possible that I could have been comparing non-orthologous genes, due to one or more undetected gene duplication and/ or extinction events in the angiosperms. Multiple sequences of cloned RT-PCR products were obtained, and any variation seen was minor and consistent with allelic variation or *Taq* polymerase errors. However, it is possible that a second copy in some angiosperms was overlooked here, or once existed and has since became extinct in one or more taxa. Oxelman and Bremer (2000) found a second copy of *RPB2* in two members of the eudicot order Gentianales. Oxelman et al. (2004) cloned and sequenced *RPB2* for a larger number of eudicots and found two copies of the gene in several taxa. The angiosperm taxa chosen here all appear to have one copy (see also Oxelman et al, 2004). The extra

copy arising from the duplication appears to have been lost at least six times (Oxelman et al., 2004). Of the eudicots considered here, only *Platanus* and possibly *Vitis* diverged before the duplication took place, although all remaining taxa appear to have retained the same copy of *RPB2*, based on the gene-tree presented by Oxelman et al. (2004). However, if one or more of *Arabidopsis, Dianthus, Spinacea,* and *Solanum* actually possess a copy of *RPB2* that is non-orthologous to the copy that all other taxa in this study possess, and that was the copy I obtained, it could have resulted in the apparent non-monophyly of the eudicots, as depicted in Figs. 2.5 and 2.6. Further taxon sampling for *RPB2*, or the use of Southern blots to search for multiple copies, may indicate whether there are more, as of yet undetected, duplication and extinction events involving *RPB2* in the angiosperms.

The maximum-likelihood tree (Fig. 2.8), however, does indicate that the eudicots sampled here are monophyletic. Thus, the maximum likelihood analysis suggests that the latter result may be also be an analytical artifact (rather than mistaken orthology assignment), as they find the monocots to be monophyletic, congruent with all recent work (reviewed in Graham et al., submitted), albeit with *Hordeum* depicted as the sister group of *Acorus* (an unlikely result). *Hordeum* belongs to the grass family, Poaceae, well known to have highly elevated substitution rates in plastid and nuclear genes (e.g., Gaut, 1992, 1996; Graham et al., submitted). Angiosperm phylogeny inferred here using maximum likelihood analysis is generally congruent with relationships that are now well accepted (e.g., the relative positions of *Platanus* and *Vitus*, Fig. 2.8, another strong conflict with the Bayesian analysis; Fig. 2.6). It would be valuable to estimate branch support in a likelihood context using bootstrap analysis, but unfortunately this is not computationally tractable with the current data set.

Major conclusions.

With the exception of the anthophyte hypothesis, type 1 error rates for different hypotheses involving Gnetales placement were generally low using the maximum parsimony criterion, and were decreased by adding more data or by using maximum likelihood inference. With the exception of the "Gnetales basal in seed plants" topology, type 2 error rates were also generally low for the hypotheses considered here. Since the SH tests support a close relationship between Gnetales and conifers, and analysis of the different codon position partitions consistently find the "gnepine" hypothesis, this lends weight to a placement of Gnetales as the sister group of Pinaceae, or at least to conifers as a whole. SH tests reject all alternative hypotheses examined concerning Psilotaceae placement, but the large error rates observed in the simulation results should make us wary about the strongly supported realtionship observed here (and elsewhere) between the whisk ferns and moonworts. Relationships among conifer families (ignoring the problem of Gnetales) and fern families seen here match those found in a variety of other studies, supporting the utility of this locus for inferring relationships in these taxa. Relationships within the angiosperms are more problematic. They are generally only weakly supported, possibly a function of low taxon density coupled with long terminal branches and short internal branches, but some strongly supported clades in the Bayesian analysis conflict with accepted ideas of relationships. However, a tree inferred using maximum likelihood analysis is largely consistent with other current studies (e.g., concerning most aspects of basal angiosperms relationship, and the monophyly of eudicots and monocots), suggesting that at least some of the conflicts observed in

the Bayesian analysis are artifacts of that analysis method rather than the result of mistaken gene orthology. Finally, despite some problems in interpretation (particularly in the flowering plants), the *RPB2* locus has again demonstrated its broad utlility for making inferences about deep vascular-plant relationships. The gene should be a valuable addition to the relatively limited tool box of nuclear genes that are currently available for inferring deep plant phylogeny.

References

Angiosperm Phylogeny Group II. 2003. An update of the angiosperm phylogeny group classification for the orders and families of flowering plants: APG II. *Botanical Journal of the Linnean Society* **141**: 399–436.

Barkman, T. J., Chenery, G., McNeal, J. R., Lyons-Weiler, J., Ellisens, W. J., Moore, G., Wolfe, A. D., and dePamphilis, C. W. 2000. Independent and combined analysis of sequences from all three genomic compartments converge on the root of flowering plant phylogeny. *Proceedings of the National Academy of Sciences USA* **97**: 13166–13171.

Barry, D., and Hartigan, J. A. 1987. Asynchronous distance between homologous DNA sequences. *Biometrics* **43**: 261–276.

Bierhorst, D. W. 1971. *Morphology of Vascular Plants*. Macmillan, New York

Bierhorst, D. W. 1977. The systematic positions of *Psilotum* and *Tmesipteris*. *Brittonia* **29**: 3–13.

Bogler, D. J., and Francisco-Ortega, J. In Press. Molecular systematic studies in cycads: Evidence from *trn*L intron and ITS2 sequences. *Memoirs of the New York Botanical Garden.*

Bowe, L. M., Coat, G., and dePamphilis, C. W. 2000. Phylogeny of seed plants based on all three genomic compartments: Extant gymnosperms are monophyletic and Gnetales' closest relatives are conifers. *Proceedings of the National Academy of Sciences USA* **97:** 4092–4097.

Brunsfeld, S. J., Soltis, P. S., Soltis, D. E., Gadek, P. A., Quinn, C. J., Strenge, D. D., and Ranker, T. A. 1994. Phylogenetic relationships among the genera of Taxodiaceae and Cupressaceae: Evidence from *rbc*L sequences. *Systematic Botany* **19**: 253–262.

Chang, J. T. 1996. Inconsistency of evolutionary tree topology reconstruction methods when substitution rates vary across characters. *Mathematical Biosciences* **134**: 189–215.

Chaw, S-M., Sung, H-M., Long, H., Zharkikh, A., and Li, W-H. 1995. The phylogenetic positions of the conifer genera *Amenotaxus*, *Phyllocladus* and *Nageia* inferred from 18S rDNA sequences. *Journal of Molecular Evolution* **41**: 224–230.

Chaw, S., Zharkikh, A., Sung, H., Lau, T., and Li, W. 1997. Molecular phylogeny of extant gymnosperms and seed plant evolution: Analysis of 18S rRNA sequences. *Molecular Biology and Evolution* **14**: 56–68.

Chaw, S-M., Parkinson, C. L., Cheng, Y., Vincent, T. M., and Palmer, J. D. 2000. Seed plant phylogeny inferred from all three plant genomes: Monophyly of extant gymnosperms and origin of Gnetales from conifers. *Proceedings of the National Academy of Sciences USA* **97:** 4086–4091.

Cheng, Y., Nicolson, R. G., Tripp, T., and Chaw, S-M. 2000. Phylogeny of Taxaceae and Cephalotaxaceae genera inferred from chloroplast *mat*K gene and nuclear rDNA ITS region. *Molecular Phylogenetics and Evolution* **14**: 353–365.

Crane, P. R. 1985. Phylogenetic relationships in seed plants. *Cladistics* **1**: 329–348.

Denton, A. L., McConaughy, B. L., and Hall, B. D. 1998. Usefulness of RNA polymerase II coding sequences for estimation of green plant phylogeny. *Molecular Biology and Evolution* **15**: 1082–1085.

Dombrovska, O., and Qiu, Y-L. 2004. Distribution of introns in the mitochondrial gene *nad*1 in land plants: phylogenetic and molecular evolutionary implications. *Molecular Phylogenetics and Evolution* **32**: 246-263.

Donoghue, M. J., and Doyle, A. D. 2000. Seed plant phylogeny: Demise of the anthophyte hypothesis? *Current Biology* **10**: R106–R109.

Doyle, J. A. 1996. Seed plant phylogeny and the relationships of Gnetales. *International Journal of Plant Sciences* **157** S3–S9.

Doyle, J. A., and Donoghue, M. J. 1986. Seed plant phylogeny and the origin of angiosperms: an experimental cladistic approach. *Botanical Reviews* **52**: 321–431.

Doyle, J. A., and Donoghue, M. J. 1992. Fossils and seed plant phylogeny reanalyzed. *Brittonia* **44:** 89–106.

Duff, R. J. and Nickrent, D. L. 1999. Phylogenetic relationships of land plants using mitochondrial small-subunit rDNA sequences. *American Journal of Botany* **86**: 372–386.

Eckenwalder, J. E. 1976. Re-evaluation of Cupressaceae and Taxodiaceae: a proposed merger. *Madrono* **23**: 237–300.

Felsenstein, J. 1978. Cases in which parsimony or compatibility methods will be positively misleading. *Systematic Zoology* **27**: 401–410.

Felsenstein, J., 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. *Journal of Molecular Evolution* **17**: 368–376.

Felsenstein, J. 1983. Parsimony in systematics: Biological and statistical issues. *Annual Review of Ecology and Systematics* **14**: 313–333.

Felsenstein, J. 1985. Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* **39**: 783–791.

Felsenstein, J. 2004. *Inferring Phylogenies.* Sinauer Associates, Sunderland, MA.

Gaut, B. S., Muse, S. V., Clark, W. D., and Clegg, M. T. 1992. Relative rates of nucleotide substitution at the *rbc*L locus of monocotyledonous plants. *Journal of Molecular Evolution* **35**: 292–303.

Gaut, B. S, Morton, B. R., McCaig, B. C., and Clegg, M. T. 1996. Synonymous rate comparisons between grasses and palms: synonymous rate differences at the nuclear gene *Adh* parallel rate difference at the plastid gene *rbcL*. *Proceedings of the National Academy of Sciences USA* **93**: 10274–10279.

Gensel, P. G. 1977. Morphologic and taxonomic relationships of the Psilotaceae relative to the evolutionary lines in early land vascular plants. *Brittonia* **29**: 14–29.

Goldman, N., Anderson, J. P., and Rodrigo, A. G. 2000. Likelihood-based tests of topologies in phylogenetics. *Systematic Biology* **49**: 652–670.

Goremykin, V., Bobrova, B., Pahnke, J., Troitsky, A., Antonov, A., and Martin, W. 1996. Noncoding sequences for the slowly evolving chloroplast inverted repeat do not support gnetalean affinities of angiosperms. *Molecular Biology and Evolution* **13**: 383–396.

Goremykin, V. V., Hirsch-Ernst, K. I., Wölf, S., and Hellwig, F. H. 2003. Analysis of the *Amborella trichopoda* chloroplast genome sequence suggests that *Amborella* is not a basal angiosperm. *Molecular Biology and Evolution* **20**: 1499–1505

Graham, S. W. and Olmstead, R. G. 2000. Utility of chloroplast genes for inferring the phylogeny of the basal angiosperms. *American Journal of Botany* **87**: 1712–1730.

Graham, S. W., Reeves, P. A., Burns, A. C. E., and Olmstead, R. G. 2000. Microstructural changes in noncoding chloroplast DNA: interpretation, evolution, and utility of indels and inversions in basal angiosperm phylogenetic inference. *International Journal of Plant Sciences* **161**: S83–S96.

Graham, S. W., Olmstead, R. G., and Barrett, S. C. H. 2002. Rooting phylogenetic trees with distant outgroups: A case study from the Commelinoid monocots. *Molecular Biology and Evolution* **19**: 1769–1781.

Graham, S. W., Zgurski, J. M., McPherson, M. A., Cherniawsky, D. M., Saarela, J. M., Horne, E. S. C., Smith, S. Y., Wong, W. A., O'Brien, H. E., Biron, V. L., Pires, J. C., Olmstead, R. G., Chase, M. W., and Rai, H. S. Submitted. Robust inference of monocot deep phylogeny using an expanded multigene plastid data set.

Guindon, S., and Gascuel, O. 2003. PHYML - A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology* **52**: 696–704.

Gurgerli, F., Sperisen, C., Buchler, U., Brunner, I., Brodbeck, S., Palmer, J. D., and Qiu, Y. L. 2001. The evolutionary split of Pinaceae from other conifers: evidence from an intron loss and a multigene phylogeny. *Molecular Phylogenetics and Evolution* **21**: 167–175.

Hart, J. A. 1987. A cladistic analysis of conifers: preliminary results. *Journal of the Arnold Arboretum* **68**: 269–307.

Hasebe, M., Wolf, P. G., Pryer, K. M., Ueda, K., Ito, M., Sano, R., Gastony, G. J., Yokoyama, J., Manhart, J. R., Murakami, N., Crane, E. H., Haufler, C. H., and Hauk, W. D. 1995. Fern phylogeny based on *rbc*L nucleotide sequences. *American Fern Journal* **85**: 134–181.

Hasegawa, M. H., Kishino, H., and Yano, T. 1985. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *Journal of Molecular Evolution* **22**: 160–174.

Hendy, M. D., and Penny, D. 1989. A framework for the quantitative study of evolutionary trees. *Systematic Zoology* **38**: 297–309.

Hernandez-Castillo, G. R., Rothwell, G. W., and Mapes, G. 2001. Compound pollen cone in a Paleozoic conifer. *American Journal of Botany* **88**: 1139–1142.

Hill, K. D., Chase, M. W., Stevenson, D. W., Hills, H. G., and Schutzman, B. 2003. The families and genera of cycads: A molecular phylogenetic analysis of cycadophyta based on nuclear and plastid DNA sequences. *International Journal of Plant Sciences* **164**: 933-948.

Hillis, D. M. 1996. Inferring complex phylogenies. *Nature* **383**: 130-131.

Hillis, D. M. 1998. Taxonomic sampling, phylogenetic accuracy and investigator bias. *Systematic Biology* **47:** 3–8.

Hilu, K. W., Borsch, T., Müller, K., Soltis, D. E., Soltis, P. S., Savolainen, V., Chase, M. W., Powell, M. P., Alice, L. A., Evans, R., Sauquet, H., Neinhuis ,C., Slotta, T. A. B., Rohwer, J. G., Campbell, C. S., and Chatrou, L. W. 2003. Angiosperm phylogeny based on *matK* sequence information. *American Journal of Botany* **90**: 1758–1776.

Huelsenbeck, J. 1995. Performance of phylogenetic methods in simulation. *Systematic Biology* **44**: 17–48.

Huelsenbeck, J. P. 1998. Systematic bias in phylogenetic analysis: Is the Strepsiptera problem solved? *Systematic Biology* **47**: 519–537.

Huelsenbeck, J. P., and Crandall, K. A. 1997. Phylogeny estimation and hypothesis testing using maximum likelihood. *Annual Review of Ecology and Systematics* **28**: 437–466.

Iwabe, N., Kuma, K., Kishino, H., Hasegawa, M., and Miyata, T. 1991. Evolution of RNA polymerases and branching patterns of the three major groups of archaebacteria. *Journal of Molecular Evolution* **32**: 70–78.

Judd, W. S., Kellogg, E. A., Stevens, P. F., and Donoghue, M. J. 2002. *Plant Systematics: A Phylogenetic Approach, second edition.* Sinauer Associates, Sunderland, MA.

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

Kenrick, P., and Crane, P. R. 1997. *The Origin and Early Diversification of Land Plants: A Cladistic Study.* Smithsonian Institution Press, Washington.

Korall, K., and Kenrick, P. 2002. Phylogenetic relationships in Selaginellaceae based on *rbc*L sequences. *American Journal of Botany* **89**: 506–517.

Korall, K., and Kenrick, P. 2004. The phylogenetic history of Selaginellaceae based on DNA sequences from the plastid and nucleus: Extreme substitution rates and rate heterogeneity. *Molecular Phylogenetics and Evolution* **31**: 852–864.

Kramer, K. U., and Green, P. S., eds. 1990. *The Families and Genera of Vascular Plants, Vol. 1: Pteridophytes and Gymnosperms.* (series ed. K. Kubitzki). Springer-Verlag, Berlin.

Lanave, C., Preparata, G., Saccone, C., and Serio, G. 1984. A new method for calculating evolutionary substitution rates. *Journal of Molecular Evolution* **20**, 86–93.

Larkin, R., and Guilfoyle, T. 1993. The second largest subunit of RNA polymerase II from *Arabidopsis thaliana*. *Nucleic Acids Research* **21**: 1038.

Loconte, H., and Stevenson, D. W. 1990. Cladistics of the Spermatophyta. *Brittonia* **42**: 197–211.

Maddison, D. R., and Maddison, W. P. 2001. MacClade 4: analysis of phylogeny and character evolution. Sinauer Associates, Sunderland, MA.

Maden, A. R., Whittier, D. P., Garbary, D. J. and Renzaglia, K. S. 1997. Ultrastructure of the spermatozoid of *Lycopodiella lateralis* (Lycopodiaceae). *Canadian Journal of Botany* **75**: 1728–1738.

Magallon, S., and Sanderson, M. J. 2002. Relationships among seed plants inferred from highly conserved genes: Sorting conflicting phylogenetic signals among ancient lineages. *American Journal of Botany* **89**: 1991–2006.

Malek, O., Lättig, K., Hiesel, R., Brennicke, A., and Knoop, V. 1996. RNA editing in bryophytes and a molecular phylogeny of land plants. *EMBO Journal* **15**: 1403–1411.

Manhart, J. R. 1995. Chloroplast 16S rDNA sequences and phylogenetic relationships of fern allies and ferns. *American Fern Journal* **85**: 182–192.

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

Nickerson, J., and Drouin, G. 2004. The sequence of the largest subunit of RNA polymerase II is a useful marker for inferring seed plant phylogeny. *Molecular Phylogenetics and Evolution* **31**: 403–415.

Nickrent , D. L., C. L. Parkinson, J. D. Palmer, and Duff, R. J. 2000. Multigene phylogeny of land plants with special reference to bryophytes and the earliest land plants. *Molecular Biology and Evolution* **17**: 1885–1895..

Nixon, K. C., Crepet, W. L., Stevenson, D. and Friis, E. M. 1994. A reevaluation of seed plant phylogeny. *Annals of the Missouri Botanical Garden* **81**: 484–533.

Oxelman, B., and Bremer, B. 2000. Discovery of paralogous nuclear gene sequences coding for the second-largest subunit of RNA polymerase II (*RPB*II) and their phylogenetic utility in Gentianales of the asterids. *Molecular Biology and Evolution* **17**: 1131–1145.

Oxelman B., Yoshikawa, N., McConaughy, B. L., Luo, J., Denton, A. L., and Hall, B. D. 2004. *RPB*2 gene phylogeny in flowering plants, with particular emphasis on asterids. *Molecular Phylogenetics and Evolution* **32**: 462–479.

Page, C. N. 1990a. Cephalotaxaceae. pp. 299-302. In: Kramer, K. U., and Green, P. S., eds. 1990. *The Families and Genera of Vascular Plants, Vol. 1: Pteridophytes and Gymnosperms*. (series ed. K. Kubitzki). Springer-Verlag, Berlin.

Page, C. N. 1990b. Taxodiaceae. pp. 353-361. In: Kramer, K. U., and Green, P. S., eds. 1990. *The Families and Genera of Vascular Plants, Vol. 1: Pteridophytes and Gymnosperms*. (series ed. K. Kubitzki). Springer-Verlag, Berlin.

Parkinson, C. L., Adams, K. L., and Palmer, J. D. 1999. Multigene analyses identify the three earliest lineages of extant flowering plants. *Current Biology* **9**: 1485–1488.

Poe, S., and Swofford, D. L. 1999. Taxon sampling revisited. *Nature* **389**: 299–300.

Pryer, K. M., Schneider, H., Smith, A. R., Cranfill, R., Wolf, P. G., Hunt, J. S., and Sipes, S. D. 2001. Horsetails and ferns are a monophyletic group and the closest living relatives to seed plants. *Nature* **409**: 618–622.

Qiu, Y. –L., Lee, J., Bernasconi-Quadroni, F., Soltis, D. E., Soltis, P. S., Zanis, M., Zimmer, E. A., Chen, Z., Savolainen, V., and Chase, M. W. 1999. The earliest angiosperms: evidence from mitochondrial, plastid and nuclear genomes. *Nature* **402**: 404–407.

Quinn, C. J., Price, R. A., and Gadek, P. A. 2002. Familial concepts and relationships in the conifers based on *rbc*L and *mat*K sequence comparisons. *Kew Bulletin* **57**: 513–531.

Rai, H. S., O'Brien, H. E., Reeves, P. A., Olmstead, R. G., and Graham, S. W. 2003. Inference of higher-order relationships in the cycads from a large chloroplast data set. *Molecular Phylogenetics and Evolution* **29**: 350–359.

Rambaut, A., 1998. "Se-Al (Sequence Alignment Editor Version 1.0)," Computer program and documentation. Department of Zoology, University of Oxford, UK.

Rambaut, A. and Grassly, N. C. 1997. Seq-Gen: an application for the Monte Carlo simulation of DNA sequence evolution along phylogenetic trees. *Computational and Applied Biosciences* **13:** 235–238.

Rannala, B., J. P. Huelsenbeck, Z. Yang, and Nielsen, R. 1998. Taxon sampling and the accuracy of large phylogenies. *Systematic Biology* **47:** 702–710.

Raubeson, L. A., and Jansen, R. K. 1992a. Chloroplast DNA evidence on the most ancient split in vascular land plants. *Science* **85**: 1697–1699.

Raubeson, L. A., and Jansen, R. K. 1992b. A rare chloroplast-DNA structural mutation is shared by all conifers. *Biochemical Systematics and Ecology* **20**: 17–24.

Renzaglia, K. S., Duff, R. J., Nickrent, D. L., and Garbary, D. J. 2000 Vegetative and reproductive innovations of early land plants: Implications for a unified phylogeny. *Philosophical Transactions of the Royal Society of London, Series B* **355**: 769–793.

Rodríguez, F., Oliver, J. L., Marín, A., and Medina, J. R., 1990. The general stochastic model of nucleotide substitution. *Journal of Theoretical Biology* **142**: 485–501.

Ronquist, F., and Huelsenbeck, J. P. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**: 1572–1574.

Rothwell, G. W. 1999. Fossils and ferns in the resolution of land plant phylogeny. *Botanical Review* **65**:188–218.

Rothwell, G.W. and Serbet, R. 1994. Lignophyte phylogeny and the evolution of spermatophytes: a numerical cladistic analysis. *Systematic Botany* **19**: 443–482.

Rothwell, G. W., and Stockey, R. A. 1989. Fossil Ophioglossales in Paleocene of western *North America. American Journal of Botany* **76**: 637–644.

Rydin, C., Kallersjo, M., and Friis, E. M. 2002. Seed plant relationships and the systematic position of Gnetales based on nuclear and chloroplast DNA: conflicting data, rooting problems and the monophyly of conifers. *International Journal of Plant Science* **163**: 197–214.

Sanderson, M. J., Wojciechowski, M. F., Hu, J. M., Sher Khan, T., and Brady, S. G. 2000. Error, bias and long-branch attraction in data for two chloroplast photosystem genes in seed plants. *Molecular Biology and Evolution* **17**: 782–797.

Schmidt, M. and Schneider-Poetsch, H. A. W. 2002. The evolution of gymnosperms redrawn by phytochrome genes: The Gnetatae appear at the base of the gymnosperms. *Journal of Molecular Evolution* **54**: 715–724.

Shimodaira, H., and Hasegawa, M. 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Molecular Biology and Evolution* **16**: 1114–1116.

Sidow, A., and Thomas, W. K. 1994. A molecular evolutionary framework for eukaryotic model organisms. *Current Biology* **4**: 596–603.

Soltis, D. E., and Soltis, P. S. 2004. *Amborella* not a "basal angiosperm"? Not so fast. *American Journal of Botany* **91**: 997–1001.

Soltis, P. S., Soltis, D. E., and Chase, M. W. 1999. Angiosperm phylogeny inferred from multiple genes as a tool for comparative biology. *Nature* **402**: 402–403.

Soltis, D. E., Soltis, P. S., and Zanis, M. J. 2002. Phylogeny of seed plants based on evidence from eight genes. *American Journal of Botany* **89**: 1670–1681.

Soltis, D. E., Soltis, P. S., Chase, M. W., Mort, M. E., Albach, D. C., Zanis, M., Savolainen, V., Hahn, W. H., Hoot, S. B., Fay, M. F., Axtell, M., Swensen, S. M., Prince, L. M., Kress, W. J., Nixon, K. C., and Farris, J. S. 2000. Angiosperm phylogeny inferred from 18S rDNA, *rbcL*, and *atpB* sequences. *Botanical Journal of the Linnean Society* **133**: 381–461.

Stefanovic, S., Jager, M., Deutsch, J., Broutin, J., and Masselot, M. 1998. Phylogenetic relationships of conifers inferred from partial 28S rRNA gene sequences. *American Journal of Botany* **85**: 688–697.

Stevenson, D. W. 1990. Morphology and systematics of the Cycadales. *Memoirs of the New York Botanical Garden* **57**: 8–55.

Stevenson, D.W. 1992. A formal classification of the extant cycads. *Brittonia* **44**: 220–223.

Stevenson, D. W., and Loconte, H. 1996. Ordinal and familial relationships of pteridophyte genera. Pp. 435–467 in J. M. Camus, M. Gibby and R. J. Johns (eds) *Pteridology in Perspective.* Royal Botanic Gardens, Kew.

Stewart, W. N., and Rothwell, G. W. 1993. *Paleobotany and the Evolution of Plants*, second edition. Cambridge University Press, Cambridge, United Kingdom.

Sullivan, J. and Swofford, D. L. 1997. Are guinea pigs rodents? The importance of adequate models on molecular phylogenetics. *Journal of Mammalian Evolution* **4**: 77–86.

Swofford, D. L., 2002. "*PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4.0b10*," Computer program and documentation. Sinauer Associates, Sunderland, MA.

Tavaré, S., 1986. Some probabilistic and statistical problems on the analysis of DNA sequences. *Lectures in Mathematics and Life Sciences* **17**: 57–86.

Treutlein, J., and Wink, M. 2002. Molecular phylogeny of cycads inferred from *rbc*L sequences. *Naturwissenschaften* **89**: 221–225.

Wagner, W. H. Jr. 1977. Systematic implications of the Psilotaceae. *Brittonia* **29**: 54–63.

Wakasugi, T., Tsudzuki, J., Ito, S., Nakashima, K., Tsudzuki, T. and Sugiura, M. 1994. Loss of all *ndh* genes as determined by sequencing the entire chloroplast genome of the black pine *Pinus thunbergii*. *Proceedings of the National Academy of Sciences USA.* **91**: 9794–9798.

Warrilow, D., and Symons, R. H. 1996. Sequence analysis of the second-largest subunit of tomato RNA polymerase II. *Plant Molecular Biology* **30**: 337–342.

Winter, K.-U., Becker, A., Munster, T., Kim, J. T., Saedler, H., and Theissen, G. 1999. MADS-box genes reveal that gnetophytes are more closely related to conifers than to flowering plants. *Proceedings of the National Academy of Sciences USA* **96**: 7342–7347.

Wolf, P. G. 1997. Evaluation of *atpB* nucleotide sequences for phylogenetic studies of ferns and other pteridophytes. *American Journal of Botany* **84**: 1429–1440.

Yang, Z. 1994. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: Approximate methods. *Journal of Molecular Evolution* **39**: 306–314.

Zanis, M. J., Soltis, D. E., Soltis, P. S., Mathews, S., and Donoghue, M. J. 2002. The root of angiosperms revisited. *Proceedings of the National Academy of Sciences USA* **99**: 6848–6853.

Zwickl, D. J., and Hillis, D. M.. 2002. Increased taxon sampling greatly reduces phylogenetic error. *Systematic Biology* **51**: 588–598.

Table 2.1. Taxon names, source information, and GenBank accession numbers for plants used in this study. Voucher abbreviations refer to standard herbarium acronyms, unless otherwise stated. In cases where no voucher exists, the location where the sample was collected is noted.[1]

Species	Family [2]	Voucher / Source Information	Genbank #
Bryophytes			
Marchantia polymorpha L.	Marchantiaceae	Denton et al. (1998)	AF020844
Lycopods			
Isoetes L. sp.	Isoetaceae	JZ73, ALTA	AY699180
Lycopodium annotinum L.	Lycopodiaceae	14-09-02-13, ALTA	AY699181
Selaginella densa Rydb.	Selaginellaceae	Oxelman et al. (2004)	AY563264
Psilophytes			
Psilotum nudum (L.) P. Beauv.	Psilotaceae	UW Greenhouse	AY699182
Arthrophytes			
Equisetum hymenale L.	Equisetaceae	14-09-02-2 ALTA.	AY699179
Eusporangiate ferns			
Botrychium virginianum (L.) Sw.	Ophioglossaceae	14-09-02-3, ALTA	AY699177[3]
Ophioglossum vulgatum L.	Ophioglossaceae	JZ69, ALTA	AY699178
Leptosporangiate ferns			
Adiantum tenerum Sw.	Pteridaceae	JZ71, ALTA	AY699183[4]
Anemia phyllitidis (L.) Sw.	Schizeaceae	JZ72, ALTA	AY699184
Cryptogramma crispa (L.) R. Br. ex Hook.	Pteridaceae	14-09-02-10, ALTA	AY699185
Dicksonia antarctica Labill.	Dicksoniaceae	JZ68, ALTA	AY699186
Dryopteris filix-mas (L.) Schott	Dryopteridaceae	14-09-02-8, ALTA	AY699187[4]
Marsilea quadrifolia L.	Marsileaceae	JZ64, ALTA	AY699188
Polystichum lonchitis (L.) Roth	Dryopteridaceae	14-09-02-6, ALTA	AY699189
Stenochlaena tenuifolia (Desv.) Moore	Blechnaceae	JZ63, ALTA	AY699190[4]
Cycads			
Ceratozamia miqueliana H. Wendl	Zamiaceae	84328C, FTG	AY699191
Cycas revoluta Thunb.	Cycadaceae	Oxelman et al. (2004)	AY563265
Dioon edule Lindl.	Zamiaceae	UW Greenhouse	AY699197
Encephalartos barteri Carruth. ex. Miquel	Zamiaceae	O'Brien 1002, ALTA	AY699192
Macrozamia moorei F. Muell	Zamiaceae	59302, FTG	AY699193[3]
Microcycas calocoma (Miq.) A. DC.	Zamiaceae	77404T, FTG	AY699194
Stangeria eriopus Nash	Stangeriaceae	651325N, FTG	AY699195
Zamia floridana A. DC.	Zamiaceae	JZ62, ALTA	AY699196
Ginkgoales			
Ginkgo biloba L.	Ginkgoaceae	Denton et al. (1998)	AF020843

Species	Family	Voucher / Source Information	Genbank #
Gnetales			
Ephedra distachya L.	Ephedraceae	JZ61, ALTA	AY699198
Gnetum gnemon L.	Gnetaceae	Oxelman et al. (2004)	AY563267
Welwitschia mirabilis Hook. f.	Welwitschiaceae	UW Greenhouse	AY699199
Conifers			
Agathis australis (D. Don) Salisb.	Araucariaceae	JZ78, ALTA	AY699200
Araucaria araucana (Molina) K. Koch	Araucariaceae	JZ67, ALTA	AY699201
Cephalotaxus fortunei Hook.	Cephalotaxaceae	UW Greenhouse	AY699209
Cunninghamia lanceolata (Lamb.) Hook.	Cupressaceae	UW Greenhouse	AY699208
Dacrydium cupressinum Soland. ex.Forst. f.	Podocarpaceae	JZ75, ALTA	AY699202[3]
Juniperus communis L.	Cupressaceae	JZ77, ALTA	AY699210
Larix occidentalis Nutt.	Pinaceae	UW Greenhouse	AY699214
Metasequoia glyptostroboides Hu & W.C. Cheng	Cupressaceae	UW Greenhouse	AY699213
Podocarpus coriaceus Rich. & A. Rich.	Podocarpaceae	JZ65, ALTA	AY699203
Pseudotsuga menziesii (Mirb.) Franco	Pinaceae	UW Greenhouse	AY699215
Saxegothaea conspicua Lindl.	Podocarpaceae	JZ76, ALTA	
Sciadopitys verticillata (Thunb.) Siebold & Zucc.	Sciadopityaceae	Graham & Denton VII-98-1, WTU	AY699205
Sequoiadendron giganteum (Lindl.) J. Buchholz	Cupressaceae	UW Greenhouse	AY699212
Sequoia sempervirens (D. Don) Endl.	Cupressaceae	Oxelman et al. (2004)	AY563266
Taxus brevifolia Nutt.	Taxaceae	UW Greenhouse	AY699207
Thuja plicata Donn. ex. *D. Don.*	Cupressaceae	JZ70, ALTA	AY699211
Torreya californica Torr.	Taxaceae	UW Greenhouse	AY699206
Angiosperms			
Acorus gramineus Aiton	Acoraceae	JZ74, ALTA	AY699222
Amborella trichopoda Baill.	Amborellaceae	RBG, Sydney	AY699216
Arabidopsis thaliana (L.) Heynh.	Brassicaceae	Larkin and Guilfoyle (1993)	Z19120
Aristolochia gigantea Mart. & Zucc.	Aristolochiaceae	Denton et al. (1998)	AF020842
Asarum caudatum Lindl.	Aristolochiaceae	UW Greenhouse	AY699219
Chloranthus spicatus (Thunb.) Makino	Chloranthaceae	Denton et al. (1998)	AF041852
Dianthus L. sp.	Caryophyllaceae	UW Greenhouse	AY699221
Dioscorea sansibarensis Pax.	Dioscoreaceae	Oxelman et al. (2004)	AY563268
Drimys winteri J. R. Forst. & G. Forst.	Winteraceae	UW Greenhouse	AY699218
Hordeum vulgare L.	Poaceae	Denton et al. (1998)	AF020839
Illicium anisatum L.	Illicaceae	UW Greenhouse	AY699220
Magnolia virginiana L.	Magnoliaceae	Denton et al. (1998)	AF020841
Musa velutina H. Wendl. & Drude	Musaceae	UW Greenhouse	AY699223
Nymphaea odorata Aiton	Nymphaeaceae	Denton et al. (1998)	AF043427
Peperomia caperata Yunck.	Piperaceae	Denton et al. (1998)	AF043426
Platanus L. sp.	Plantanaceae	Oxelman et al. (2004)	AY566618

Species	Family	Voucher / Source Information	Genbank #
Angiosperms (contd.)			
Solanum lycopersicon L.	Solanaceae	Warrilow and Symons (1996)	U28403
Spinacea oleracea L.	Amaranthaceae	Denton et al. (1998)	AF020840
Trimenia moorei (Oliv.) Philipson	Trimeniaceae	P Weston 433770 (NSW)	AY699217
Vitis piasezkii Maxim.	Vitaceae	Oxelman et al. (2004)	AJ556992

1 Abbreviations: U W = University of Washington, RBG = Royal Botanic Gardens

2 Classification schemes used: Angiosperms, APG II (2003); cycads, Stevenson (1992); all others, Kramer and Green (1990), except I follow the recommendation of Eckenwalder (1976) and Stefanovic et al. (1998) and place all Taxodiaceae into Cupressaceae *s.l.*

3 Sequence is missing the first 100-125 bp.

4 Sequence is missing the last 200-250 bp.

Table 2.2. Likelihood ratio test (LRT) for various substitution models. Likelihood scores and parameter estimates are based on the MP topology presented in Fig. 2.1[1]. In all cases, $P < 0.01$.

Substitution Model [2]	**-ln likelihood**	**Comparison**	**-2 ln Λ** [3]	**d.f.**
JC69	42286.74	-----	-----	
F81	42228.15	JC69 vs. F81	58.59094	3
HKY85	41167.69	F81 vs. HKY85	1060.515	1
GTR	40928.71	HKY85 vs. GTR	238.926	5
GTR + Γ	34858.70	GTR vs. GTR + Γ	6070.017	1
GTR + Γ + I	34830.80	GTR + Γ vs. GTR + Γ + I	27.89206	1

[1]The significance value for rejection of the null hypothesis was adjusted using a Bonferroni correction and set to 0.01.

[2] Abbreviations: JC69 = Jukes-Cantor (1969); F81 = Felsenstein (1981); HKY85 = Hasegawa et al. (1985); GTR = General Time-Reversible (Lanave et al., 1984; Tavaré, 1986; Barry and Hartigan, 1987; Rodríguez et al., 1990). Γ = Gamma. I= proportion of invariable sites.

[3] Likelihood-ratio test statistic.

Table 2.3. Estimated model parameters (GTR + Γ + I) for various partitions of the *RPB2* data set derived from the ML tree depicted in Fig. 2.8 [1].

	All codon positions	Codon positions 1 & 2	Codon position 3
Base frequencies			
A	0.295403	0.310012	0.262005
C	0.234440	0.245402	0.202803
G	0.209260	0.210436	0.202345
T	0.260897	0.234150	0.332846
Rate matrix			
AC	1.07306	3.10539	1.99867
AG	2.82834	3.66045	7.19996
AT	1.62784	1.38492	2.72307
CG	0.52548	1.33166	0.75634
CT	4.01450	4.90830	7.52130
GT	1.00000	1.00000	1.00000
Γ	0.280887	0.185194	1.58389
I	0.138020	0.227586	0.038416

1. Abbreviations: Γ = gamma shape parameter; I = proportion of invariable sites.

Table 2.4. Error rates estimated using maximum parsimony (MP) and likelihood (ML) for each of five hypotheses of seed-plant relationship. The hypothesis names on each row represent the model tree from which the data were simulated (shown in Fig. 2.2); those heading columns are the hypotheses inferred from simulated data sets. All details of relationship were ignored for scoring, apart from the hypothesis under consideration (i.e., constrained branches in Fig. 2.2).

	Gnepine	**Gnetales basal (in gymnosperms)**	**Anthophyte**	**Gnetales basal (in seed plants)**	**Gnetifer**
MP (1,773 bp)					
Gnepine	**0.99**	-	-	-	0.01
Gnetales basal (in gymnosperms)	-	**0.97**	-	-	0.01
Anthophyte	-	0.16	**0.43**	0.41	-
Gnetales basal (in seed plants)	-	0.06	0.05	**0.89**	-
Gnetifer	0.20	-	-	-	**0.71**
MP (5,319 bp)					
Gnepine	**1.00**	-	-	-	-
Gnetales basal (in gymnosperms)	-	**1.00**	-	-	-
Anthophyte	-	0.07	**0.53**	0.40	-
Gnetales basal (in seed plants)	-	-	-	**1.00**	-
Gnetifer	0.04	-	-	-	**0.96**
MP (10 kb)					
Gnepine	**1.00**	-	-	-	-
Gnetales basal (in gymnosperms)	-	**1.00**	-	-	-
Anthophyte	-	0.05	**0.52**	0.43	-
Gnetales basal (in seed plants)	-	-	-	**1.00**	-
Gnetifer	-	-	-	-	**1.00**
ML (1,773 bp)					
Gnepine	**1.00**	-	-	-	-
Gnetales basal (in gymnosperms)	-	**0.98**	-	-	-
Anthophyte	-	0.13	**0.80**	0.07	-
Gnetales basal (in seed plants)	-	0.04	0.05	**0.91**	-
Gnetifer	0.08	-	-	-	**0.89**

Table 2.5. Error rates estimated using maximum parsimony (MP) and likelihood (ML) for each of four hypotheses of Psilotaceae relationship. The hypothesis names on each row represent the model tree from which the data were simulated (shown in Fig. 2.3); those heading columns are the trees inferred from simulated data sets. All details of relationship were ignored for scoring, apart from the hypothesis under consideration (i.e., the constrained branches in Fig. 2.3).

	a	b	c	d
MP (1,773 bp)				
a	**1.00**	0	0	0
b	0.01	**0.16**	0.15	0.02
c	0.44	0	**0.24**	0.32
d	0.30	0	0.34	**0.36**
MP (5,319 bp)				
a	**1.00**	0	0	0
b	0	**0.24**	0.12	0
c	0.32	0	**0.37**	0.31
d	0.41	0	0.22	**0.37**
MP (10 kb)				
a	**1.00**	0	0	0
b	0	**0.29**	0.06	0
c	0.46	0	**0.25**	0.29
d	0.51	0	0.21	**0.28**
ML (1,773 bp)				
a	**1.00**	0	0	0
b	0.05	**0.2**	0.09	0.01
c	0.3	0	**0.32**	0.34
d	0.31	0	0.28	**0.37**

Figure Legends

Figure 2.1. A map of primers used to amplify and sequence *RPB2* from cDNA for: (a) Seed plants (exons 12-24, 1,662 bp) and (b) Pteridophytes (exons 11-24, 1,637 bp). Exons are numbered following Larkin and Guilfoyle (1993) (see also Denton et al., 1998). Lighter areas indicate intron positions; introns and primers are not drawn to scale. All primer sequences are shown in 5'- to 3'-orientation; ambiguous nucleotides follow the IUBMB code. "F" and "R" indicate forward and reverse primers, respectively. A scale-bar is shown below each map.

Figure 2.2. Five major hypotheses of Gnetales relationship: (a) The "gnepine hypothesis" of Bowe et al. (2000) and Chaw et al. (2000); (b) The "Gnetales basal in gymnosperms hypothesis," with angiosperms and Gnetales depicted as successive sister taxa of the remaining seed plants (see Schmidt et al., 2002); (c) The "anthophyte hypothesis" (see Crane, 1985; Doyle and Donoghue, 1986, 1992; Loconte and Stevenson, 1990; Doyle, 1996); (d) The "Gnetales basal in seed plants hypothesis," with Gnetales and angiosperms depicted as successive sister taxa of the remaining seed plants (see Rydin et al., 2002; Rai et al., 2003), and; (e) The "gnetifer hypothesis" (Chaw et al., 1997). Asterisks (corresponding to each hypothesis) indicate the branches constrained in parsimony searches that were used to find the model trees for the simulation study (Table 2.3) and SH tests. The major branches inferred in the resulting constrained tree searches are also shown.

Figure 2.3. Four major hypotheses of Psilotaceae relationship (a–d). Asterisks (corresponding to each major hypothesis) indicate the branches

constrained in parsimony searches that were used to find the model trees for the simulation study (Table 2.4) and SH tests. The major branches inferred in the resulting constrained tree searches are also shown.

Figure 2.4. A portion of the nucleotide and amino-acid alignments of *RPB2* exon 16 (starting at bp 6 in the exon) displaying a frame-shifted duplication for a subset of taxa. Shaded areas in the nucleotide alignment display the duplicated portion. Note the shift in reading frame (nucleotides) and resulting change in amino acids in the duplicated portion. A neighbouring multi-taxon deletion in the same is also shown.

Figure 2.5. One of 26 most parsimonious trees found using *RPB2* exon sequences, length = 8,247, CI = 0.203, RI = 0.518. MP bootstrap values are shown above the branches, values below 30% are represented by an asterisk. Branches shown with a dotted line are those that did not appear in a strict consensus tree of all MP trees obtained. The inset shows angiosperm relationships obtained in a similar analysis after *Arabidopsis* (a eudicot), *Aristolochia, Hordeum* (a monocot) and *Peperomia* are eliminated from consideration.

Figure 2.6. Majority-rule consensus of trees sampled in the Bayesian analyses of *RPB2* exon sequences (all codon positions). Numbers above the branches indicate the frequency of recovery of each clade (= posterior probability estimates) when one set of estimated model parameters was used to describe all data subsets (HKY85 + Γ model). The numbers below the branches indicate the posterior probabilities of each clade obtained when the estimated model parameters for codon positions 1 and 2 were estimated separately from those for

codon position 3 for the same general substitution model. Values below 30 are represented by an asterisk.

Figure 2.7. Majority-rule consensus of trees sampled in the Bayesian analyses of *RPB2* exon sequences using the GTR + Γ + I model of sequence evolution: (a) codon positions 1 and 2 only, and; (b) codon position 3 only. Numbers above the branches indicate the frequency of recovery of each clade (= estimated posterior probability).

Figure 2.8. Tree obtained in maximum-likelihood analysis of *RPB2* exon sequences using the HKY + Γ model of sequence evolution (-lnL = 34,836.82).

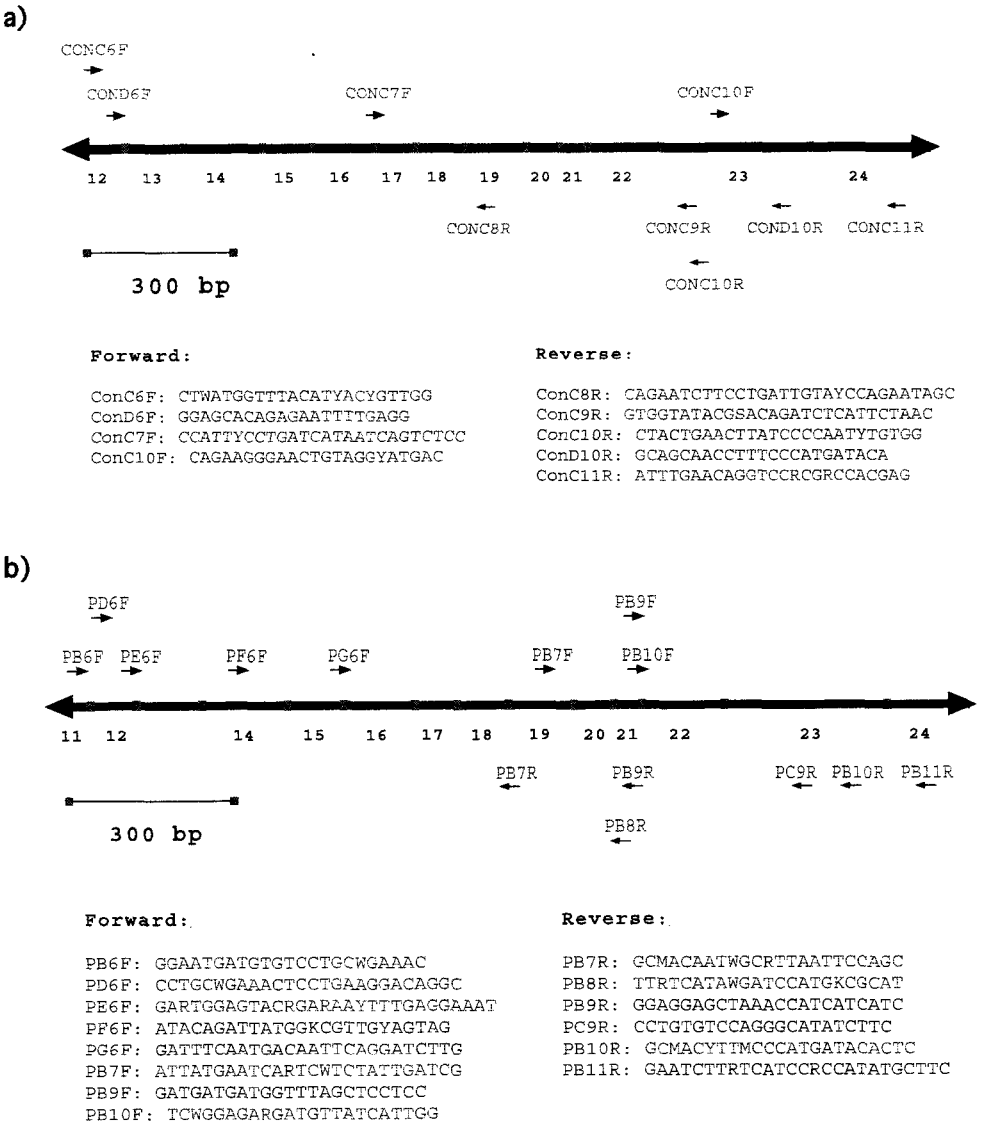


Fig. 2.1

(a) Gnepine hypothesis

(b) Gnetales basal in gymnosperms

(c) Anthophyte hypothesis

(d) Gnetales basal in seed plants

(e) Gnetifer hypothesis

Fig. 2.2

Fig. 2.3

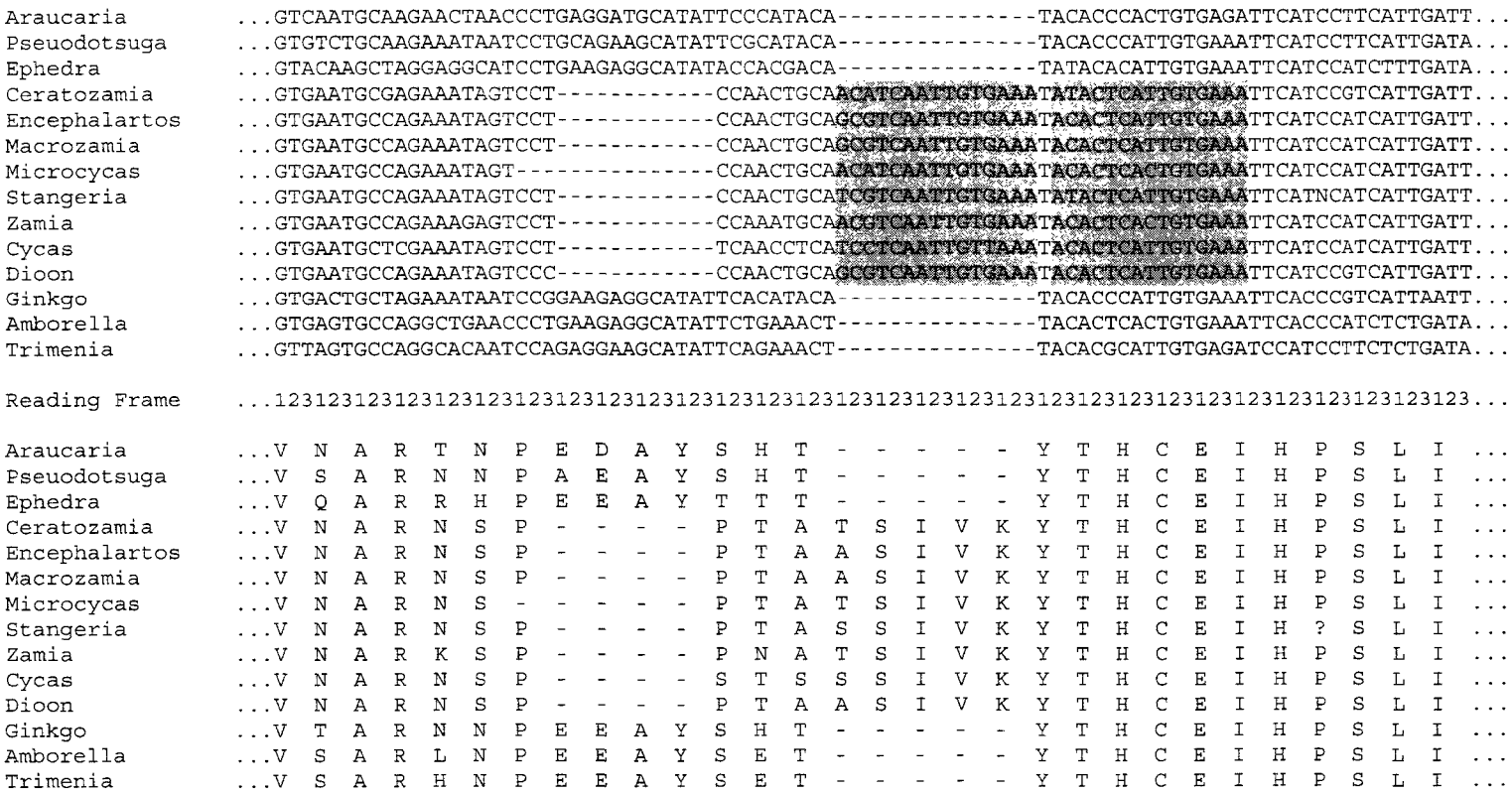


Fig. 2.4

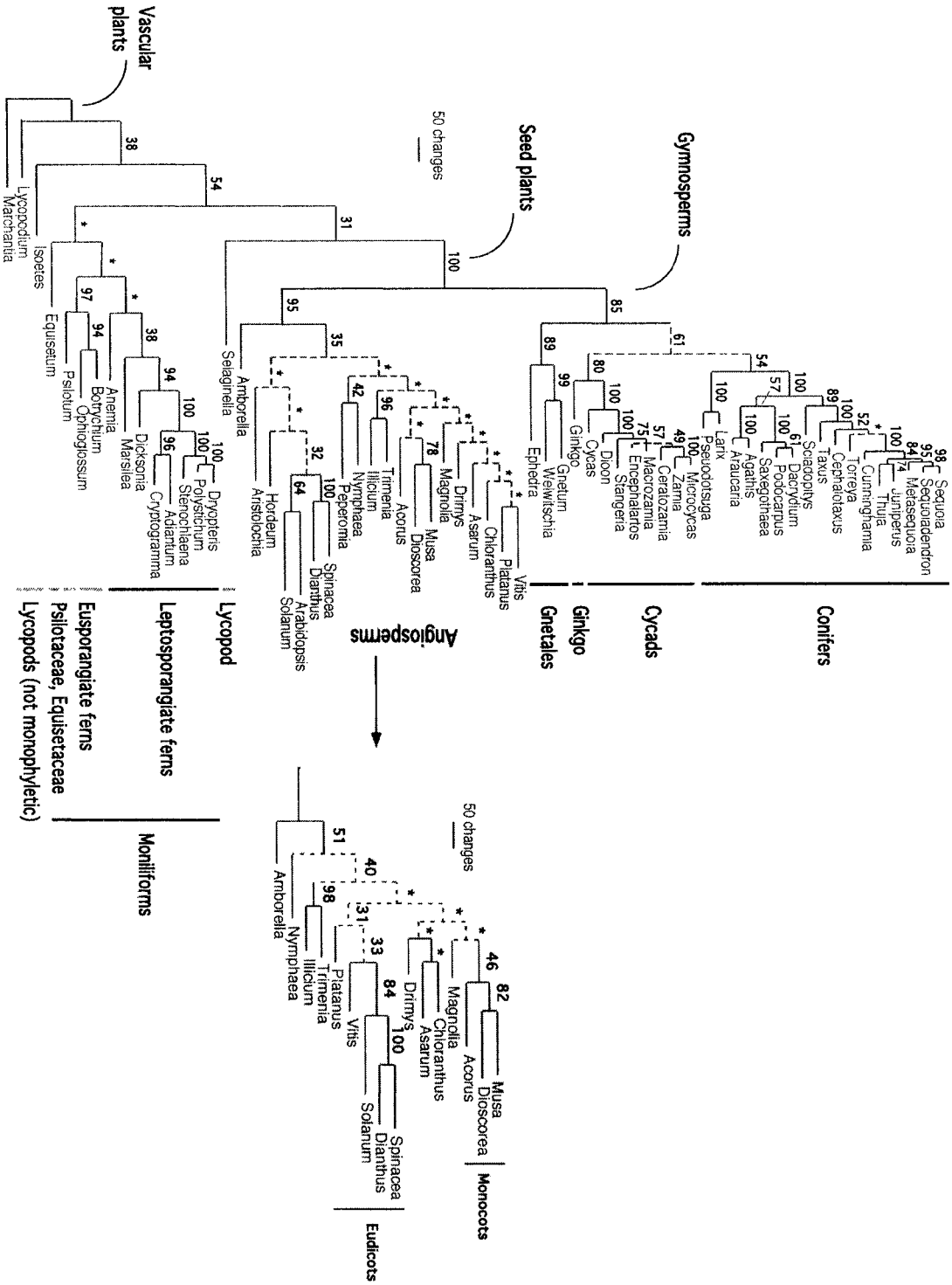


Fig. 2.5

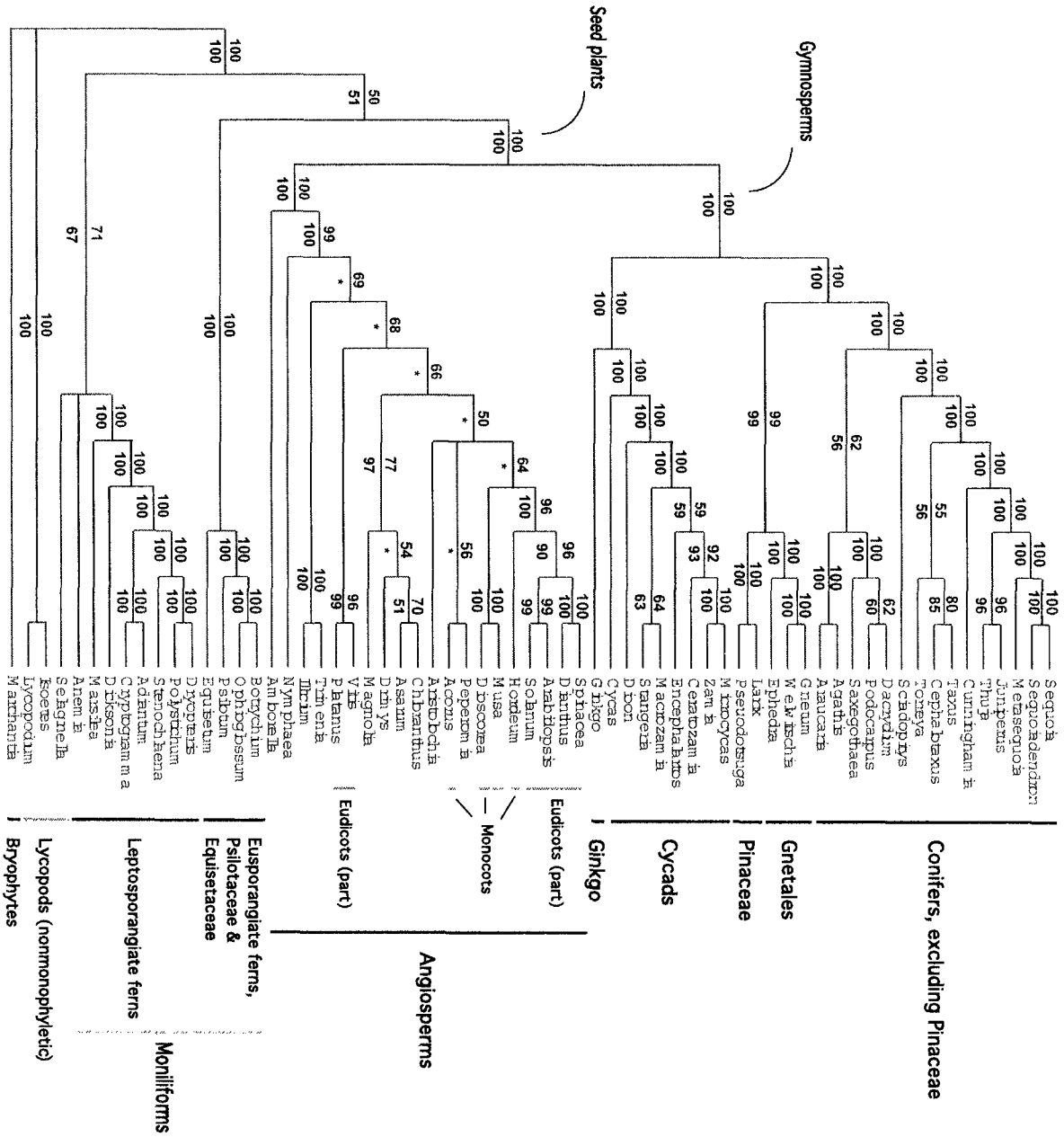
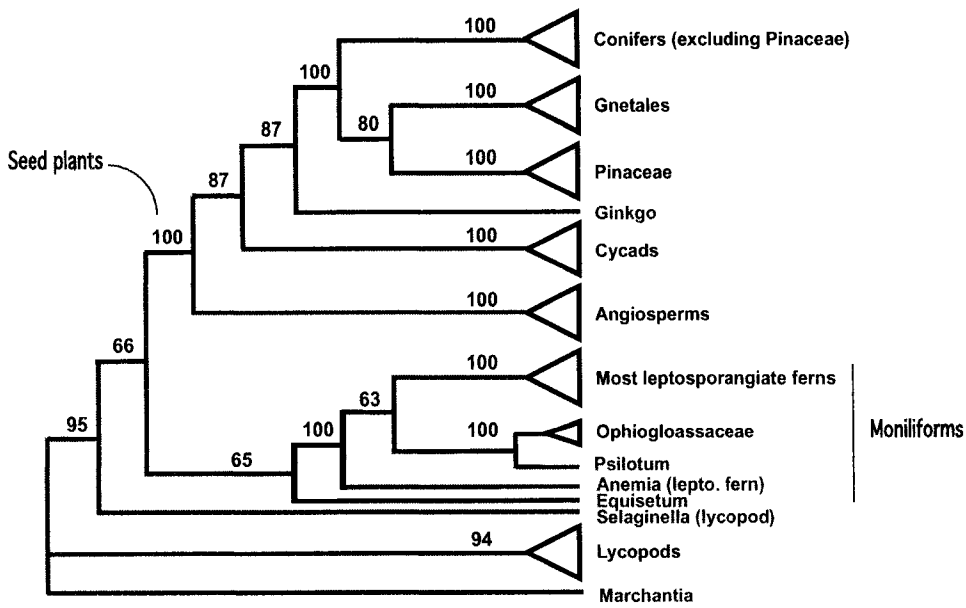


Fig. 2.6

a) Codon positions 1 + 2



b) Codon position 3

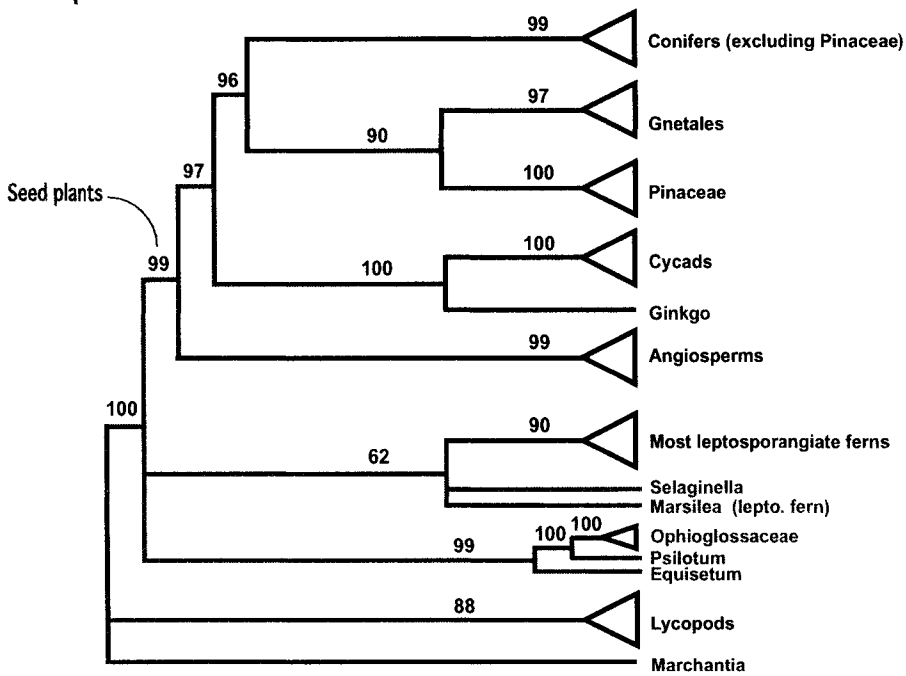


Fig.2.7

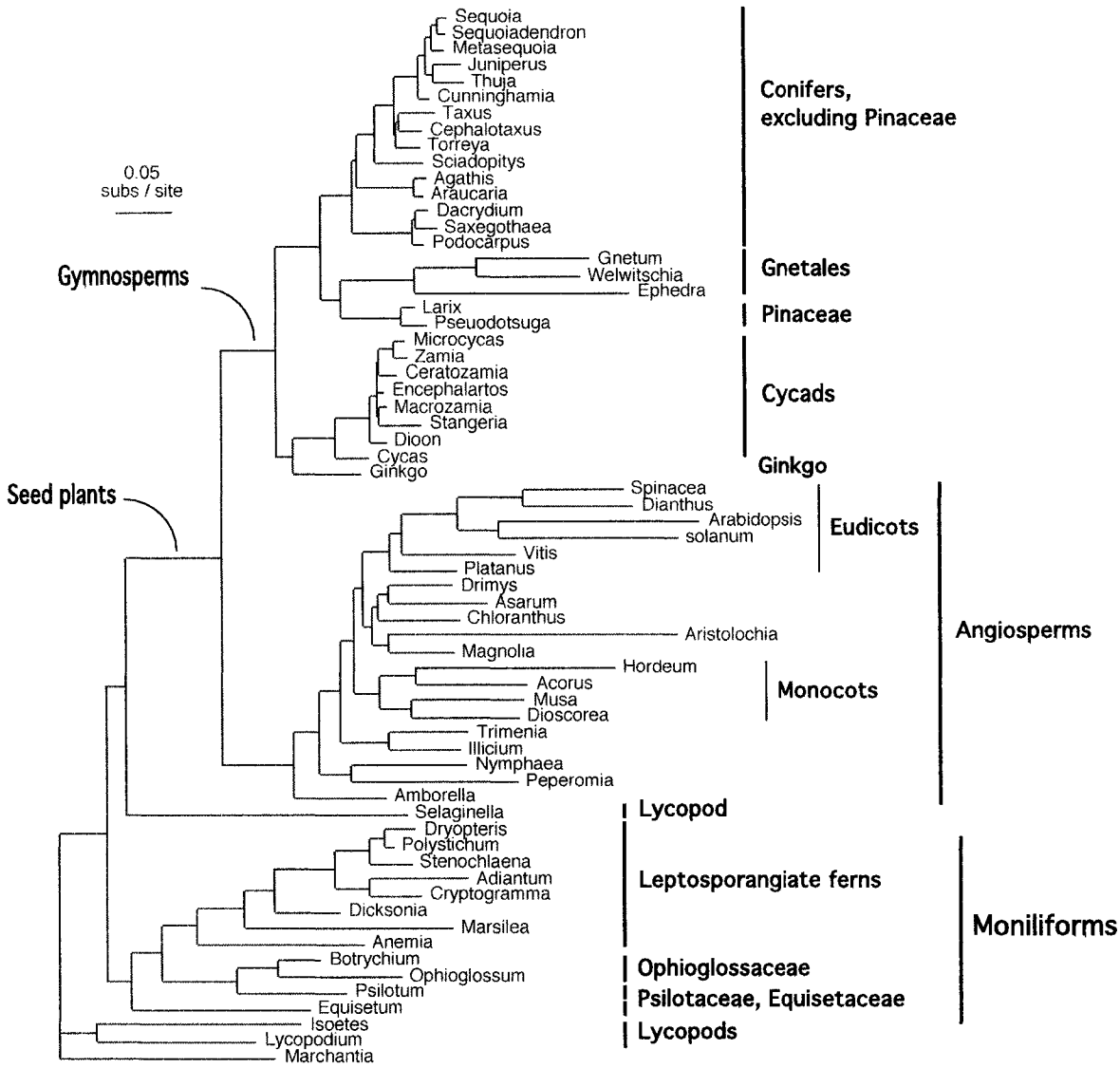


Fig.28

~Chapter Three~

Reconstruction of Cycad Deep Phylogenetic Relationships Using a Large Plastid Data Set and Two Nuclear Loci

Introduction

There are ~185–300 extant cycad species (Jones,1993; Hill et al., 2003) making them the largest extant group of gymnosperms after the conifers (~650 species). They are all long-lived, dioecious, woody species. Although they make up a small proportion of the earth's terrestrial flora, they play a central role in our understanding of seed-plant evolution (e.g., Doyle, 1998), and have a substantial and ancient fossil record that dates back to the early Permian (Mamay, 1976; Gao and Thomas, 1989). During the Mesozoic, their diversity and overall abundance was as its peak, and they were widespread across both hemispheres (Stewart and Rothwell, 1993). They currently have a much narrower distribution in the Old and New World, with centres of diversity in Mexico, South Africa, and northeast Australia (Jones, 1993). They have also attracted considerable interest from horticulturists, and are popular exotic additions to greenhouses and botanical gardens.

Cycad classification has changed several times in the past half-century. Johnson (1959) recognized three families and ten genera: Cycadaceae (one genus, *Cycas*), Stangeriaceae (one genus, *Stangeria*) and Zamiaceae (eight genera: *Bowenia, Ceratozamia, Dioon, Encephalartos, Lepidozamia, Macrozamia,*

Microcycas and *Zamia*). Stevenson (1990) added *Chigua*, and de Laubenfels and Adema (1998) described a new genus, *Epicycas*, for seven Asian species in *Cycas*. However, Hill et al. (2003) found *Chigua* and *Epicycas* to be nested within *Zamia* and *Cycas*, respectively, and proposed that these new genera need not be recognized. Stevenson (1992) moved *Bowenia* to Stangeriaceae. A more recent classification by de Laubenfels (1999) recognizes only *Dioon, Encephalartos, Lepidozamia,* and *Macrozamia* in Zamiaceae, and houses *Bowenia, Ceratozamia, Chigua, Microcycas, Stangeria* and *Zamia* in a new family, Ultracycadaceae.

Recent molecular studies (Rai et al., 2003; Hill et al., 2003; Bogler and Francisco-Ortega, in press) suggest all of the above cycad families are nonmonophyletic as currently construed, except Cycadaceae in its traditional sense. Based on the results of a molecular phylogenetic study, Hill et al. (2003) suggest that only two cycad families should be recognized: Cycadaceae (containing only *Cycas*) and Zamiaceae (containing all other genera). This classification is congruent with molecular (Treutlein and Wink, 2002; Hill et al., 2003; Rai et al., 2003; Bogler and Francisco-Ortega, in press) and morphological evidence (e.g., Stevenson, 1990), which supports the hypothesis that *Cycas* is the sister-group of all remaining taxa. *Cycas* has long been regarded as distinctive from other cycads, based on a variety of characters (Hill et al. 2003), the most prominent of which are its leaf-like, indeterminate megasporophylls. *Dioon* is likely the sister group of the remaining cycads, excluding *Cycas* (Rai et al., 2003; Bogler and Francisco-Ortega, in press). However, a morphology-based study by Stevenson (1990) suggested that *Stangeria* and *Bowenia* are a clade that is the sister group of all cycads except *Cycas*. Molecular studies indicate that these two genera are not closely related (Hill et al., 2003; Rai et al., 2003; Bogler and Francisco-Ortega, in press). A close relationship between *Ceratozamia*, *Zamia*

and *Microcycas* is supported in most studies, as is a close relationship between *Encephalartos*, *Lepidozamia*, and *Macrozamia*, but generally without strong bootstrap support for either clade in studies that performed formal phylogenetic analysis (see Hill et al., 2003). Early treatments classified the latter three taxa under *Macrozamia* (Miquel, 1861)

Crane (1985) found cycads and medullosan pteridosperms (one group of seed ferns, all now extinct) to be sister groups, using morphological data. Doyle and Donoghue (1992), however, found that cycads were distantly related to medullosans. Their most-parsimonious trees found them to be either the sister group of *Peltaspermum*, or the sister group of a large clade composed of all other seed plants excluding lyginopterids and medullosans. In another morphological analysis, Rothwell and Serbet (1994) also found that cycads were distantly related to medullosans, but their most- parsimonious trees placed cycads in several different places, and so no definite conclusions regarding their position among extant and extinct taxa could be made.

The placement of cycads among extant seed-plant groups, [i.e., the other gymnosperms (conifers, *Ginkgo biloba*, Gnetales) and angiosperms] is uncertain. Rai et al. (2003) found *Ginkgo* to be the sister group of cycads, among living taxa. They noted that *Ginkgo* and cycads share a low rate of molecular evolution, and an elevated transition-transversion ratio compared to other seed plants, and suggested that these characteristics may be additional synapomorphies for this relationship. This relationship is also supported by other molecular studies (Boivin et al., 1996; Goremykin et al., 1996; Chaw et al., 1997, 2000). However, several morphological cladistic analyses instead suggest that *Ginkgo* and conifers, not *Ginkgo* and cycads, are sister taxa (Parenti, 1980; Crane, 1985; Doyle and Donoghue, 1986), and that cycads may be the sister group of all other extant seed plants (Crane, 1985; Loconte and Stevenson, 1990); the latter result is found in

some molecular studies (Bowe et al., 2000; Chaw et al., 2000; Magallon and Sanderson, 2002). The molecular analyses of Rydin et al. (2002) and Rai et al. (2003) instead found Gnetales (*Gnetum, Ephedra* and *Welwitschia*) to be the sister group of all remaining seed plants, and in both cases this result was well-supported by very large data sets. However, both authors noted that this result could reflect misinference of relationship, due to problematic long branches in the seed plants and elsewhere (see Felsenstein, 1978; Hendy and Penny, 1989). The same caveat could be applied to all current studies of seed-plant relationships, regardless of data source or taxa considered (see also Chapter 2).

In this study, I focus primarily on resolving relationships within the cycads, but I also briefly address the position of the cycads in seed-plant phylogeny. I build on the large data set collected by Rai et al. (2003), who surveyed 17 photosynthetic, chlororespiratory and ribosomal plastid genes and associated noncoding regions, for exemplar taxa representing seven cycad genera, 17 other seed plants, a pteridophyte (*Psilotum*) and a bryophyte (*Marchantia*). Here I add comparable plastid data for the three remaining widely accepted cycad genera (*Microcycas, Macrozamia* and *Lepidozamia*), and include several new outgroup taxa. I also add previously published data (Hill et al., 2003) from another plastid region (the *trn*L[UAA]–*trn*F[GAA] region), and consider two nuclear genes, one representing previously published data (the 26S rDNA locus [Stefanovic et al., 1998; Hill et al., 2003]), the other new data from the *RPB2* locus (presented in Chapter 2 of this thesis).

Materials and Methods

Genomic and taxonomic sampling.

The taxa and genomic regions surveyed here are an expansion of those examined in Rai et al. (2003), who sequenced 17 protein-coding plastid genes (*atp*B, *ndh*B, *ndh*F, *psb*B, *psb*C, *psb*D, *psb*E, *psb*F, *psb*H, *psb*J, *psb*L, *psb*N, *psb*T, *rbc*L, *rpl*2, *rps*7, 3'-*rps*12), six short intergenic regions spanning genes in two photosystem II gene clusters (*psb*E-*psb*F-*psb*L-*psb*J and *psb*B-*psb*T-*psb*N-*psb*H), three spacer regions between two ribosomal small subunit genes (3'-*rps*12, *rps*7) and *ndh*B, and several introns (one each in *ndh*B, *rpl*2 and 3'-*rps*12). The additional taxa included here (Table 3.1) are exemplars representing three cycad genera not examined in Rai et al. (2003) (i.e., *Lepidozamia, Macrozamia,* and *Microcycas*), two additional bryophyte genera [a hornwort, *Anthoceros* (Kugita et al., 2003) and a moss, *Physcomitrella* (Sugiura at al., 2003)], an additional pteridophyte genus [a leptosporangiate fern, *Adiantum* (Wolf et al., 2003) and two additional angiosperms [(a eudicot, *Arabidopsis* and a member of Austrobaileyales, *Trimenia* (SW Graham, HS Rai, PA Reeves, ACE Burns & RG Olmstead, unpubl. data)]. For the cycads and *Ginkgo* I also added the contiguous plastid genome region spanning *trn*L(UAA) and *trn*F(GAA), using data generated by Hill et al. (2003) (Table 3.2). This region consists primarily of the *trn*L(UAA) intron and the intergenic region between *trn*L(UAA) and *trn*F(GAA). In a few instances, closely related alternative taxa were sampled for this region (see Table 3.2). With these and a few other minor exceptions noted in Table 3.1 and 3.2, all taxa were completely represented for all comparable plastid regions.

Two nuclear genes, the 26S rDNA locus and *RPB2*, were also considered. Data for *RPB*2, the second largest subunit of nuclear RNA polymerase II were derived using methods described in Chapter 2, and include previously published sequences (Table 3.2). *RPB2* was chosen because it is generally single- to very-low copy number (one or two copies in vascular plants: Larkin and Guilfoyle, 1993; Denton et al., 1998; Oxelman and Bremer, 2000; Oxelman et al., 2004) and is slowly evolving in its protein-coding portions (Denton et al., 1998). Sequences representing the first ~0.7 kb of the 26S rDNA locus were also included. These were obtained primarily by Stefanovic et al. (1998) and Hill et al. (2003). All other sources are listed in Table 3.2. The 26S rDNA locus codes for 26S rRNA, a major structural component of the large ribosomal subunit. This gene (part of the nuclear 18S–5.8S–26S rRNA transcription unit) is multi-copy, but evolves in a highly concerted manner within species (Hillis and Dixon, 1991). It is also relatively slowly evolving (Stefanovic et al. 1998), but is generally less conservative than 18S rDNA sequences (Kuzoff et al., 1998). In some cases, the 26S rDNA or *RPB*2 sequences were obtained for a different species from the same genus examined for the plastid data (Tables 3.1, 3.2). All sequences for exemplar genera were concatenated (i.e., for the purpose of analysis, those from closely related taxa were treated as the same taxon).

DNA extraction, amplification and sequencing.

I extracted genomic DNA from fresh or silica-dried material using the basic protocol of Doyle and Doyle (1987), with a modification noted in Rai et al. (2003). DNA amplification of plastid genes was carried out with a GeneAmp PCR system 9700 thermocycler (Applied Biosystems, Foster City, CA), using the following profile: (1) Initial denaturing at 94 C for five minutes; (2) 35 cycles of

a 95 C denaturation for 30 seconds, followed by annealing at 45 C for one minute, and a 72 C extension for two minutes, and; (3) a final extension at 72 C for seven minutes. 25 pmole of primer were used in each 50 ul reaction volume. Sequencing products were generated using a "Big Dye Terminator v 3.1" cycle sequencing kit (Applied Biosystems, Foster City, CA), following manufacturer instructions. Sequencing reactions were cleaned using Sephadex G50 columns, dried on a vacuum centrifuge and run on an ABI Prism 377 automated DNA sequencer. For most regions, both forward and reverse strands were sequenced. In a few minor cases where this was not possible, at least two sequences were generated from the same strand. For each amplified region and taxon, one extra sequence was generated from a second PCR product generated on a different day, typically using a second extraction from the same source, to control for possible errors in extraction or amplification. No such errors were found. Further details of DNA amplification and sequencing are described in Graham and Olmstead (2000a). I generated cDNAs for the *RPB2* sequences using methods described in Chapter 2.

Data compilation and alignment.

Sequences were compiled and base-called using Sequencher 4.1 (Gene Codes Corporation; Ann Arbor, MI) and then exported and added to a previous alignment (Graham et al., submitted), including those sequences obtained from GenBank (Tables 3.1, 3.2). Alignments were adjusted using Se-Al version 1.0 (Rambaut, 1998), following criteria laid out in Graham et al. (2000). Coordinates for gene, intron and exon boundaries and codon positions were determined using *Nicotiana tabacum* sequences, following Graham and Olmstead (2000a). Major details of exon, intron and spacer lengths and boundaries for most regions

included here can be found in Graham and Olmstead (2000a). For details regarding the remaining regions, see Chapter 2 of this thesis (*RPB2*) and Hill et al. (2003) (for most of the 26S rDNA and *trn*L-F sequences; Table 3.2). Noncoding regions were not included in the alignments for three bryophytes (*Anthoceros*, *Marchantia* and *Physcomitrella*) or the two pteridophytes (*Adiantum* and *Psilotum*) considered here.

Phylogenetic analysis.

Maximum parsimony analyses.— I conducted heuristic maximum-parsimony (MP) searches using PAUP* version 4.0b10 (Swofford, 2002). All characters and character-state changes were equally weighted. I performed searches using tree-bisection-reconnection (TBR) branch swapping, with 100 random addition replicates and no tree number limits. Tree searches were conducted using: (1) The entire data set; (2) Plastid data only; (3) Plastid first- and second-codon positions only; (4) Plastid third codon positions only; (5) Plastid transversions only; and (6) Plastid transitions only. To perform the penultimate analysis, the entire data set was converted to purines (R) and pyrimidines (Y). To emulate a "transitions-only" parsimony analysis, I weighted very heavily against transversions (800:1; cost of transversions vs. transitions), using a step-matrix approach in PAUP.* I performed a maximum parsimony bootstrap analysis (Felsenstein, 1985) on the entire data set, and for each data partition or character treatment described above. These analyses used the search conditions described above, but had one random addition replicate for each of the 100 bootstrap replications.

Bayesian Analyses.— I performed Bayesian phylogenetic inference using the program MrBayes version 3.0 (Ronquist and Huelsenbeck, 2003). Each analysis was run using four chains for 1.5 million generations, using the default temperature (0.2). I included only cycads and *Ginkgo* to allow analyses to be completed in a reasonable time frame. I used the general time reversible (GTR; Lanave et al., 1984; Tavaré, 1986; Barry and Hartigan, 1987; Rodrîguez et al., 1990) model of DNA sequence evolution, with a gamma distribution of rates (Yang, 1994; four rate categories) and with the proportion of invariable sites considered separately. This model (GTR + Γ + I) was chosen based on the results of a likelihood-ratio test (Huelsenbeck and Crandall, 1997; see Table 3.2). The test was performed on an MP tree obtained for cycads and *Ginkgo* only. I discarded the first 200 trees (representing 20,000 generations), because the chains were stable after this point. Clade posterior probabilities were estimated as the number of times a particular relationship was observed among the sampled trees, as determined on a majority-rule consensus tree. This basic analysis was repeated for plastid data only, and for each nuclear gene by itself.

I also performed separate Bayesian analyses in which DNA substitution model parameters were assigned separately (unlinked ; I use the term in a non-genetic sense) for various codon-based and multi-region data partitions. The first analysis considered plastid data only, with unlinked GTR + Γ + I substitution model parameter estimates for: (1) First and second codon positions; (2) Third codon positions, and; (3) Noncoding data. A second analysis considered all plastid and nuclear regions, but unlinked the substitution model parameters for six major partitions: (1) All photosystem II (*psb*) genes; (2) All genes found in the plastid genome inverted repeat (IR) region (*rpl* 2, 3 *rps*12, *rps*7 and *ndhB*); (3) Three of the most widely used loci in plant phylogenetic analysis (*atp*B, *ndh*B and *rbc*L); (4) The plastid noncoding regions, including those present in the IR region;

(5) The nuclear gene *RPB2*, and; (6) The nuclear 26S rDNA locus. The plastid loci are all part of the same genetic linkage group, but the partitions considered here may be subject to different molecular evolutionary dynamics. For example, the *psb* gene products are all part of the photosystem II complex and have some of the lowest synonymous substitution rates of single-copy plastid genes (Olmstead and Palmer, 1994); the IR genes evolve substantially more slowly than the rest of the plastid genome (e.g., Wolfe et al. 1987; Graham et al., 2000). A third analysis repeated the second one, but considered only plastid data. Separate Bayesian analyses were also performed for each nuclear locus.

Results

The final aligned data set was 31,463 base pairs long. For comparison, the unaligned data set was 17,169 bp for *Cycas revoluta.* 6,140 nucleotides were variable and parsimony informative (5,178 from plastid data, 638 from *RPB2* and 324 from 26S rDNA). For cycads there were 536 parsimony-informative sites (374 from plastid data, 63 from *RPB2* and 99 from 26S rDNA). The large size of this matrix (almost double the length of any individual taxon) is partly the result of large indels or unalignable regions found in some taxa, including some in monocot taxa that were included in the overall alignment, but not considered in this study.

Seed-plant relationships.

Maximum parsimony inference of seed-plant relationships. — A single tree was inferred when all data were combined (tree length = 25,942; Fig. 3.1).

Two trees were found when only plastid data was used (tree length = 21,412; Fig. 3.1).

Cycads are moderately well supported (88% BV) as the sister group of *Ginkgo* when all data (plastid and nuclear) are analyzed together. This clade (cycads + *Ginkgo*) is the sister group of conifers, and Gnetales and angiosperms are the respective successive sister taxa of the remaining seed plants. Each of these relationships is well supported by bootstrap analysis (100% BV), and seed-plant monophyly is also well supported. Relationships among the major seed-plant groups found using only chloroplast data are identical to those found using all the data (Fig. 3.1). However, trees inferred using different subsets of the data (the first two vs. the third codon position) or by considering different classes of character-state change (transitions vs. transversions) have some substantial conflicts with regards to the major details of seed-plant relationship, and some of these conflicts are also well supported. For example, the relationship between cycads and *Ginkgo* is moderately to strongly supported by each of the two codon position partitions considered here (Fig. 3.2a, b), but consideration of transitions only, or transversions only, leads to inference of a weakly supported relationship between cycads and conifers (Fig. 3.2c, d). The conifers are moderately well supported as monophyletic in analyses of all data combined and of codon position 3 (Figs. 3.1, 3.2b), but Gnetales are nested in conifers as the sister group of Pinaceae in analyses of the first two codon positions or of only transitions or transversions (Fig. 3.2a, c, d) with moderate to strong support. Gnetales are instead placed as the sister group of all other seed plants in the analysis of the third codon position data (Fig. 3.2b), and angiosperms are the sister group of all other angiosperms in all other maximum parsimony analyses considered here (Figs. 3.1, 3.2a, b, c) – all relationships that find moderate to strong support from their respective data partitions of classes of character-state change.

Higher-order relationships in the cycads.

Relationships inferred among the different cycad exemplar species are highly comparable among almost all different maximum parsimony and Bayesian analyses involving plastid data (Figs. 3.1–3.3). *Cycas* and *Dioon* are, respectively, the successive sister groups of all other cycads. *Bowenia* is the sister group of a clade consisting of *Encephalartos*, *Lepidozamia* and *Macrozamia*, and within the latter clade *Encephalartos* and *Lepidozamia* are sister taxa. *Stangeria* is the sister group of a clade consisting of *Ceratozamia*, *Microcycas* and *Zamia*, and within the latter clade *Microcycas* and *Zamia* are sister taxa. Most of these branches find moderate to strong support in maximum parsimony analysis of all data combined (Fig. 3.1), although the relative arrangement of *Stangeria* and *Ceratozamia* has only weak support in the parsimony analyses (Figs. 3.1, 3.2). Maximum parsimony analysis of different data subsets are generally highly congruent with these findings, but where they are not (the placement of *Dioon* and *Stangeria* in analysis of the first two codon positions; the placement of *Stangeria* in analysis of transitions or transversion only), any conflicting relationships are not strongly supported. The position of *Bowenia* as the sister group of *Encephalartos-Lepidozamia-Macrozamia* is seen in all parsimony analyses, but has weak bootstrap support in the reduced analyses (Fig. 3.2), and the analysis that considers only plastid data (Fig. 3.1). However, this relationship is moderately well supported when all data are considered together. The average bootstrap support for higher-order relationships within the cycads (including support for the cycads as a whole) is improved by data combination. The average bootstrap support value among cycads when all data are used is 88%. When only plastid data are used it is 83%, and when codon positions 1 and 2 only or codon

position 3 only are used, the average values are 76% and 75%, respectively. The combined plastid and nuclear data provide 536 parsimony-informative characters, the plastid data alone provides 374 characters, plastid codon positions 1 and 2 provide only 52 characters, and plastid codon position 3 provides 159 characters.

The Bayesian analyses of all plastid and nuclear data combined, or of all plastid data combined, find exactly the same higher-order relationships in the cycads as the main parsimony analysis (Figs. 3.1, 3.3). Estimated posterior probabilities for individual branches in the cycads are nearly uniformly high, regardless of whether plastid data are considered alone or in combination with nuclear data, or whether parameter values in the DNA substitution model are applied uniformly across all data or permitted to vary by major data partition or plastid codon position.

The only exception to this uniform pattern of strong support and congruence concerns the analysis where DNA substitution model parameters for the plastid data are allowed to vary across the two codon position partitions considered here (i.e., the first two vs. third position) and the noncoding data. In this case, support for two branches drops: the branch supporting the clade consisting of *Bowenia*, *Encephalartos*, *Lepidozamia* and *Macrozamia* (from ~100% to 85% PP), and the branch supporting the clade consisting of *Stangeria*, *Ceratozamia*, *Microcycas* and *Zamia* (from 100% to 36% PP). The latter result indicates that there may be some moderate conflict within the plastid data concerning the precise position of *Stangeria*, and in fact this analysis placed *Stangeria* as the sister group of *Zamia* and *Microcycas* with weak support (53% PP).

Each nuclear data set has relatively few characters (99 and 63 informative characters for the 26S rDNA locus and the *RPB2* locus within cycads, respectively, the latter with two fewer taxa included), and most relationships are

either congruent with the analyses of plastid data or all data together, or only weakly supported if incongruent. For both genes, *Cycas* is well supported as the sister group of all other cycads (Fig. 3.4). *Dioon* is well supported as the next successive sister group in the Bayesian analysis of the *RPB2* data, although its precise position as a lineage emerging from basal cycad phylogeny is less certain for the 26S rDNA analysis, with less than 50% PP for any particular relationship. *Zamia* and *Microcycas* are well supported as sister taxa by both nuclear genes, but a sister group relationship of these two to *Ceratozamia* is strongly supported only by the *RPB2* data (92% PP). The relative positions of *Encephalartos, Macrozamia* and *Stangeria* are poorly supported in the Bayesian analysis of the *RPB2* data (*Bowenia* and *Lepidozamia* were not sampled for this gene), as are the relative positions of *Stangeria, Bowenia, Ceratozamia* and *Dioon* in the Bayesian analysis of the 26S rDNA locus. The latter data set supports a close relationship between *Encephalartos, Lepidozamia* and *Macrozamia*, but in contrast with all other analyses, it has *Lepidozamia* and *Macrozamia* strongly supported as sister taxa, a local arrangement that represents the only strongly supported conflict among cycads in the current study.

Discussion

The parsimony analysis of all data combined, the plastid data alone (Fig. 3.1) and the two analyses of different plastid codon partitions (Fig. 3.2a, b) place *Ginkgo* as the sister group of cycads. This result is congruent with other molecular analyses (Boivin et al.,1996; Goremykin et al., 1996; Chaw et al., 1997, 2000; and Rai et al., 2003). *Ginkgo* places as the sister group of a clade composed of conifers and Gnetales when either only transitions or transversions

are considered, but this relationship has only weak bootstrap support in these analyses. In contrast, there are some robustly supported conflicts in parsimony analysis of the various plastid data with regards to the position of Gnetales in seed-plant phylogeny (Figs. 3.1, 3.2). Similar conflicts have been observed elsewhere (e.g., Sanderson et al., 2000; Magallon and Sanderson, 2002; Rydin et al. 2002). If the placement of Gnetales as the sister group of Pinaceae observed with some analyses (Fig. 3.2) is an artifact of long-branch attraction (see Chapter 2), then we might expect it to be observed more readily with more rapidly evolving classes of data (Felsenstein, 1983). However, this relationship is observed in analyses here that focus on both transitions and transversions, and the former class of character-state change is known to occur substantially more frequently than the former, both in general and for plastid data in particular (e.g., Rai et al., 2003). The "gnepine" relationship is also seen in an analysis that considers the first two codon positions of plastid protein-coding genes, but not the third, yet the latter partition evolves on the whole substantially more rapidly than the former two codon positions, because for comparable plastid data, nucleotide substitutions in the first two codon positions are predominantly nonsynonymous, whereas those in the third are predominantly synonymous (e.g., Sanderson et al., 2000). The more rapidly evolving subset of nucleotides in the plastid protein-coding regions support conifer monophyly, yet place the branch leading to Gnetales deep in seed-plant phylogeny, a position seen when all data are analyzed together (Fig. 3.1; Rai et al. 2003).

There is thus substantial conflict in the plastid genome regarding the relationships among the five extant groups of seed plants, and no clear indication that any particular subset of the data can be trusted with greater confidence, regardless of overall rates of molecular evolution. Because of the mutual strong conflict, at least some of the relationships among the five major seed-plant groups

must be wrong, and perhaps all of them are. The problem of overall seed-plant relationships deserves more thorough study (using the sorts of simulation analyses I performed in Chapter 2, for example), and the addition of more taxa (denser outgroup sampling may be beneficial, for example). It is also possible that model-based approaches (e.g., maximum likelihood and Bayesian inference) will be less prone to the conflicts observed with maximum parsimony analysis, although preliminary results from maximum likelihood analysis suggest otherwise (SW Graham, HS Rai et al., unpublished data). This thorny issue will be addressed in more detail elsewhere.

In contrast, higher-level relationships within cycads show few signs of substantial conflict between different classes of data. The major exception is that the plastid data indicate that *Encephalartos* and *Lepidozamia* are sister taxa, whereas the nuclear 26S rDNA data instead indicate that *Macrozamia* and *Lepidozamia* are sister groups. In both instances, the results were robustly supported by Bayesian analysis. Incongruence such as this may result from an analytical problem (such as systematic error, see Chapter 2), or it may instead reflect some underlying biological phenomenon: "real" or "hard" incongruence, rather than conflicts that arise because of sampling error due to too few data ("spurious" or soft incongruence).

Apparently well-supported conflicts such as these may be the result of a variety of evolutionary processes (Maddison, 1997; Wendel and Doyle, 1998) such as introgression and lineage sorting of ancestral polymorphisms. Occasional hybridization is known in extant cycads, but rarely between species from different genera (Johnson and Wilson, 1990). This does not rule out that gene transfer might have occurred among the ancestors of *Encephalartos*, *Lepidozamia* or *Macrozamia,* as appears to have occurred in *Gossypium gossypioides*, for example (Wendel et al., 1995). Incomplete lineage sorting of ancestral polymorphisms can

also result in incongruent gene/ species trees (e.g., Wendel and Doyle, 1998). Since only one nuclear gene was used in this study for all cycad genera (attempts to sample *RPB2* for *Lepidozamia* and *Bowenia* failed), I did not have sufficient data to determine whether *RPB2* was also incongruent with the plastid data in this part of the tree.

All analyses performed here strongly support a sister-group relationship between *Cycas* and all remaining cycad genera. This is congruent with other recent molecular (Caputo et al., 1991; Treutlein and Wink, 2002; Hill et al., 2003; Rai et al., 2003; Bogler and Francisco-Ortega, in press) and morphological (Petriella and Crisci, 1977; Crane, 1988; Stevenson, 1990) studies. Bogler and Francisco-Ortega (in press) note that sequences of the plastid *trn*L intron and the nuclear internal transcribed spacer between 5.8S and 26S rDNA for several species of *Cycas* were more divergent than those within any other cycad genus. They suggest that this indicates that *Cycas* has been isolated from the other genera for an extended period of time. *Cycas* is also morphologically very distinct from other cycad genera. It has indeterminate female cones, platyspermic seeds, diffuse secondary leaf vasculature, ascending ovules, and multi-ovulate megasporophylls (Stevenson, 1990). In contrast, all other cycad genera have indeterminate female cones, radiospermic seeds, regular secondary leaf vasculature, inverted ovules and biovulate megasporophylls (Stevenson, 1990).

The combined nuclear and chloroplast data set used by Hill et al. (2003) indicates that *Stangeria* is the sister group of all remaining cycad genera, excluding *Cycas*, a weakly supported result in their study. A morphological analysis instead indicated that a clade composed of *Stangeria* and *Bowenia* occupies this position (Stevenson, 1990). However, the plastid data and the *RPB2* data clearly show that *Dioon* is the sister group of the remaining cycads, excluding *Cycas* (the 26S rDNA data do not support any one arrangement). This

finding is well-supported by both parsimony and Bayesian analyses (Figs. 3.1, 3.3), and is congruent with other molecular studies (Rai et al. 2003; Bogler and Francisco-Ortega, in press) and a morphological study (Petriella and Crisci, 1977).

There was no hint of a close relationship between *Bowenia* and *Stangeria* from the plastid or nuclear data examined here. Other molecular analyses (Hill et al., 2003; Rai et al., 2003; Bogler and Francisco-Ortega, in press) have also failed to find a close relationship between these two taxa, and so the family Stangeriaceae as circumscribed by Stevenson (1990) is not monophyletic. Apparent synapomorphies linking these two genera (the presence of concentric vascular bundles in the cotyledon with endarch protoxylem, petioles with a circular arrangement of vascular bundles, and vascularized stipules), therefore probably represent characters that evolved independently in these two taxa or that evolved once and were subsequently lost in the other taxa found to be close relatives of these two genera (Figs. 3.1–3.3).

My data indicate two large cycad clades, one composed of *Ceratozamia, Microcycas, Stangeria* and *Zamia,* and the other of *Bowenia, Encephalartos, Lepidozamia* and *Macrozamia.* In the first clade, there is apparently minor incongruence between codon position partitions with regards to the placement of *Ceratozamia* and *Stangeria.* The parsimony analysis that considered all data combined depicts *Ceratozamia* as the sister group of a clade composed of *Microcycas* and *Zamia*, strongly supported in most Bayesian analyses. However, the parsimony analysis of codon positions one and two alone, and the Bayesian analysis with parameter estimates for the two codon position partitions unlinked, instead places *Stangeria* as the sister group of the *Microcycas-Zamia* clade. However, in the latter cases there is only very weak support for the discordant placement of *Stangeria.* Codon positions one and two provide very few

parsimony-informative characters when compared to third codon positions, so the discordant parsimony result, at least, may be due to insufficient data in the first two codon positions, and it should be noted that *Stangeria* and *Ceratozamia* are separated by very short branches (Fig. 3.1). However, although the first two codon positions provide only one-third the number of parsimony-informative characters compared to the third codon positions, the average bootstrap support value when relationships among cycads are considered is very similar between these two data partitions (76% and 75% for codon positions 1 and 2 vs. 3, respectively).

There are a number of morphological synapomorphies that support the grouping of *Ceratozamia, Microcycas* and *Zamia*. These include: articulated leaflets, absence of terminal leaflets in seedlings, foveolate pollen, peduncles without cataphylls, peltate megasporophylls and the presence of stipules (Hill et al., 2003). A relationship between these three taxa is also supported by phylogenetic analyses using morphological data (Stevenson, 1990) and by other molecular studies (Hill et al., 2003; Bogler and Francisco-Ortega, in press). These three genera are also all found in the neotropics. *Ceratozamia* occurs in Central America from Mexico to Belize, *Microcycas* is endemic to Cuba, and *Zamia* occurs in the West Indies, the southeastern United States, Central America and northern South America.

There are no clear synapomorphies supporting the clade composed of *Bowenia, Encephalartos, Lepidozamia,* and *Macrozamia*, although Hill et al. (2003) also found this clade in their analyses. The morphological analyses of Stevenson (1990) and Petriella and Crisci (1977) recovered a clade composed of the latter three genera. They all possess megasporophylls with lateral lobes, and have medullary vascular bundles (Stevenson, 1990), potential synapomorphies for this clade.

Major conclusions.

Relationships among cycads are resolved here with moderate to strong support. *Cycas* and *Dioon* are, respectively, successive sister taxa to the remaining genera. *Ceratozamia, Microcycas, Stangeria* and *Zamia* form a clade that is the sister group to one composed of *Bowenia, Encephalartos, Lepidozamia* and *Macrozamia.* The results presented here are largely congruent with other recent studies, provide substantial support for the backbone of cycad relationships (more than any previous study), and support the suggestion by Hill et al. (2003) that cycads should be circumscribed to contain only two families: Cycadaceae, with the relatively isolated genus *Cycas* and Zamiaceae, containing all remaining genera.

References

Barry, D., and Hartigan, J. A. 1987. Asynchronous distance between homologous DNA sequences. *Biometrics* **43**: 261–276.

Bogler, D. J., and Francisco-Ortega, J. in press. Molecular systematic studies in cycads: Evidence from *trnL* intron and ITS2 sequences. *Memoirs of the New York Botanical Garden.*

Boivin, R., Richard, M., Beauseigle, D., Bousquet, J., and Bellemare, G. 1996. Phylogenetic inferences from chloroplast *chl*B gene sequences of *Nephrolepis exaltata* (Filicopsida), *Ephedra altissima* (Gnetopsida), and diverse land plants. *Molecular Phylogenetics and Evolution* **6**: 19–29.

Borsch, T., Hilu, K. W., Quandt, D., Wilde, V., Neinhuis, C. and Barthlott, W. 2003. Noncoding plastid *trn*T-*trn*F sequences reveal a well resolved phylogeny of basal angiosperms. *Journal of Evolutionary Biology* **16***:* 558–576.

Bowe, L. M., Coat, G., and dePamphilis, C. W. 2000. Phylogeny of seed plant based on all three genomic compartments: Extant gymnosperms are monophyletic and Gnetales' closest relatives are conifers. *Proceedings of the National Academy of Sciences USA* **97:** 4092–4097.

Caputo, P., Stevenson, D. W., and Wurtzel, E. T. 1991. A phylogenetic analysis of American Zamiaceae (Cycadales) using chloroplast DNA restriction fragment length polymorphisms. *Brittonia* **43**: 135–145.

Chaw, S-M., Parkinson, C. L., Cheng, Y., Vincent, T. M., and Palmer, J. D. 2000. Seed plant phylogeny inferred from all three plant genomes: Monophyly of extant gymnosperms and origin of Gnetales from conifers. *Proceedings of the National Academy of Sciences USA.* **97:** 4086–4091.

Chaw, S., Zharkikh, A., Sung, H., Lau, T., and Li, W. 1997. Molecular phylogeny of extant gymnosperms and seed plant evolution: Analysis of 18S rRNA sequences. *Molecular Biology and Evolution* **14**: 56–68.

Crane, P. R. 1985. Phylogenetic relationships in seed plants. *Cladistics* **1**: 329–348.

Crane, P. R. 1988. Major clades and relationships in the "higher" gymnosperms. pp. 218–272. in: *Origin and Evolution of Gymnosperms* (C. B. Beck, ed.) Columbia University Press, New York, NY.

Denton, A. L., McConaughy, B. L., and Hall, B. D. 1998. Usefulness of RNA polymerase II coding sequences for estimation of green plant phylogeny. *Molecular Biology and Evolution* **15**: 1082–1085.

Doyle, J. A., 1998. Phylogeny of vascular plants. *Annual Review of Ecology and Systematics* **29**: 567–599.

Doyle, J. A., and Donoghue, M. J. 1986. Seed plant phylogeny and the origin of angiosperms: an experimental cladistic approach. *Botanical Reviews* **52**: 321–431.

Doyle, J. A., and Donoghue, M. J. 1992. Fossils and seed plant phylogeny reanalyzed. *Brittonia* **44:** 89–106.

Felsenstein, J. 1978. Cases in which parsimony or compatibility methods will be positively misleading. *Systematic Zoology* **27:** 401–410.

Felsenstein, J., 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. *Journal of Molecular Evolution* **17**: 368–376.

Felsenstein, J. 1983. Parsimony in systematics: Biological and statistical issues. *Annual Review of Ecology and Systematics* **14**: 313–333.

Felsenstein, J. 1985. Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* **39**: 783–791.

Gao, Z-F., and Thomas B. A. 1989. A review of fossil cycad megasporophylls, with new evidence of *Crossozamia* Pomel and it associated leave from the lower Permian of Taiyuan, China. *Review of Palaeobotany and Palynology* **60**: 205–223.

Goremykin, V., Bobrova, B., Pahnke, J., Troitsky, A., Antonov, A., and Martin, W. 1996. Noncoding sequences for the slowly evolving chloroplast inverted repeat do not support gnetalean affinities of angiosperms *Molecular Biology and Evolution* **13**: 383–396.

Graham, S. W. and Olmstead, R. G. 2000a. Utility of chloroplast genes for inferring the phylogeny of the basal angiosperms. *American Journal of Botany* **87**: 1712–1730.

Graham S. W., and Olmstead R. G. 2000b. Evolutionary significance of an unusual chloroplast DNA inversion in two basal angiosperm lineages. *Current Genetics* **37**: 183–188.

Graham, S. W., Reeves, P. A., Burns, A. C. E., and Olmstead, R. G. 2000. Microstructural changes in noncoding chloroplast DNA: interpretation, evolution, and utility of indels and inversions in basal angiosperm phylogenetic inference. *International Journal of Plant Sciences* **161**: S83–S96.

Graham, S. W., Zgurski, J. M., McPherson, M. A., Cherniawsky, D. M., Saarela, J. M., Horne, E. S. C., Smith, S. Y., Wong, W. A., O'Brien, H. E., Biron, V. L., Pires, J. C., Olmstead, R. G., Chase, M. W., and Rai, H. S. Submitted. Robust inference of monocot deep phylogeny using an expanded multigene plastid data set.

Hasegawa, M. H., Kishino, H., and Yano, T. 1985. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *Journal of Molecular Evolution* **22**: 160–174.

Hendy, M. D., and Penny, D. 1989. A framework for the quantitative study of evolutionary trees. *Systematic Zoology* **38**: 297–309.

Hill, K. D., Chase, M. W., Stevenson, D. W., Hills, H. G., and Schutzman, B. 2003. The families and genera of cycads: A molecular phylogenetic analysis of cycadophyta based on nuclear and plastid DNA sequences. *International Journal of Plant Sciences* **164**: 933–948.

Hillis, D. M., and Dixon, M. T. 1991. Ribosomal RNA: Molecular evolution and phylogenetic inference. *Quarterly Review of Biology* **66**: 411–453.

Huelsenbeck, J. P., and Crandall, K. A. 1997. Phylogeny estimation and hypothesis testing using maximum likelihood. *Annual Review of Ecology and Systematics* **28**: 437–466.

Johnson, L. A. S. 1959. The families of cycads and the Zamiaceae of Australia. *Proceedings of the Linnean Society of New South Wales* **84**: 64–117.

Johnson, L. A. S., and Wilson, K. L. 1990. General traits of the Cycadales. pp. 363-369. In: Kramer, K. U., and Green, P. S., eds. *The Families and Genera of Vascular Plants, Vol. 1: Pteridophytes and gymnosperms* (series ed. K. Kubitzki). Springer-Verlag, Berlin.

Jones, D. L. 1993. *Cycads of the World.* Smithsonian Institution Press, Washington, D. C.

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

Kim, S., Soltis, D. E., Soltis, P. S., Zanis, M. J. and Suh, Y. 2004. Phylogenetic relationships among early-diverging eudicots based on four genes: were the eudicots ancestrally woody? *Molecular Phylogenetics and Evolution* **31**: 16–30.

Kugita, M., Kaneko, A., Yamamoto, Y., Takeya, Y., Matsumoto, T., and Yoshinaga, K. 2003. The complete nucleotide sequence of the hornwort (*Anthoceros formosae*) chloroplast genome: insight into the earliest land plants. *Nucleic Acids Research* **31**: 716–721.

Kuzoff, R. K., Sweere, J. A., Soltis, D. E., Soltis, P. S., and Zimmer, E. A. 1998. The phylogenetic potential of entire 26S rDNA sequences in plants *Molecular Biology and Evolution* **15**: 251–263.

Lanave, C., Preparata, G., Saccone, C., and Serio, G. 1984. A new method for calculating evolutionary substitution rates. *Journal of Molecular Evolution* **20**: 86–93.

Larkin, R., and Guilfoyle, T. 1993. The second largest subunit of RNA polymerase II from *Arabidopsis thaliana. Nucleic Acids Research* **21**: 1038.

de Laubenfels, D. J., and Adema, F. 1998. A taxonomic revision of the genera *Cycas* and *Epicycas* gen. nov. (Cycadaceae). *Blumea* **43:** 351-400.

de Laubenfels, D. J. 1999. The families of the Cycadaceae. *Encephalartos* **59**: 7–9.

Loconte, H., and Stevenson, D. W. 1990. Cladistics of the Spermatophyta. *Brittonia* **42**: 97–211.

Maddison, W. P. 1997. Gene trees in species trees. *Systematic Biology* **46**: 523–536.

Magallon, S., and Sanderson, M. J. 2002. Relationships among seed plants inferred from highly conserved genes: Sorting conflicting phylogenetic signals among ancient lineages. *American Journal of Botany* **89**: 1991–2006.

Mamay, S. H. 1976. Paleozoic origin of the cycads. *U. S. Geological Survey Professional Paper 934* Washington, D. C.

Miquel, F. A. W. 1861. *Prodromus Systematis Cycadearum.* van der Post, Utrecht.

Neyland, R. 2002. A phylogeny inferred from large-subunit (26S) ribosomal DNA sequences suggests that Burmanniales are polyphyletic. *Australian Systematic Botany* **15:** 19–28.

Olmstead, R. G., and Palmer, J. D. 1994. Chloroplast DNA systematics: A review of methods and data analysis. *American Journal of Botany* **81**: 1205–1224

Oxelman, B., and Bremer, B. 2000. Discovery of paralogous nuclear gene sequences coding for the second-largest subunit of RNA polymerase II (*RPB2*) and their phylogenetic utility in Gentianales of the asterids. *Molecular Biology and Evolution* **17**: 1131–1145.

Oxelman B., Yoshikawa, N., McConaughy, B. L., Luo, J., Denton, A. L., and Hall, B. D. 2004. *RPB2* gene phylogeny in flowering plants, with particular emphasis on asterids. *Molecular Phylogenetics and Evolution* **32**: 462–479.

Parenti, L. R. 1980. A phylogenetic analysis of the land plants. *Biological Journal of the Linnean Society* **13:** 225–242.

Petriella, B., and Crisci, J. V. 1977. Estudios numericos en Cycadales. II. Cycadales actuales: simulacion de arboles evolutivos. *Obra Mus La Plata* **3**: 151–159.

Rai, H. S., O'Brien, H. E., Reeves, P. A., Olmstead, R. G., and Graham, S. W. 2003. Inference of higher-order relationships in the cycads from a large chloroplast data set. *Molecular Phylogenetics and Evolution* **29**: 350–359.

Rambaut, A., 1998. "*Se-Al (Sequence Alignment Editor Version 1.0)*," Computer program and documentation. Department of Zoology, University of Oxford, UK.

Rodríguez, F., Oliver, J. L., Marín, A., and Medina, J. R., 1990. The general stochastic model of nucleotide substitution. *Journal of Theoretical Biology* **142**: 485–501.

Ronquist, F., and Huelsenbeck, J. P. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**: 1572–1574.

Rothwell, G.W., and R. Serbet. 1994. Lignophyte phylogeny and the evolution of spermatophytes: A numerical cladistic analysis. *Systematic Botany* **19**: 443–482.

Rydin, C., Kallersjo, M., and Friis, E. M. 2002. Seed plant relationships and the systematic position of Gnetales based on nuclear and chloroplast DNA: conflicting data, rooting problems and the monophyly of conifers. *International Journal of Plant Science* **163**: 197–214.

Sanderson, M. J., Wojciechowski, M. F., Hu, J. M., Sher Khan, T., and Brady, S. G. 2000. Error, bias and long-branch attraction in data for two chloroplast photosystem genes in seed plants. *Molecular Biology and Evolution* **17**: 782–797.

Seelanan, T., Schnabel, A., and Wendel, J. F. 1997. Congruence and consensus in the cotton tribe. *Systematic Botany* **22**: 259–290.

Soltis, D. E., Senters, A. E., Zanis, M. J., Kim, S., Thompson, J . D., Soltis, P .S., Ronse de Craene, L. P., Endress, P. K. and Farris, J. S. 2003. Gunnerales are sister to other core eudicots: implications for the evolution of pentamery. *American Journal of Botany* **90**: 461–470.

Stefanovic, S. M.. Jager, J., Deutsch, J.. Broutin, J., and Masselot, M. 1998. Phylogenetic relationships of conifers inferred from partial 28S rRNA gene sequences. *American Journal of Botany* **85**: 688–697.

Stevenson, D. W. 1990. Morphology and systematics of the Cycadales. *Memoirs of the New York Botanical Garden* **57**: 8–55.

Stevenson, D.W. 1992. A formal classification of the extant cycads. *Brittonia* **44**: 220–223.

Stewart, W. N., and Rothwell, G. W. 1993. *Paleobotany and the Evolution of Plants*. Second edition. Cambridge University Press, Cambridge, United Kingdom

Sugiura, C., Kobayashi, Y., Aoki, S., Sugita, S., and Sugita, M. 2003. Complete chloroplast DNA sequence of the moss *Physcomitrella patens*: Evidence for the loss and relocation of *rpoA* from the chloroplast to the nucleus *Nucleic Acids Research* **31**: 5324–5331.

Swofford, D. L., 2002. "PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4.0b10," Computer program and documentation. Sinauer Associates, Sunderland, MA.

Tavaré, S., 1986. Some probabilistic and statistical problems on the analysis of DNA sequences. *Lectures in Mathematics and Life Sciences* **17**: 57–86.

Treutlein, J., and Wink, M. 2002. Molecular phylogeny of cycads inferred from *rbc*L sequences. *Naturwissenschaften* **89**: 221–225.

Wendel, J. F. and Doyle, J. J. 1998. Phylogenetic incongruence: Window into genome history and molecular evolution. pp. 265-296 In: *Molecular Systematics of Plants II: DNA sequencing*. (D. E. Soltis, P. S. Soltis and J. J. Doyle, eds). Kluwer Academic Publishers, Norwell, MS.

Wendel, J. F., Schnabel, A., and Seelanan, T. 1995. An unusual ribosomal DNA sequence from *Gossypium gossypioides* reveals ancient, cryptic, intergenomic introgression. *Molecular Phylogenetics and Evolution* **4**: 298–313.

Wheeler, J.A. 2000. Molecular phylogenetic reconstructions of the marchantioid liverwort radiation. *Bryologist* **103**: 314–333.

Wolf, P. G., Rowe, C. A., Sinclair, R. B., and Hasebe, M. 2003. Complete nucleotide sequence of the chloroplast genome from a leptosporangiate fern, *Adiantum capillus-veneris* L. *DNA Research* **10**: 59–64.

Wolfe, K. H., Li. W-H., and Sharp, P. M. 1987 Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast and nuclear DNAs. *Proceedings of the National Academy of Sciences, USA* **84**: 9054–9058.

Yang, Z. 1994. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: Approximate methods. *Journal of Molecular Evolution* **39**: 306–314.

Zanis, M. J., Soltis, P. S., Qiu, Y. L., Zimmer, E. and Soltis, D. E. 2003. Phylogenetic analyses and perianth evolution in basal angiosperms. *Annals of the Missouri Botanic Garden* **90**: 129—150.

Table 3.1. GenBank accession numbers and vouchers for new sequences of exemplar cycad and angiosperm taxa. [1] Sequences in bold were generated by other lab groups using different source material. Voucher abbreviations refer to standard herbarium acronyms

	Gene or region							
Taxon [Voucher (herbarium)]	*atp*B	*ndh*F	*psb*B, T, N, & *psb*H	*psb*D & C	*psb*E, F, L & *psb*J	*rbc*L	*rpl*2	3'-*rps*12, *rps*7, *ndh*B & *trn*L
Cycads								
Lepidozamia hopei Regel [69428M (FTG)]	AY699124	AY699157	AY699145	AY699172	AY699148	AY699175	AY699151	AY699154
Macrozamia moorei F. Muell [59302 (FTG)]	AY699125	AY699158	AY699146	AY699173	AY699149	AY699176	AY699152	AY699155
Microcycas calocoma (Miq.) A. DC. [77404T (FTG)]	AY699126	AY699159	AY699147	AY699174	AY699150	AF531214	AY699153	AY699156
Angiosperms								
Arabidopsis thaliana (L.) Heynh. [SWG IX 97 (1) (WTU)	**NC_000932**	AF238049	AY007458	AF239774	AY007473	**NC_000932**	AY007488	AF238063

Taxon [Voucher (herbarium)]	Gene or region							
	*atp*B	*ndh*F	*psb*B, T, N, & *psb*H	*psb*D & C	*psb*E, F, L & *psb*J	*rbc*L	*rpl*2	3'-*rps*12, *rps*7, *ndh*B & *trn*L
Angiosperms (contd.)								
Trimenia moorei (Oliv.) Philipson [P Weston 433770 (NSW)]	AY116653	AY116655	AY116656	AY116657	AY116652	AY116658	AY116659	AY116654

[1] The following species were used by Rai et al. (2003): Angiosperms (*Amborella trichopoda, Acorus calamus, Austrobaileya scandens, Cabomba caroliniana, Ceratophyllum demersum, Hydrastis canadensis, Illicium parviflorum, Nymphaea odorata*); Conifers (*Cedrus deodara, Metasequoia glyptostroboides, Pinus thunbergii, Podocarpus chinensis, Sciadopitys verticillata*); Cycads (*Bowenia serrulata, Ceratozamia miqualiana, Cycas revoluta, Dioon purpusii, Encephalartos barteri, Stangeria eriopus, Zamia furfuracea*); *Ginkgo biloba*; Gnetales (*Ephedra nevadensis, Gnetum gnemon, Welwitschia mirabilis*); *Marchantia polymorpha* and *Psilotum nudum;* see Graham and Olmstead (2000a, b) and Rai et al. (2003) for GenBank accession numbers and source details. The following sequences come from other studies: *Adiantum capillus-veneris* (Wolf et al. 2003), *Anthoceros formosae* (Kugita et al. (2003) and *Physcomitrella patens* (Sugiura at al. 2003).

Table 3.2. Genbank accession numbers for the *trn*L-F region and the *RPB2* and 26S rDNA loci. Voucher information for the *RPB2* sequences is provided in Chapter 2. Source information for all other sequences can be found in the original publications, noted in the footnotes.

	*trn*L-F [1]	*RPB2* [2]	26S rDNA [3]
Bryophyte			
Marchantia polymorpha L.	-	AF020844	AF226020
Pteridophyte			
Psilotum nudum (L.) Beauv.	-	AY699182	-
~ Seed plants ~			
Angiosperms			
Acorus calamus L.	-	-	AF203679
& *A. gramineus* Aiton	-	AY699222	-
Amborella trichopoda Baill.	-	AY699216	AF479238
Arabidopsis thaliana (L.) Heynh.	-	Z19120	-
Austrobaileya scandens C. T. White	-	-	AY292886
Cabomba caroliniana A. Gray	-	-	AF479239
Ceratophyllum demersum L.	-	-	AF479228
Hydrastis canadensis L.	-	-	AF389268
Illicium anisatum L.	-	AY699220	
Nymphaea odorata Aiton	-	AF043427	
& *N. nouchali* Burm. f.	-	-	U90711
Trimenia moorei (Oliv.) Philipson	-	AY699217	AY095470
Conifers			
Cedrus libani A. Rich.	-	-	AY056507
Metasequoia glyptostroboides Hu & W. C. Cheng	-	AY699213	AY056512
Pinus peuce Griseb.	-	-	AY056499
Podocarpus coriaceus Rich. & A. Rich.	-	AY699203	-
& *P. macrophyllus* (Thunb.) Sweet	-	-	U90685
Sciadopitys verticillata (Thunb.) Siebold & Zucc.	-	AY699205	U90698
Cycads			
Bowenia serrulata (W. Bull) Chamb.	AF531185	-	AF531247
Ceratozamia miqueliana H. Wendl.	AF531192	AY699191	AF531252
Cycas revoluta Thunb.	AF531181	AY563265	U90673
Dioon edule Lindl.	AF531186	AY699197	AY056483

	*trn*L-F [1]	*RPB2* [2]	26S rDNA [3]
Cycads (contd)			
Encephalartos arenarius R. A. Dyer	-	-	AF531249
& *E. barteri* Carruth. ex. Miquel	-	AY699192	-
& *E. ferox* Bertol. f.	AF531190	-	-
Lepidozamia peroffskyana Regel	AF531191	-	AF53125
Macrozamia moorei F. Muell	-	AY699193	-
& *M. pauli-guilielmi* W. Hill & F. Muell.	AF531189.1	-	AF531248
Microcycas calocoma (Miq.) A. DC.	AF531194	AY699194	AF531254
Stangeria eriopus Nash.	AF531184	AY699195	U90675
Zamia floridana A. DC.	-	AY699196	-
& *Z. skinneri* Warsz. ex A. Dietr.	AF531197.1	-	AF531257.1
Ginkgo			
Ginkgo biloba L.	AY145323	AF020843	AY095475
Gnetales			
Ephedra distachya L.	-	AY699198	AF036489
Gnetum gnemon L.	-	AY563267	AF036488
Welwitschia mirabilis Hook. f.	-	AY699199	AY056484

[1] The *trn*L-F region consists primarily of the *trn*L(UAA) intron and the intergenic spacer region between *trn*L(UAA) and *trn*F(GAA). All cycad *trn*L-F sequences are from Hill et al. (2003) and for *Ginkgo* from Borsch et al. (2003).

[2] References for *RPB2* sequences: *Arabidopsis* (Larkin and Guilfoyle, 1993); *Cycas* and *Gnetum* (Oxelman and Bremer, in press); *Ginkgo, Marchantia* and *Nymphaea* (Denton et al., 1998).

[3] References for 26S rDNA sequences: *Acorus* (Neyland, 2002); *Amborella, Cabomba* and *Ceratophyllum* (Soltis et al., 2003); *Austrobaileya* (Qiu et al., unpublished data); *Bowenia, Ceratozamia, Encephalartos, Lepidozamia* and *Microcycas* (Hill et al., 2003); *Cedrus, Dioon, Metasequoia, Pinus* and *Welwitschia* (Rydin et al., 2000); *Cycas, Nymphaea, Podocarpus, Sciadopitys, Stangeria* and *Zamia* (Stefanovic et al., 1998); *Ephedra* and *Gnetum* (Kuzoff et al., 1998); *Ginkgo* and *Trimenia* (Zanis et al., 2003); *Hydrastis* (Kim et al., 2004); *Marchantia* (Wheeler, 2000)

Table 3.3. Likelihood ratio test (LRT) for various substitution models. Likelihood scores and parameter estimates were determined on an MP topology for cycads and *Ginkgo* only. The topology was identical to the one shown in Fig. 3.2. The significance value for rejection of the null hypothesis was adjusted using a Bonferroni correction and set to 0.01. In all cases, $P < 0.01$.

Substitution Model [1]	**-ln likelihood**	**Comparison**	**-2 ln Λ** [2]	**d.f.**
JC69	49576.98	-----	-----	
F81	49295.99	JC69 vs. F81	561.98	3
HKY85	47866.85	F81 vs. HKY85	2858.28	1
GTR	47746.55	HKY85 vs. GTR	240.60	5
GTR + Γ	47211.32	GTR vs. GTR + Γ	1070.46	1
GTR + Γ + I	47194.80	GTR + Γ vs. GTR + Γ + I	33.04	1

[1] Abbreviations: JC69 = Jukes-Cantor (1969); F81 = Felsenstein (1981); HKY85 = Hasegawa et al. (1985); GTR = General Time-Reversible (Lanave et al., 1984; Tavaré, 1986; Barry and Hartigan, 1987; Rodríguez et al., 1990). Γ = Gamma.

[2] Likelihood ratio test statistic.

Figure Legends

Figure 3.1. Inference of seed-plant phylogeny using maximum-parsimony analysis of a large plastid data set and two nuclear loci (*RPB2* and 26S rDNA) combined. Tree length = 25,942 steps, consistency index (CI) = 0.530, retention index (RI) = 0.620. Bootstrap values are indicated above branches. The bootstrap values below branches, or on the right, were obtained from an analysis using plastid data only. The branch subtending the clade consisting of *Ceratozamia*, *Microcycas* and *Zamia* collapses in a strict consensus of two most-parsimonious trees inferred using only plastid data (tree length = 21,412, CI = 0.541, RI = 0.643).

Figure 3.2. Maximum-parsimony trees inferred for cycads and other seed plants based on different data codon-position data partitions (a, b) or different types of character-state change across all of the protein-coding data (c, d). (a) Plastid codon positions 1 and 2 combined. Two most-parsimonious trees were inferred (tree length = 26,077; CI = 0.527; RI = 0.615). (b) Plastid codon position 3. Two most-parsimonious trees were inferred (tree length = 25,985; CI = 0.529; RI = 0.618). (c) Plastid transversions only. One most parsimonious trees was inferred (tree length = 8,748; CI = 0.550; RI = 0.597). (d) Plastid transitions only. One most parsimonious tree was inferred (tree length = 26,041; CI = 0.258; RI = 0.616). The arrowheads point to nodes that collapse in strict consensus trees.

Figure 3.3. Tree inferred using Bayesian phylogenetic inference of a large plastid data set, either alone or in combination with two nuclear genes, the *RPB2* and 26S rDNA loci. Estimated posterior probabilities (expressed as percentages) are shown beside branches (above branches for all data combined, below for all

plastid data combined). The lefthand values in each case are those inferred when one set of estimated model parameters was used to describe all data subsets (GTR + Γ + I model). The second set of values in each case are those inferred when DNA substitution model parameters are estimated separately for six partitions (photosystem II genes; plastid inverted repeat genes; all other chloroplast genes; plastid noncoding regions; *RPB*2; 26S rDNA) or for plastid data alone using only four partitions (the same ones). The lower righthand values are those inferred when DNA substitution model parameters are estimated separately for two different codon partitions (positions 1 and 2 combined, vs. codon position 3) and the plastid noncoding data. An asterisk indicates when all values for a particular data set were estimated to be 100%.

Figure 3.4. Tree inferred using Bayesian phylogenetic inference for two nuclear genes: (a) 26S rDNA; (b) *RPB*2. Estimated posterior probabilities (expressed as percentages) are shown beside branches.

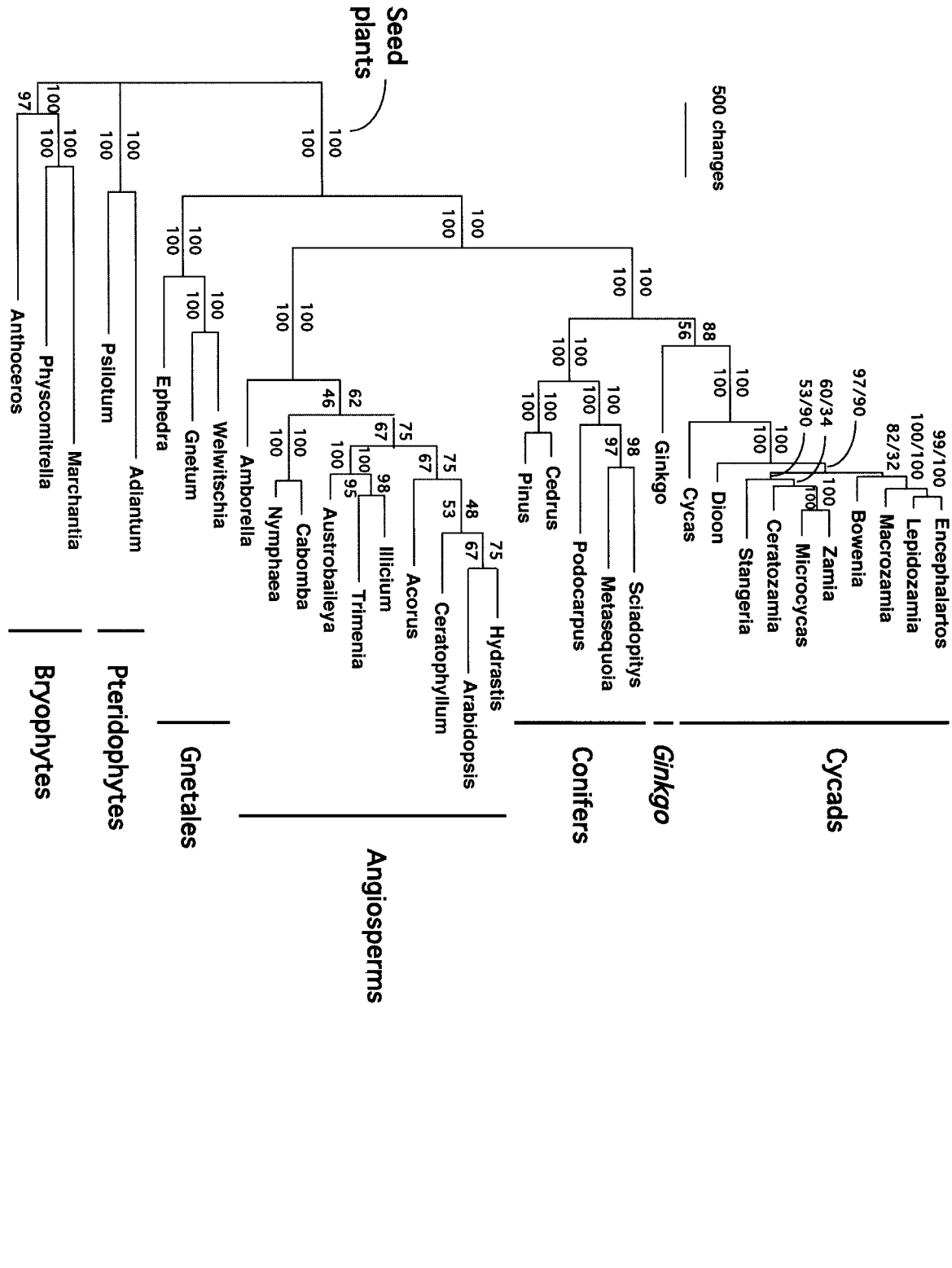


Fig. 3.1

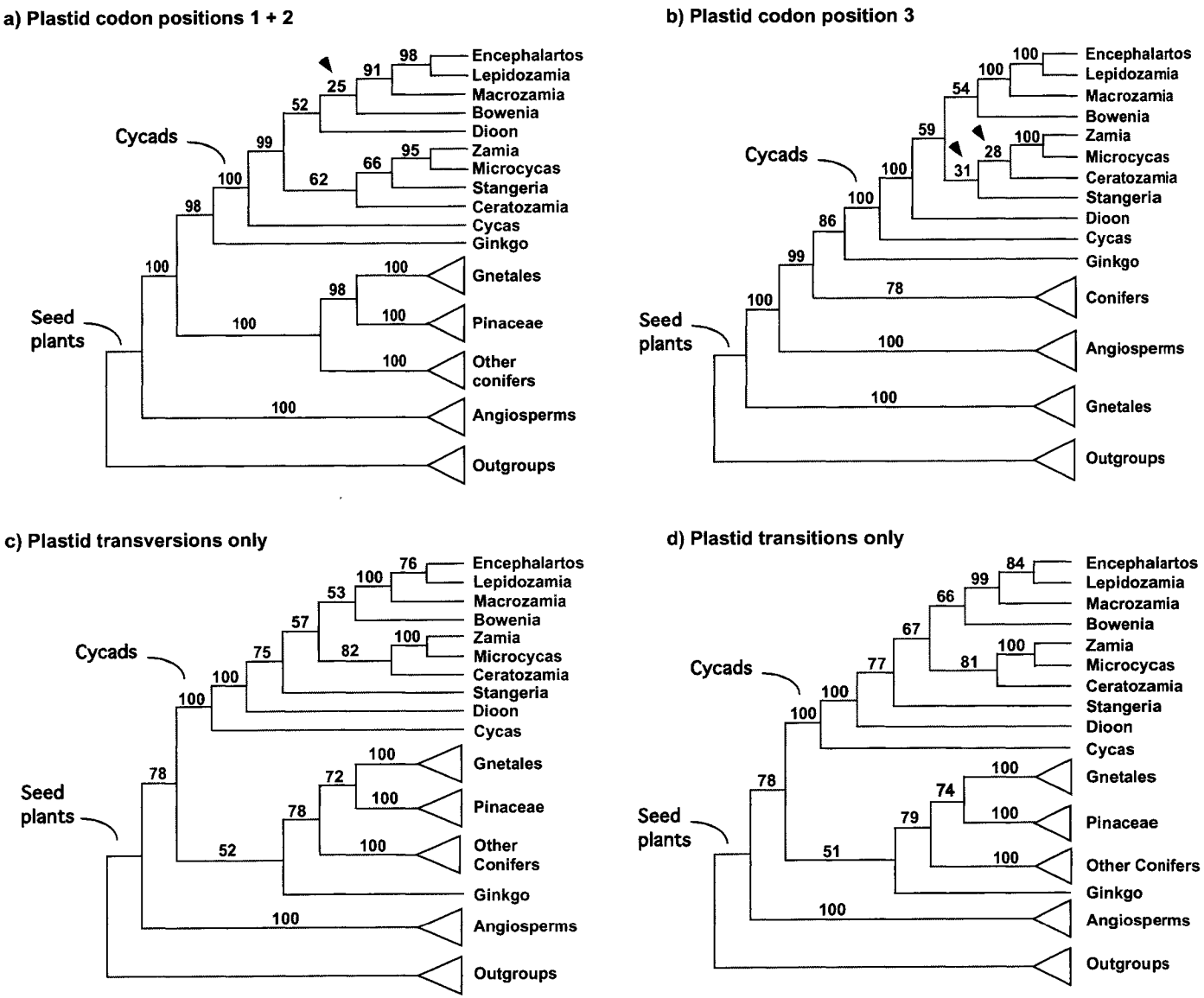


Fig. 3.2

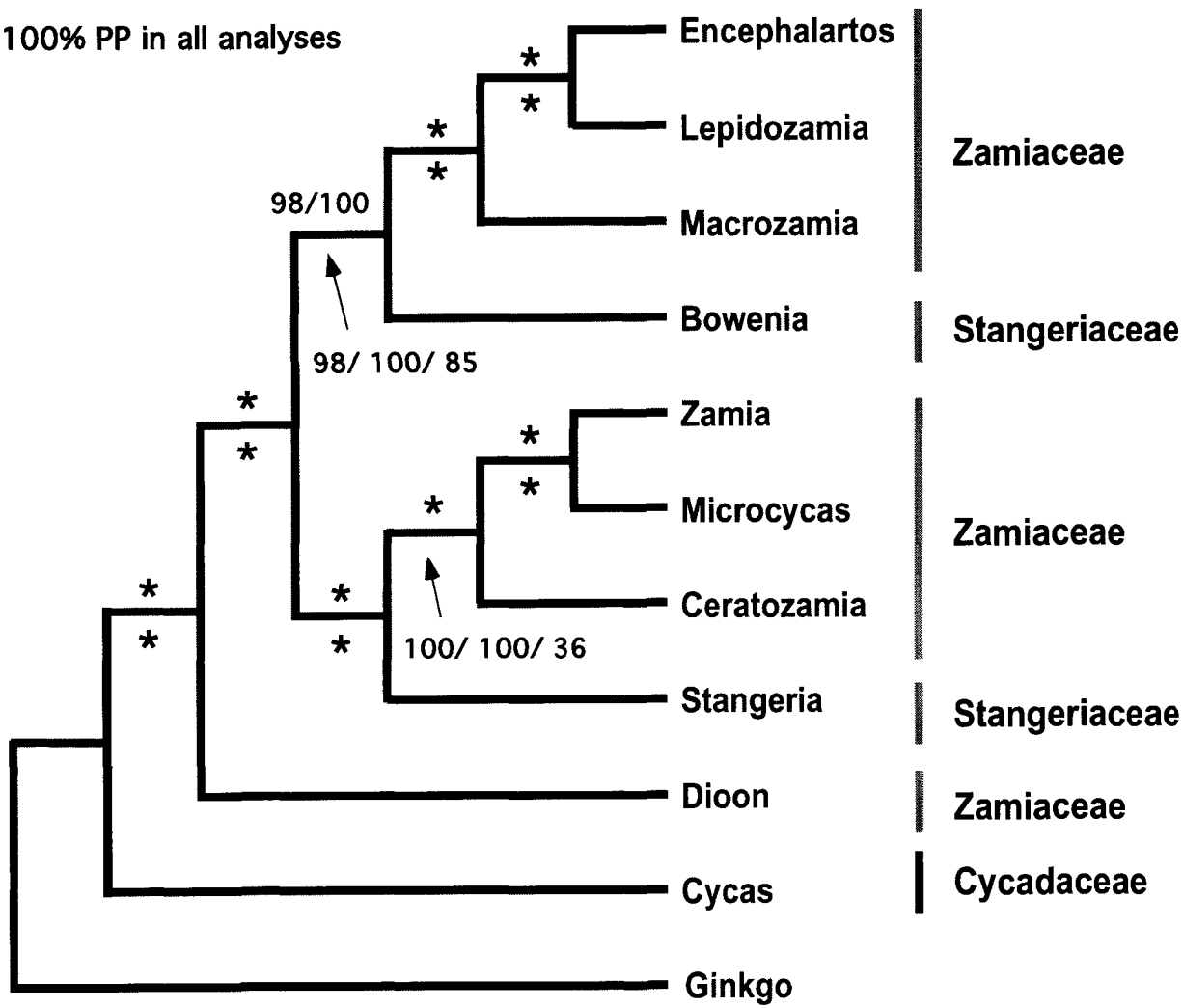


Fig. 3.3

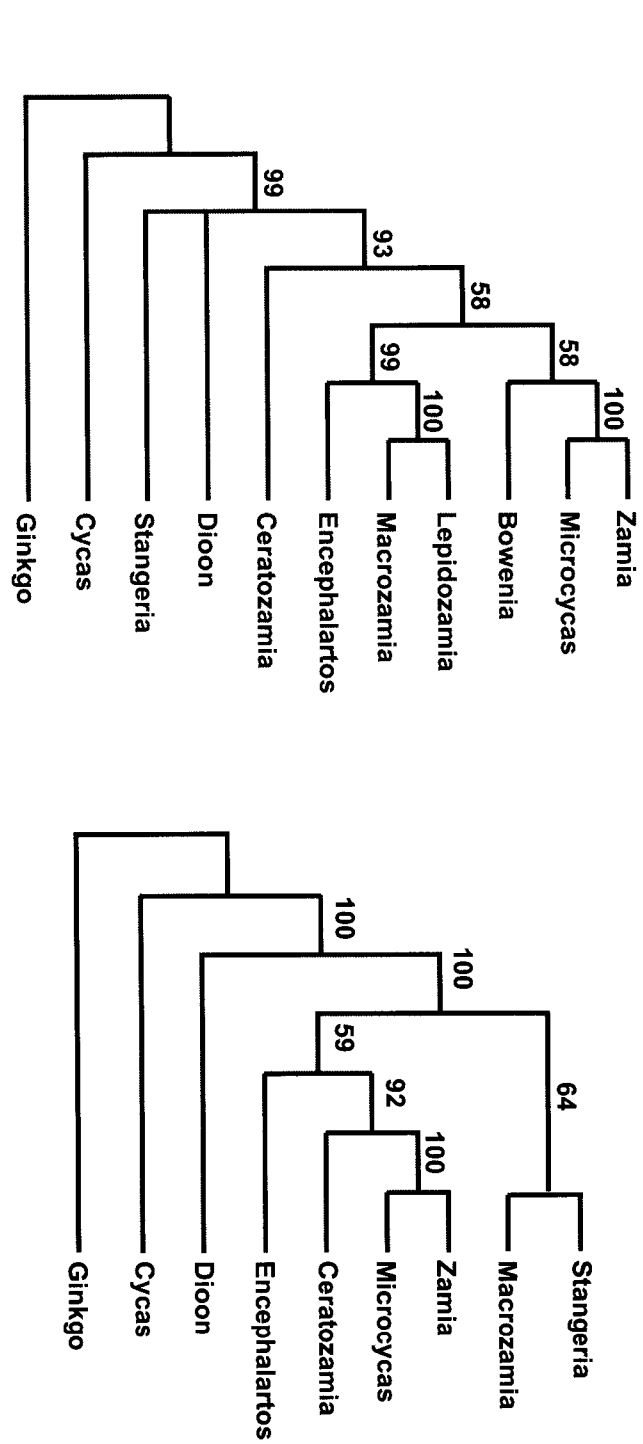


Fig. 3.4

~ Chapter Four ~

Higher-order Phylogenetic Relationships in the Monocot Order Liliales, as Inferred from a Large Plastid Data Set

Introduction

The order Liliales is a well-known group of petaloid monocots, the members of which typically have striking floral displays associated with insect pollination. It includes numerous horticultural favorites, such as *Alstroemeria, Gloriosa, Colchicum, Lilium, Trillium* and *Tulipa*. The order as currently recognized (Angiosperm Phylogeny Group II, APG II, 2003) consists of approximately 1,300 species in ten families: Alstroemeriaceae, Campynemataceae, Colchicaceae, Corsiaceae, Liliaceae, Luzuriagaceae, Melanthiaceae, Philesiaceae, Rhipogonaceae and Smilacaceae. This delineation departs substantially from well known and widely taught circumscriptions based on pre-cladistic morphological methodology and data, which place many superficially similar, but distantly related taxa into one large family, Liliaceae (e.g., Cronquist, 1988). Cronquist defined the order to include 15 families with a total of ~8,000 species. He placed approximately half of these species, representing ~280 genera, in Liliaceae. In his view, this family consisted largely of perennial, geophytic and herbaceous plants, with perfect flowers possessing six stamens, three fused carpels and a superior ovary.

More recent studies based on morphology (e.g., Dahlgren et al., 1985) and molecular data (Duvall et al., 1993a; Chase et al., 1995a; APG II, 2003) have placed

many of these "lilioid" monocots into other orders, and Dahlgren et al. (1985) and APG II (2003) have defined Liliales and its constituent families (particularly Liliaceae) much more narrowly. Many "lilioid" taxa now find a home in Asparagales, an order not recognized by Cronquist (1988). The treatment of Asparagales and Liliales by Dahlgren et al. (1985), who overhauled monocot classification primarily using morphological evidence, is largely congruent with recent, largely molecular-phylogenetic studies (e.g., APG, 1998; Chase et al., 1995b; Fay et al., 2000; Rudall et al., 2000; APG II, 2003; McPherson, 2003). The recognition of Asparagales in its modern sense goes back to Huber (1969), who promoted it based on several seed-coat characteristics. Most Asparagales (including many taxa placed in Liliales by Cronquist) have a collapsed inner seed coat and a phytomelan crust on the outer epidermis, although this crust is lacking in most taxa with fleshy fruits, and the entire outer epidermis is lacking in taxa with baccate fruits (Dahlgren et al., 1985).

However, no single morphological feature has been found to suffice for distinguishing members of Asparagales and Liliales from each other, in part because of homoplasy in several putative morphological synapomorphies of each order. In practice, however, several micro- and macromorphological characters considered together aid in differentiating members of the two orders (see Dahlgren et al., 1985). For example, members of Asparagales can be succulent or may possess an unusual mode of secondary thickening, usually have nectaries in the septa of their ovaries, and always lack a sarcotesta. In contrast, members of Liliales are never succulent (although some are woody), and they usually have nectaries present on the base of their tepals or filaments, and occasionally have a sarcotesta. Additionally, most Liliales have extrorse anther dehiscence, and many have spotted tepals. These observations are with respect to the ordinal definitions of Dahlgren et al. (1985), but

generally hold for the APG II (2003) circumscriptions, with some further exceptions. For example, spotted tepals are rare at the family level in Asparagales, but are well known in Orchidaceae.

The monocot classifications of Dahlgren et al. (1985) and APG II (2003) are highly congruent with each other, particularly when either is compared to Cronquist's (1988) system (see Table 4.1). Nonetheless, multiple differences exist between these two classifications. Dahlgren et al. (1985), for example, included Iridaceae and Orchidaceae in Liliales (families now recognized as Asparagales), and several taxa now considered to be part of Liliales by APG II (2003) were thought by Dahlgren et al. (1985) to belong to other orders. Examples of these include Campynemataceae, Corsiaceae, Luzuriagaceae and Philesiaceae. The latter two were placed in Asparagales and the former two were placed in Melanthiales and Burmanniales, respectively. In addition, several taxa with family status in Liliales according to Dahlgren et al. (1985) have since been subsumed in other families of Liliales. Examples include Calochortaceae, which is now a part of Liliaceae *sensu* APG II (2003), and Uvulariaceae, members of which have been largely been placed in Colchicaceae or Liliaceae (see Table 4.1).

Although the familial composition of Liliales is now relatively clear (APG II, 2003), and the monophyly of the order is generally well supported (Chase et al., 2000), the precise relationship of Liliales to the other monocot orders has been difficult to determine (Chase et al., 1993; Soltis et al., 2000; Hilu et al., 2003). However, two recent studies have found the order to be the sister group of a large clade composed of Asparagales and the commelinid monocots (Chase et al., 2000; Fuse and Tamura, 2000) with somewhat weak bootstrap support. Fay et al. (2000) found the same relationship, with better support, but they focussed on Asparagales and included representatives of fewer orders of monocots. McPherson (2003), who

also focussed on Asparagales, found a well-supported sister group relationship between the commelinids and Asparagales, and found Liliales to be the sister group of this clade, but with only moderate support for the latter relationship from maximum-parsimony bootstrap analyses.

Several recent molecular and morphological studies have clarified phylogenetic relationships within Liliales. For example, Rudall et al. (2000) used morphological characters and plastid sequences (*rbc*L and the *trn*L-*trn*F region) to infer relationships within the order. Four major clades were recognized: (1) A group centred around Liliaceae that also included Philesiaceae, Smilacaceae, Calochortaceae and a portion of Uvulariaceae; (2) Campynemataceae; (3) A group centred around Colchicaceae, that also included *Petermannia cirrosa*, Alstroemeriaceae, a redefined Luzuriagaceae, and portions of Uvulariaceae, and; (4) Melanthiaceae. However, relationships among and within each of these groups were often unclear or poorly supported by bootstrap analyses.

The major goals of this study are to address current outstanding problems regarding Liliales phylogeny using an expanded plastid data set relative to previous studies for representatives of all of the major lineages of the monocotyledons. While the majority of studies addressing monocot phylogenetic relationships use sequences from one to several genes or regions for a moderately large number of taxa, our study substantially expands the number of characters sampled per taxon by including plastid data that represent approximately one-tenth of the genome. Increasing the amount of data per taxon has demonstrated potential for increasing our ability to infer accurate, well-supported trees, based on theoretical (e.g., Poe and Swofford, 1999; Sanderson et al., 2000) and empirical studies (e.g., Graham et al., 1998; Bremer et al., 1999). The plastid data set used in the current study has provided new and robust inferences for deep phylogenetic relationships of a number

of groups, including the basal angiosperms (Graham et al., 2000; Graham and Olmstead, 2000), the cycads (Rai et al., 2003), and Asparagales (McPherson, 2003). Here, it is used to infer the placement of Liliales along the backbone of monocot phylogeny and to determine the higher-order relationships among its constituent families.

Materials and Methods

Taxon sampling.

Taxa were chosen to exemplify the taxonomic and phylogenetic diversity of the Liliales based primarily on the classification scheme of APG II (2003), which we follow for the remainder of this paper. A minimum of one species was chosen to represent each family in the order, with the exception of the achlorophyllous family Corsiaceae. When multiple species were chosen for a family, they were selected to exemplify its phylogenetic and morphological diversity. These families include Colchicaceae, Liliaceae, and Melanthiaceae. One species of Colchicaceae was chosen from the "uvularioid" group (*Tripladenia cunninghamii*) and one was chosen from the "wurmbaeoid" group (*Wurmbaea pygmaea*), as defined by Nordenstam (1998). An additional member of Colchicaceae, *Petermannia cirrosa,* was chosen because, although it is included in Colchicaceae by APG II (2003), it has been considered sufficiently distinctive to be placed in its own family, Petermanniaceae (Conrad and Clifford, 1998).

Outgroups were chosen to exemplify monocot diversity at the ordinal level based on APG II (2003). Many of these sequences were first obtained and published by McPherson (2003). The final matrix contained 16 Liliales, 28 Asparagales, 20 commelinids, three Dioscoreales, four Pandanales, five Alismatales, two Petrosaviaceae, and *Acorus calamus* (Acorales). The taxonomic placement of the genera included here according to three widely known taxonomic systems (Cronquist, 1988; Dahlgren et al., 1985; and APG II, 2003) is shown in Table 4.1. The monocot rooting assumed here (at the branch leading to *Acorus*) is that seen in most recent studies (Chase et al., 1993; Duvall et al., 1993b; Chase et al., 2000; Graham and Olmstead, 2000).

Genomic sampling.

Substantial or complete portions of 17 protein-coding plastid genes were sampled here, together with 14 of their associated intergenic spacers and four of their introns. Voucher and GenBank accession numbers for sequences used in this study are provided in Table 4.2. Coding regions used include: subunit ß of ATP synthase (*atp*B), two NADH dehydrogenase subunit genes (*ndh*B and *ndh*F), ten photosystem II genes (*psb*B, *psb*C, *psb*D, *psb*E, *psb*F, *psb*H, *psb*J, *psb*L, *psb*N, and *psb*T), three ribosomal protein genes (*rpl*2, *rps*7,and 3'- *rps*12), and the large subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase (*rbc*L). The non-coding regions sampled here include: the introns from *ndh*B, *3'-rps*12, and *rpl*2, and the intergenic regions spanning protein coding genes in the *psb*B–*psb*T–*psb*N–*psb*H cluster, the *psb*E–*psb*F–*psb*L–*psb*J operon and the 3'*rps*12-*rps*7-*ndh*B region. Three primarily noncoding regions were also included for members of Asparagales, Liliales and a subset of the commelinid taxa. These

noncoding regions are the intergenic spacer region between *ndh*B and *trn*L (CAA), the intergenic spacer between *rbc*L and *atp*B, and the *trn*L (UAA)-*trn*F(GAA) region [including the *trn*L(UAA) intron and the intergenic spacer between *trn*L (UAA) and *trn*F (GAA)].

In a few cases in Asparagales, GenBank sequences for the *trn*L-F region were included from a different species in the same genus, or from a different genus in the same family (Table 4.2). Several instances where more than 50 bp of a region are missing for a taxon are also noted in Table 4.2. Most taxa had nearly complete coverage for all the regions considered here. The major exceptions are in *Kingia* and *Ixiolirion* (where sequencing of *rpl*2 was not attempted due to limited material), *Burmannia* (an *ndh*F sequence could not be amplified), and *Petrosavia* (due to our inability to retrieve any *ndh*F or *ndh*B second exon amplification products; several photosystem regions were also amplified for this taxon, but these require further characterization before inclusion in phylogenetic analysis).

Details of exon, intron and spacer lengths and boundaries for most of the regions are provided in Graham and Olmstead (2000), with the exception of the *trn*L-*trn*F region, the *ndh*B-*trn*L intergenic spacer and the *rbc*L-*atp*B intergenic spacer. For members of Liliales examined in this study, the *trn*L-*trn*F region ranged from 1,010 bp in *Alstroemeria aurea* to 721 bp in *Tricyrtis* sp., the *rbc*L-*atp*B intergenic spacer ranged in length from to 599 bp in *Anticlea elegans* to 827 bp in *Luzuriaga radicans*, and the *ndh*B-*trn*L intergenic spacer ranged from 498 bp in *Luzuriaga radicans* to 588 bp in *Petermannia cirrosa*.

DNA extraction, amplification and sequencing.

DNA was extracted from silica-dried or pressed herbarium specimens using the basic protocol of Doyle and Doyle (1987), or was obtained from the Royal Botanic Gardens, Kew (see Fay et al. 2000, for their methods of DNA extraction). DNA amplification was performed using a Gene-E Techne cycler (Techne, Inc., Burlington,

NJ) and a GeneAmp PCR system 9700 (Applied Biosystems, Foster City, CA) with the following profile: (1) Initial denaturing at 94 C for five minutes; (2) 35 cycles of a 95 C denaturation for 30 seconds, followed by annealing at 45 C for one minute and a 72 C extension for two minutes, and; (3) a final extension at 72 C for seven minutes. 25 pmole of primer were used in each 50 μL reaction. Most primers used for amplification and sequencing are those designed by Graham and Olmstead (2000). Other published primers used were those for *atp*B (Hoot et al., 1995), *ndh*F (Olmstead and Sweere, 1994; Kim and Jansen, 1995; Olmstead and Reeves, 1995; Neyland and Urbatsch, 1996 a, b; Graham et al., 1998), *rbc*L (Zurawski et al.,1984), the intergenic spacer between *ndh*B and *trn*L (CAA) (McPherson, 2003), the *trn*L-*trn*F region (Taberlet et al., 1991) and the intergenic spacer region between *rbc*L and *atp*B (Chiang et al., 1998; McPherson, 2003).

A nested PCR approach was used to generate sequences of most regions for *Campynema* and *Petermannia*, because low quantities of amplification product were produced in the initial PCR reactions, and I had a very limited amount of DNA for these taxa. The PCR steps outlined above were performed, and 1 to 2 µL of the PCR product produced were amplified again using primers internal to the ones initially used and the same PCR cycle described above. *Pfu* polymerase was used instead of *Taq* polymerase during the second round of amplification because it has a

better proofreading ability, which minimizes the quantity of error introduced during the multiple rounds of amplification.

All amplification products were purified using QIAquick columns (QIAgen, Inc. Valencia, CA), following the manufacturer instructions. Sequencing products were generated using either the "DYEnamic ET" terminator cycle sequencing kit (Amersham Biosciences, Piscataway, NJ) or the "Big Dye Terminator v 3.1" cycle sequencing kit (Applied Biosystems, Foster City, CA), following manufacturer instructions. Sequencing reactions were cleaned using Sephadex, dried on a vacuum centrifuge and run on an ABI Prism 377 automated sequencer.

All regions were sequenced at least twice, and in the great majority of cases both forward and reverse strands were sequenced. For each amplified region for each taxon, one extra sequence was generated from a second PCR product generated on a different day, typically using a second extraction from the same source, to control for possible errors in extraction or amplification. No such errors were found.

Data assembly.

Contiguous sequences were compiled and called using Sequencher 4.1 (Gene Codes Corporation; Ann Arbor, MI). The consensus sequences for each taxon were added to a previous alignment (McPherson, 2003) and were adjusted manually using Se-Al version 1.0 (Rambaut, 1998). Co-ordinates for gene, intron, and exon boundaries were determined using *Nicotiana tabacum* sequences, following Graham and Olmstead (2000). Two regions that were difficult to align within the *atp*B-*rbc*L intergenic spacer and the *trn*L-*trn*F region were excluded from analyses. The final alignment contained 32,416 bp, or 31,900 bp with exclusion of

the difficult-to-align portions. This corresponds to 16,955 bp of unaligned sequences (reference taxon = *Alstroemeria aurea*). 8,104 of the aligned characters were variable, and 5,061 of these were parsimony informative.

Assessment of phylogenetic congruence.

Various subsets of the data set were assessed for incongruency using the incongruence length difference (ILD) test (Farris et al., 1994) as implemented in PAUP*, and by visual inspection of the bootstrap values. In both cases, comparisons were made between first and second vs. third codon positions and between coding and non-coding regions. The ILD test was completed for each of these with the entire data set and using the taxa in Liliales alone. The test was implemented with a heuristic search using simple step-wise addition, and tree-bisection reconnection (TBR) branch swapping for 100 partition replicates.

Tree search characteristics for parsimony analyses.

Maximum-parsimony (MP) tree searches were conducted using PAUP* version 4.0b10 (Swofford 2002). All characters and character-state changes were equally weighted, and heuristic searches were conducted using TBR branch swapping and 100 random addition replicates. The MP analyses were run on a data set that included all of the data (but excluding the two short unaligned regions), referred to as here as the "complete" data set, and on a somewhat reduced data set that included only the major regions that were sequenced for most taxa (i.e., excluding the *atp*B-*rbcL* intergenic spacer, the *ndh*B-*trn*L (CAA) intergenic spacer and the entire *trn*L-*trn*F region); the latter data set is referred to here as the

"reduced" data set. Several heuristic searches were also run with constraints applied to several taxa of interest that were not found in the most parsimonious trees.

Additional MP analyses were run for data sets containing : (1) the first and second codon positions; (2) only the third codon positions; (3) all codon positions, and; (4) non-coding regions included in the "reduced" data set. A parsimony-based bootstrap (BP) analysis (Felsenstein, 1985) was performed on the complete data set, on the reduced data set and for these four data partitions, using the same search conditions, except that only one random addition replicate was employed for each of the 1,000 bootstrap replicates.

Bayesian analysis.

A Bayesian analysis was performed using the program "MrBayes v. 3.0" (Ronquist and Huelsenbeck, 2003). The data set was reduced to 58 taxa to facilitate analysis in a reasonable time frame. The reduced data set contained: *Acorus calamus*, *Japanolirion osense*, five Alismatales, two Pandanales, two Dioscoreales, 16 Liliales, 14 commelinds and 17 Asparagales. All nucleotide data were used. The analysis was run using four chains, one million generations, and the default temperature (0.2). The general time reversible (GTR; Lanave et al., 1984; Tavaré, 1986; Barry and Hartigan, 1987; Rodríguez et al., 1990) model of sequence evolution with a gamma-distribution of rates was used, based on the result of a likelihood-ratio test (Huelsenbeck and Crandall, 1997; see Table 4.3). The first 20,000 trees were rejected and the results were summarized on a majority-rule consensus tree.

Results

Phylogenetic position of Liliales in the monocots.

We found two most parsimonious trees using the "complete" data set (Fig. 4.1). The two topologies differ only in the relative positions of several closely related families of Asparagales. The sampled members of Liliales (*sensu* APG II, 2003) together comprise a well-supported clade (bootstrap value, BV = 99%; posterior probability, PP = 100%). There is moderate to robust support (79% BV, 100% PP) for Liliales as the sister group of a large clade consisting of Asparagales and the commelinid monocots. The Asparagales-commelinid clade is very well supported (BV = 97%, PP = 100%). All deeper relationships in the monocots are also robustly supported, with one possible exception—the sister-group relationship inferred here between Dioscoreales and Pandanales. The latter clade has only weak bootstrap support (BV = 62%) but has substantial support from the Bayesian analysis (PP = 96%).

A large clade consisting of Asparagales, commelinids, Dioscoreales, Liliales and Pandanales was robustly supported (BV = 92%; PP = 100%). The two members of Petrosaviaceae (once considered part of Melanthiaceae, Dahlgren et al., 1985) sampled here form a well supported clade (BV = 100%), and this small family is robustly supported (BV = 100%; PP = 100%) as the sister group of all monocots excluding *Acorus* and Alismatales. The relationships within the large clades representing each order of monocots are consistent with earlier observations for these data (McPherson, 2003) and are not discussed further here.

Higher-order relationships in Liliales.

The deepest branches in Liliales all find robust support from parsimony bootstrap and Bayesian analyses (with bootstrap support values from 88–100% and posterior probabilities from 93–100%; Fig. 4.1). Campynemataceae, represented here by *Campynema lineare*, arise from the deepest bifurcation in the order, and Melanthiaceae (represented here by *Anticlea elegans* and *Trillium grandiflorum*) are the sister group of the remaining families in the order. A large clade consisting of Alstroemeriaceae, Colchicaceae, Luzuriagaceae and *Petermannia cirrosa* is the sister group of the clade of families consisting of Liliaceae, Philesiaceae, Rhipogonaceae and Smilacaceae. The two sampled taxa from Alstroemeriaceae and Luzuriagaceae are sister groups, and this clade is the sister group of Colchicaceae. *Petermannia cirrosa* is not part of Colchicaceae, but is well supported as the sister group of the clade consisting of Alstroemeriaceae, Luzuriagaceae, and Colchicaceae.

Most relationships within Liliales are very well supported by the "complete" data set in parsimony and Bayesian analyses. The most poorly supported relationships concern the relationships among three of the five exemplar taxa sampled for Liliaceae, and the position of Smilacaceae among a cluster of four families including Liliaceae, Philesiaceae, Rhipogonaceae and Smilacaceae. In the former case, only a *Lilium-Medeola* clade is robustly supported by parsimony bootstrap and Bayesian analysis (BV = 100%; PP = 100%). However, Liliaceae as a whole are very well supported (BV = 100%, PP = 100%). A clade consisting of *Prosartes* and *Tricyrtis* has poor support from bootstrap analysis (BV = 42%) although it is well supported in the Bayesian analysis (PP = 99%). *Calochortus* has substantial support as the sister taxon of *Lilium-Medeola* in the

Bayesian analysis (PP = 81%), but not in the parsimony analysis (BV = 41%). The shortest tree found with *Prosartes, Calochortus* and *Tricyrtis* constrained as a clade was only two steps longer than the shortest MP trees.

Philesiaceae and Rhipogonaceae are together supported as the sister group of Liliaceae, but with only weak to moderate support (BV = 60%; PP = 89%). However, a large clade consisting of Liliaceae, Philesiaceae, Rhipogonaceae and Smilacaceae is robustly supported (BV = 100%; PP = 100%), and the Philesiaceae-Rhipogonaceae clade is well-supported in its own right (BV = 100%; PP = 100%). The shortest tree found with Smilacaceae, Rhipogonaceae and Philesiaceae constrained as a clade was only three steps longer than the shortest MP trees.

Congruence among different data partitions.

The incongruence length difference (ILD) test did not indicate any significant incongruence between codon positions 1 + 2 vs. codon position 3 (P = 0.86), or for the coding vs. non-coding partition (P = 0.31) when only the Liliales were considered, although significant results were obtained when all monocots were considered (P = 0.02 and P = 0.01, respectively). The latter results should be viewed cautiously, as the test is known to have an excessively high Type I error rate (e.g., Graham et al. 1998; Barker and Lutzoni 2002).

A comparison of bootstrap values for clades in Liliales and in other monocot clades nearby based on the different data partitions (Figs. 4.2, 4.3) illustrates that there is little evidence of conflict among the subsets of the phylogenetic data. Figure 4.3 depicts bootstrap support values for all four data partitions, and Bayesian posterior probabilities for the complete data set (this is not to imply that bootstrap support values and Bayesian posterior probabilities are equivalent; however, they

share the same absolute scale, can be conveniently plotted together, and their values generally appear to be correlated with each other.)

All branches found in the shortest maximum parsimony trees have high posterior probabilities from the "complete" data set (81–100%; Figs. 4.2, 4.3). Bootstrap support values from this data set for the clades considered here are all high, with the exception of three branches that have only weak to moderate support (branches "o," "r," and "s"; Figs. 4.2, 4.3). With the exception of branch "o," none of these branches is well supported by individual subsets of the data. This branch has less bootstrap support from the fully combined data set (60%) than from the analysis of noncoding data alone (99%). This is the strongest hint of incongruence among the partitions. The other data partitions support branches that weakly conflict with branch "o."

Eight clades have > 90% bootstrap support from the two codon partitions considered here and from the noncoding data (Fig. 4.3). All branches considered in Figs. 4.2 and 4.3 had some bootstrap support from all the data subsets under consideration. The "complete" data set (all of the former regions combined) had a mean bootstrap value of 88.9% (and mean posterior probability of 98.1% across branches "a" – "s"). The mean bootstrap value for branches "a" through "s" was highest for codon position 3 (77.3%) and lowest for the first two codon positions combined (56.8%). The comparable mean bootstrap support for the noncoding data was 73.8%, and 79.5% for coding regions alone (the latter is not shown in Fig. 4.3).

The "reduced" data set (all regions in the complete data set except for three spacer regions not shared by all taxa; see Materials and Methods and Table 4.1) had a mean bootstrap value of 84.4% for these clades and in general the three non-coding regions contributed only a small amount to the robustness of inferred

branches. However, the three extra noncoding regions appear to have contributed to placing Melanthiaceae robustly in Liliales (branch "i" has only 43% support without these regions and 88% with them; not shown in Fig. 4.3).

Discussion

Liliales are robustly supported as a monophyletic group that is the sister group of a clade composed of Asparagales and commelinids. This result is congruent with other recent studies of monocot phylogeny (Fay et al., 2000; McPherson et al., 2003). The relationship between this large clade and the Dioscoreales and Pandanales has remained unclear in other studies of monocot phylogeny (Chase et al., 2000). Here, Dioscoreales and Pandanales form a moderately-supported clade that is the sister group to the Liliales-Asparagales-commelinid clade. Petrosaviaceae (*Petrosavia* and *Japanolirion*) is an isolated group that is the sister group of the clade composed of Asparagales, commelinids, Dioscoreales, Liliales and Pandanales. This relationship was also seen in a study by Fuse and Tamura (2001). Alismatales are well-supported as the sister group of the remaining monocots.

Campynemataceae, a small family with a range limited to Tasmania and New Caledonia, has been hypothesized to be related to various other taxa, including Iridaceae (Takhtajan, 1980) and Melanthiaceae (Dahlgren et al., 1985; Dahlgren and Lu, 1985). I find it to be the sister taxon of the rest of the Liliales, a result that is congruent with other molecular studies on Liliales phylogeny (e.g. Vinnersten and Bremer, 2002). The family contains two genera, *Campynema* and *Campynemanthe*, which both molecular (Vinnersten and Bremer, 2002) and morphological data

(Rudall et al., 2000; Dahlgren and Lu, 1985) have confirmed are each other's closest relatives.

Melanthiaceae are the sister group of the remaining Liliales, excluding Campynemataceae (Fig. 4.1). The results presented here are consistent with a circumscription of Melanthiaceae that includes *Trillium*: We sampled *Trillium* and one other exemplar taxon in Melanthiaceae (*Anticlea*), and the two are well-supported as a clade with 100% BV and 100% PP. *Trillium* and three other genera (*Daiswa, Kinugasa* and *Paris*), all of which are placed in Melanthiaceae by APG II (2003), have been treated by some authors (Tamura, 1998a) as belonging to a separate family, Trilliaceae. Several other molecular studies have presented compelling evidence that *Trillium* and relatives are nested within Melanthiaceae (Fuse and Tamura, 2000; Zomlefer et al., 2001).

Despite including Trilliaceae within the family, APG II (2003) circumscribed Melanthiaceae more narrowly than Dahlgren et al. (1985), who included several additional subfamilies, represented here by *Tofieldia*, *Petrosavia*, and *Narthecium* (Dahlgren et al., 1985). The data presented here suggest that these genera are not part of Melanthiaceae and belong in positions consistent with the APG II (2003) classifications of these taxa (*Tofieldia* in Alismatales, and *Narthecium* in Dioscoreales). Petrosaviaceae (including *Petrosavia* and *Japanolirion*) are isolated from all other orders, suggesting that the family should be placed in its own order, Petrosaviales.

A robustly supported sister-group relationship was found between the two Southern Hemisphere families, Luzuriagaceae and Alstroemeriaceae, a result also noted by Vinnersten and Bremer (2001). This clade is well-supported as the sister group of Colchicaceae. A detailed phylogeny of Colchicaceae is presented by Vinnersten and Reeves (2003). The position of the Australian species *Petermannia*

cirrosa has generally been difficult to resolve, although it has usually been regarded as related to Luzuriagaceae, or other climbing taxa in the Liliales, such as Philesiaceae, Smilacaceae or Rhipogonaceae (Conran and Clifford, 1986; Conran, 1988; Stevenson and Loconte, 1995). Cronquist (1988) placed it within Smilacaceae, and Dahlgren et al. (1985) placed it in Petermanniaceae, which they referred to Dioscoreales (Table 4.1). APG II (2003) place *Petermannia* within Colchicaceae, a decision based on molecular evidence that was the result of a misidentification of a *Tripladenia* specimen (Vinnersten and Reeves, 2003; P. Rudall, personal communication, 2003). Here, *Petermannia* is the sister group of the clade consisting of Alstroemeriaceae, Luzuriagaceae and Colchicaceae, and is not closely related to Smilacaceae, Rhipogonaceae or Philesiaceae. These results support the separation of Petermanniaceae, as a family distinct from Colchicaceae.

Liliaceae, Smilacaceae, Philesiaceae and Rhipogonaceae together form a robustly supported clade that is the sister group of the clade composed of Alstroemeriaceae, Luzuriagaceae, Colchicaceae and *Petermannia.* However, the exact position of Smilacaceae is somewhat unclear from the results obtained in this analysis. The three families Smilacaceae, Rhipogonaceae and Philesiaceae are generally regarded as being closely related. Cronquist (1988) placed members of Philesiaceae and Rhipogonaceae into Smilacaceae, but also included species now regarded as belonging to other families, including Lomandraceae and Hemerocallidaceae (see Table 4.1). Rudall et al. (2000) also suggested that the three families be put into one family, Smilacaceae, based on the results of their analyses of morphological and molecular data. Other authors (Conran 1998) have suggested that *Rhipogonum* (Rhipogonaceae), but not Philesiaceae should be placed in Smilacaceae. In the MP tree presented here (Fig. 4.1) Smilacaceae are the sister group of a clade consisting of Rhipogonaceae, Philesiaceae and Liliaceae. If this is

correct, the expanded Smilacaceae as proposed by Rudall et al. (2000) would be paraphyletic. However, the placement of Smilacaceae receives only low to moderate support here from the bootstrap analysis using the complete data set. The posterior probability value is much higher, but this should be viewed with some caution, as these values are generally higher than comparable bootstrap values (Huelsenbeck et al., 2002). The non-coding data considered alone (Fig. 4.3) give high support to branch "o," the placement of Smilacaceae as the sister group of a clade composed of Philesiaceae, Rhipogonaceae and Liliaceae. However, constraining the results of the MP tree search so that Smilacaceae, Philesiaceae and Rhipogonaceae form a monophyletic group adds only three steps to the tree length. The branches separating the three families from each other are very short (Fig. 4.2) which likely contributes to the difficulty in resolving the relationships between them.

Although Liliaceae are well-supported as a monophyletic group, the relationships among the five exemplar Liliaceae included here generally remain unresolved, with the exception of *Lilium* and *Medeola,* which are clearly closely related (Fig. 4.1). The three additional taxa (*Prosartes, Tricyrtis* and *Calochortus*) have been placed in a separate family, Calochortaceae, by some authors (e.g. Tamura 1998b). Patterson and Givnish (2002) support the division of Liliaceae *sensu* APG II into two families, Liliaceae and Calochortaceae, and their analysis of *rbc*L and *ndh*F sequences support the monophyly of the two groups. The maximum parsimony tree presented here (Fig. 4.1) does not support the monophyly of Calochortaceae, as *Calochortus* is sister to *Lilium* and *Medeola*, not *Prosartes* and *Tricyrtis*. This result is not well supported by the bootstrap analysis but is moderately supported by the Bayesian analysis. Constraining the maximum parsimony tree search so that Calochortaceae is a monophyletic group adds only

two steps to the tree length, indicating that the monophyly of the group cannot confidently be rejected.

Conclusions.

Several conclusions can be made from the analyses presented here, as relationships among major monocot clades and among families within Liliales were resolved with moderate to robust support and there was no substantial evidence of incongruence between data partitions seen here. Liliales are the sister group of a clade composed of Asparagales and the commelinids. Within Liliales, Campynemataceae and Melanthiaceae are successive sister groups of all other families in the order. *Petermannia cirrosa* is isolated from Colchicaceae and should be recognized as its own family, Petermanniaceae. Petermanniaceae is the sister taxon of a clade composed of Colchicaceae, Luzuriagaceae and Alstroemeriaceae. This clade is the sister group of one composed of Smilacaceae, Philesiaceae, Rhipogonaceae and Liliaceae.

References

Angiosperm Phylogeny Group. 1998. An ordinal classification for the families of flowering plants. *Annals of the Missouri Botanical Garden* **85**: 531–553.

Angiosperm Phylogeny Group II. 2003. An update of the angiosperm phylogeny group classification for the orders and families of flowering plants: APG II. *Botanical Journal of the Linnean Society* **141**: 399–436.

Barker, F. K., and Lutzoni, F. M. 2002. The utility of the incongruence length difference test. *Systematic Biology* **51**: 625–637.

Barry, D., and Hartigan, J. A. 1987. Asynchronous distance between homologous DNA sequences. *Biometrics* **43**: 261–276.

Bremer, B., Jansen, R. K., Oxelman, B., Backlund, M., Lantz, H., and Kim, K-J. 1999. More characters or more taxa for a robust phylogeny – case study from the coffee family. *Systematic Biology* **48**: 413–435.

Brummitt, R. K. 1992. *Vascular Plant Families and Genera.* Kew: Royal Botanic Gardens, Great Britain.

Chase, M. W., Soltis, D. E., Olmstead, R. G., Morgan, D., Les, D. H., Mishler, B., Duvall, M. R., Price, R. A., Hills, H. G., Qiu, Y-L., Kron, K. A., Rettig, J. H., Conti, E. Palmer, J. D., Manhart, J. R., Systma, K. J., Michaels, H. J., Kress, W. J., Karol, K. G., Clark, W. D., Hedren, M., Gaut, B. S., Jansen, R. K., Kim, K-J., Wimpee, C. F., Smith, J. F., Soltis, P. S., Swenson, S. M., Williams, S. E., Gadek, P. A., Quinn, C. J., Eguiarte, L., Golenburg, E., Learn, G. H., Graham, S. W., Barret, S. C. H., Dayanandan, S., and Albert, V. A. 1993. Phylogenetics of seed plants: an analysis of nucleotide sequences from the chloroplast gene *rbc*L. *Annals of the Missouri Botanical Garden* **80**: 528–580.

Chase, M. W., Duvall, M. R., Hills, H. G., Conran, J. G., Cox, A. V., Eguiarte, L. E., Hartwell, J., Fay, M. F., Caddick, L. R., Cameron, K. M., and Hoot, S. 1995a. *Molecular Phylogenetics of Lilianae.* Pp. 109–137 in: *Monocotyledons: Systematics and Evolution.* (P. J. Rudall, P. J. Cribb, D. F. Cutler, & C. J. Humphries. eds.), Royal Botanic Gardens, Kew.

Chase, M. W., Stevenson, D. W., Wilkin, P., and Rudall, P. J. 1995b. Monocot systematics: a combined analysis. Pp. 685–730 in: *Monocotyledons: Systematics and Evolution* (P. J. Rudall, P. J. Cribb, D. F. Cutler, & C. J. Humphries. eds.) Royal Botanic Gardens, Kew.

Chase, M. W., Soltis, D. E., Soltis, P. S., Rudall, P. J., Fay, M. F., Hahn, W. H., Sullivan, S., Joseph, J., Molvray, M., Kores, P. J., Givnish, T. J., Sytsma, K. J., and Pires, J. C. 2000. Higher-level systematics of the monocotyledons: an assessment of current knowledge and a new classification. Pp. 3–16 in *Monocots: Systematics and Evolution* (K. L. Wilson & D. A. Morrison, eds.). CSIRO, Melbourne.

Chiang, T.-Y., Schaal, B. A., and Peng, C. I. 1998. Universal primers for amplification and sequencing a noncoding spacer between the *atp*B and *rbc*L genes of chloroplast DNA. *Botanical bulletin of Academia Sinica.* (Taipei) **39**: 245–250.

Conran, J. G. 1988. Embryology and possible relationships of *Petermannia cirrosa* (Petermanniaceae) *Nordic Journal of Botany* **8**: 13–17.

Conran, J. G. 1998. Smilacaceae. Pp. 417–422. In: *The Families and Genera of Vascular Plants. Volume III: Flowering Plants, Monocotyledons: Lilianae (except Orchidaceae).* (K. Kubitzki, ed.). Springer-Verlag, Berlin.

Conran, J. G., and Clifford, H. T. 1985. The taxonomic affinities of the genus *Rhipogonum* J. R. et. G. Forst. *Nordic Journal of Botany* **5**: 215–219.

Conran, J. G., and Clifford, H. T. 1998. Petermanniaceae. Pp. 406–408. In: *The Families and Genera of Vascular Plants. Volume III: Flowering Plants, Monocotyledons: Lilianae (except Orchidaceae).* (K. Kubitzki, ed,). Springer-Verlag, Berlin.

Cronquist, A. 1988. *The Evolution and Classification of Flowering Plants,* 2nd edition. The New York Botanical Garden, Bronx, New York.

Dahlgren, R. M. T., Clifford, H. T., & Yeo, P. F. 1985. *The Families of Monocotyledons: Structure, Evolution and Taxonomy* Springer-Verlag, Berlin.

Dahlgren, R., and Lu, A. 1985. *Campynemanthe* (Campynemataceae): Morphology, microsporogenesis, early ovule ontogeny and relationships. *Nordic Journal of Botany* **5**: 321–330.

Doyle, J.J., and Doyle, J. L. 1987. A rapid isolation procedure for small quantities of fresh leaf tissue. *Phytochemistry Bulletin* **19:** 11–15.

Duvall, M. R., Clegg, M. T., Chase, M. W., Clark, W. D., Kress, W. J., Hills, H. G., Eguiarte, L. E., Smith, J. F., Gaut, B. S., Zimmer, E. A., and Learn, G. H. 1993a. Phylogenetic hypothesis for the monocotyledons constructed from *rbc*L sequence data. *Annals of the Missouri Botanical Garden* **80**: 607–619.

Duvall, M. R., Learn, G. H., Eguiarte, L. E., and Clegg, M. T. 1993b. Phylogenetic analysis of *rbc*L sequences identifies *Acorus calamus* as the primal extant monocotyledon. *Proceedings of the National Academy of Sciences, USA* **90**: 4641–4644.

Farris, J. S., Källersjö, M., Kluge, A. G., and Bult, C. 1994. Testing significance of incongruence. *Cladistics* **10**: 315–319.

Fay, M. F., Rudall, P. J., Sullivan, S., Stobart, K. L., de Bruijn, A. Y., Reeves, G., Qamaruz-Zaman, F., Hong, W-P., Joseph, J., Hahn, W. J., Conran, J. G., and Chase, M. W. 2000. Phylogenetic studies of Asparagales based on four plastid DNA regions. Pp. 360–371. in: *Monocots: Systematics and Evolution* (K. L. Wilson & D. A. Morrison, eds.). CSIRO, Melbourne.

Felsenstein, J. 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. *Journal of Molecular Evolution* **17**: 368–376.

Felsenstein, J. 1985. Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* **39**: 783–791.

Fuse, S. & Tamura, M. N. 2000. A phylogenetic analysis of the plastid *mat*K gene with emphasis on Melanthiaceae *sensu lato*. *Plant Biology* **2**: 415–427.

Graham, S. W., Kohn, J. R., Morton, B. R., Eckenwalder, J. E., and Barrett, S. C. H. 1998. Phylogenetic congruence and discordance among one morphological and three molecular data sets from Pontederiaceae. *Systematic Biology* **47**: 545–567.

Graham, S. W., and Olmstead, R. G. 2000. Utility of 17 chloroplast genes for inferring the phylogeny of the basal angiosperms. *American Journal of Botany* **17**: 1712–1730.

Graham, S. W., Reeves, P. A., Burns, A. C. E., Olmstead, R. G. 2000. Microstructural changes in noncoding chloroplast DNA: interpretation, evolution, and utility of indels and inversions in basal angiosperm phylogenetic inference. *International Journal of Plant Science* **161**: S83–S96.

Hasegawa, M., Kishino, H., Yano T. 1985. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *Journal of Molecular Evolution* **21**: 160–174.

Hilu, K. W., Borsch, T., Muller, K., Soltis, D. E., Soltis, P. S., Savolainen, V., Chase, M. W., Powell, M. P., Alice, L. A., Evans, R., Sauquet, H., Neinhuis, C., Slotta, T. A. B., Rohwer, J. G., Campbell, C. S., and Chatrou, L. W. 2003. Angiosperm phylogeny based on *mat*K sequence information. *American Journal of Botany* **90**: 1758–1776.

Hoot, S. B., Culham, A., and Crane, P. R. 1995. The utility of *atp*B gene sequences in resolving phylogenetic relationships: comparison with *rbc*L and 18S ribosomal DNA sequences in Lardizabalaceae. *Annals of the Missouri Botanical Garden* **82**: 194–207.

Huber, H., 1969. Die Samenmerkmale und Verwandtschaftsverhältnisse der Liliiflorae. *Mitteilungen der botanischen Staatssammlung München* **8**: 219–538.

Huelsenbeck, J. P., and Crandall, K. A. 1997. Phylogeny estimation and hypothesis testing using maximum likelihood. *Annual Review of Ecology and Systematics* **28**: 437–466.

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

Kim, K.-J., and Jansen, R. K 1995. *ndh*F sequence evolution and the major clades in the sunflower family. *Proceedings of the National Academy of Sciences, USA* **92**: 10379–10383.

Kubitzki, K. 1998a. *The Families and Genera of Vascular Plants. Volume IV, Alismatanae and Commelinanae (except Gramineae.)* Springer-Verlag, Berlin.

Kubitzki, K. 1998b. *The Families and Genera of Vascular Plants. Volume III, Flowering plants. Monocotyledons: Lilianae (except Orchidaceae).* Springer-Verlag, Berlin.

Lanave, C., Preparata, G., Saccone, C., and Serio, G. 1984. A new method for calculating evolutionary substitution rates. *Journal of Molecular Evolution* **20**: 86–93.

McPherson, M. A. 2003. Phylogenetics of Asparagales (Liliopsida) as inferred from a large plastid DNA sequence data set. M.Sc. Thesis. University of Alberta, Edmonton, Alberta, Canada.

Neyland, R., and Urbatsch, L. E. 1996a. Phylogeny of subfamily Epidendroideae (Orchidaceae) inferred from *ndh*F chloroplast gene sequences. *American Journal of Botany* **83**: 1195–1206.

Neyland, R., and Urbatsch, L. E. 1996b The *ndh*F chloroplast gene detected in all vascular plant divisions. *Planta* **20:** 273–277.

Neyland, R., and Hennigan, M. 2003. A phylogenetic analysis of large-subunit (26S)

ribosome DNA sequences suggests that the Corsiaceae are polyphyletic. *New Zealand Journal of Botany* **41**: 1–11.

Nordenstam, B. 1998. Colchicaceae. Pp. 175–185. In: *The Families and Genera of Vascular Plants. Volume III: Flowering Plants, Monocotyledons: Lilianae (except Orchidaceae)* (K. Kubitzki, ed.). Springer-Verlag, Berlin.

Olmstead, R. G., and Reeves, P. A. 1995. Evidence for the polyphyly of the Scrophulariaceae based on chloroplast *rbc*L and *ndh*F sequences. *Annals of the Missouri Botanical Garden* **82**: 149–175.

Olmstead, R. G., and Sweere, J. A. 1994. Combining data in phylogenetic systematics: an empirical approach using three molecular data sets in the Solanaceae. *Systematic Biology* **43**: 467–481.

Patterson, T. B., and Givnish, T. J. 2002. Phylogeny, concerted convergence and phylogenetic niche conservatism in the core Liliales: insights from *rbc*L and *ndh*F sequence data. *Evolution* **56**: 233–252.

Poe, S., and Swofford, D. L. 1999. Taxon sampling revisited. *Nature* **398:** 299–300.

Rai, H. S., O'Brien, H. E., Reeves, P. A., Olmstead, R. G., and Graham, S. W. 2003. Inference of higher-order relationships in the cycads from a large chloroplast data set. *Molecular Phylogenetics and Evolution* **29**: 350–359.

Rambaut, A., 1998. "*Se-Al (Sequence Alignment Editor Version 1.0)*," Computer program and documentation. Department of Zoology, University of Oxford, UK.

Rodríguez, F., Oliver, J. L., Marín, A., and Medina, J. R. 1990. The general stochastic model of nucleotide substitution. *Journal of Theoretical Biology* **142**: 485–501.

Ronquist, F., and Huelsenbeck, J. P. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**: 1572–1574.

Rudall, P. J., Stobart, K. L., Honh, W., Conran, J. G., Furness, C. A., Kite, G. C., and Chase, M. W. 2000. Consider the Lilies: systematics of Liliales. Pp. 348–356 in *Monocots: Systematics and Evolution* (K. L. Wilson & D. A. Morrison, eds.). CSIRO, Melbourne.

Sanderson, M. J., Wojciechowski, M. F., Hu, J. M., Sher Khan, T., and Brady, S. G. 2000. Error, bias and long-branch attraction in data for two chloroplast photosystem genes in seed plants. *Molecular Biology and Evolution* **17**: 782–797.

Soltis, D. E., Soltis, P. S., Chase, M. W., Mort, M. E., Albach, D. C., Zanis, M., Savolainen, V., Hahn, W., H., Hoot, S. B., Fay, M. F., Axtell, M., Swenson, S., M., Prince, L.M., Kress, W. J., Nixon, K. C., and Farris, J. S. 2000. Angiosperm phylogeny inferred from 18S rDNA, *rbc*L, and *atp*B sequences. *Botanical Journal of the Linnean Society* **133**: 381–461.

Stevenson, D. W., and Loconte, H. 1995. Cladistic analysis of monocot families. Pp. 543–578 In: *Monocotyledons: Systematics and Evolution* (P. J. Rudall, P. J. Cribb, D. F. Cutler, & C. J. Humphries. eds.) Royal Botanic Gardens, Kew.

Swofford, D. L. 2002. "*PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4.0b10*," Computer program and documentation. Sinauer Associates, Sunderland, MA.

Taberlet, P., Gielly, L., Pautou, G., and Bouvet, J. 1991. Universal primers for amplification of three non-coding regions of chloroplast DNA. *Plant Molecular Biology* **17**: 1105–1109.

Takhtajan, A. 1997. *Diversity and Classification of Flowering Plants*. Columbia University Press, New York.

Tamura, M. N. 1998a. Trilliaceae. Pp. 445–452. In: *The Families and Genera of Vascular Plants. Volume III: Flowering Plants, Monocotyledons: Lilianae (except Orchidaceae)*. (K. Kubitzki, ed,). Springer-Verlag, Berlin.

Tamura, M. N. 1998b. Calochortaceae. Pp. 164–172. In: *The Families and Genera of Vascular Plants. Volume III: Flowering Plants, Monocotyledons: Lilianae (except Orchidaceae)* (K. Kubitzki, ed,). Springer-Verlag, Berlin.

Tavaré, S., 1986. Some probabilistic and statistical problems on the analysis of DNA sequences. *Lectures in Mathematics and Life Sciences* **17**: 57–86.

Vinnersten, A., and Bremer, K. 2001. Age and biogeography of major clades in Liliales. *American Journal of Botany* **88**: 1695–1703.

Vinnersten, A., and Reeves, G. 2003. Phylogenetic relationships within Colchicaceae. *American Journal of Botany* **90**: 1455–1462.

Zurawski, G., Clegg, M. T., and Brown, A. H. D. 1984. The nature of nucleotide sequence divergence between barley and maize chloroplast DNA. *Genetics* **106**: 735–749.

Zomlefer, W. B., Williams, N. H., Whitten, W. M., and Judd, W. S. 2001. Generic circumscription and relationships in the tribe Melanthieae (Liliales: Melanthiaceae), with emphasis on *Zigadenus*: evidence from ITS and *trn*L-F sequence data. *American Journal of Botany* **88**: 1657–1669.

Table 4.1. Classification of the monocot genera examined here, according to three widely used taxonomic schemes [1]. All "bracketed," or optionally recognized families in the APG system are recognized at the family level here. Genera of Liliales *sensu* APG (2003) are underlined; the number of orders and families are only noted (in parentheses) if some taxa were not sampled at that rank.

Cronquist (1988) [2] [5 subclasses]	Dahlgren, Clifford and Yeo (1985) [3] [10 superorders]	APG II (2003) [4] [10 orders; 2 unplaced families]
ALISMATIDAE (4 orders)	ALISMATIFLORAE	Family unplaced in monocots
Alismatales (3 families)	Alismatales (5 families)	Petrosaviaceae
Alismataceae	Alismataceae	*Japonolirion*
Sagittaria	*Sagittaria*	*Petrosavia*
Butomaceae	Butomaceae	
Butomus	*Butomus*	Acorales
Najadales (10 families)	Najadales (8 families)	Acoraceae
Scheuchzeriaceae	Scheuchzeriaceae	*Acorus*
Scheuchzeria	*Scheuchzeria*	
Triuridales (2 families)		Alismatales (14 families)
Petrosaviaceae	ARECIFLORAE	Araceae
Petrosavia	Arecales	*Spathiphyllum*
	Arecaceae	Alismataceae
ARECIDAE	*Roystonea*	*Sagittaria*
Arales (3 families)		Butomaceae
Acoraceae	ARIFLORAE	*Butomus*
Acorus	Arales (2 families)	Scheuchzeriaceae
Araceae	Araceae	*Scheuchzeria*
Spathiphyllum	*Acorus*	Tofieldiaceae
Arecales	*Spathiphyllum*	*Tofieldia*
Arecaceae (Palmae)		
Roystonea	BROMELIIFLORAE	Asparagales (29 families)
Cyclanthales	Bromeliales	Agapanthaceae
Cyclanthaceae	Bromeliaceae	*Agapanthus*
Carludovica	*Ananas*	Agavaceae
Pandanales	Haemodorales	*Yucca*
Pandanaceae	Haemodoraceae	Alliaceae
Pandanus	*Xiphidium*	*Allium*
	Philydrales	Amaryllidaceae
COMMELINIDAE (7 orders)	Philydraceae	*Narcissus*
Commelinales (4 families)	*Philydrum*	Anthericaceae
Mayacaceae	Pontederiales	*Chlorophytum*
Mayaca	Pontederiaceae	Asparagaceae
Xyridaceae	*Hydrothrix*	*Asparagus*
Xyris	Typhales	Asphodelaceae
Cyperales	Sparganiaceae	*Asphodelus*
Cyperaceae	*Sparganium*	Asteliaceae
Cyperus	Typhaceae	*Astelia*
Poaceae (Graminae)	*Typha*	Blandfordiaceae
Oryza	Velloziales	*Blandfordia*
Triticum	Velloziaceae	Boryaceae
Zea	*Talbotia*	*Alania*
Restionales (4 families)		Doryanthaceae
Flagellariaceae	COMMELINIFLORAE (4 orders)	*Doryanthes*
Flagellaria	Commelinales (5 families)	Hemerocallidaceae
Restionaceae	Mayacaceae	*Hemerocallis*
Ecdeiocolea	*Mayaca*	*Phormium*
Elegia	Xyridaceae	Hyacinthaceae
Typhales	*Xyris*	*Muscari*
Sparganiaceae		
Sparganium		

Cronquist (1988)	Dalgren et al. (1985)	APG II (2003)
Typhales (contd.)	COMMELINIFLORAE (contd.)	Asparagales (contd.)
Typhaceae		Hypoxidaceae
Typha	Cyperales (3 families)	*Curculigo*
	Cyperaceae	Iridaceae
LILIIDAE	*Cyperus*	*Iris*
Liliales (15 families)	Poales (7 families)	*Sisyrinchium*
Agavaceae	Ecdeiocoleaceae	Ixioliriaceae
Doryanthes	*Ecdeiocolea*	*Ixiolirion*
Phormium	Flagellariaceae	Lanariaceae
Xeronema	*Flagellaria*	*Lanaria*
Yucca	Poaceae	Laxmanniaceae
Cyanastraceae	*Oryza*	*Lomandra*
Cyanastrum	*Triticum*	Orchidaceae
Dioscoreaceae	*Zea*	*Amerorchis*
Dioscorea	Restionaceae	*Coelogyne*
Haemodoraceae	*Elegia*	*Cypripedium*
Xiphidium		Ruscaceae
Iridaceae	CYCLANTHIFLORAE	*Maianthemum*
Iris	Cyclanthales	Tecophilaeaceae
Sisyrinchium	Cyclanthaceae	*Cynastrum*
Liliaceae	*Carludovica*	Themidaceae
Agapanthus		*Muilla*
Alania	LILIIFLORAE	Xanthorrhoeaceae
Allium	Asparagales (30 families)	*Xanthorrhoea*
Alstroemeria	Agavaceae	Xeronemataceae
Anticlea	*Yucca*	*Xeronema*
Asparagus	Alliaceae	
Asphodelus	*Allium*	Dioscoreales
Astelia	*Agapanthus*	Burmanniaceae
Blandfordia	*Muilla*	*Burmannia*
Calochortus	Amaryllidaceae	Dioscoreaceae
Campynema	*Narcissus*	*Dioscorea*
Chlorophytum	Anthericaceae	Nartheciaceae
Curculigo	*Alania*	*Narthecium*
Hemerocallis	*Chlorophytum*	
Ixiolirion	Asparagaceae	Liliales (10 families)
Lilium	*Asparagus*	Alstroemeriaceae
Maianthemum	Asphodelaceae	*Alstroemeria*
Medeola	*Asphodelus*	Campynemataceae
Muilla	Asteliaceae	*Campynema*
Muscari	*Astelia*	Colchicaceae
Narcissus	Blandfordiaceae	*Petermannia*
Narthecium	*Blandfordia*	*Tripladenia*
Prosartes	Convallariacae	*Wurmbea*
Tricyrtis	*Maianthemum*	Liliaceae
Trillium	Cyanastraceae	*Calochortus*
Tripladenia	*Cynastrum*	*Lilium*
Tofieldia	Dasypogonaceae	*Medeola*
Wurmbea	*Dasypogon*	*Prosartes*
Philydraceae	*Kingia*	*Tricyrtis*
Philydrum	*Lomandra*	Luzuriagaceae
Pontederiaceae	Doryanthaceae	*Luzuriaga*
Hydrothrix	*Doryanthes*	Melanthiaceae
	Hemerocallidaceae	*Anticlea*
	Hemerocallis	*Trillium*

Cronquist (1988)	Dahlgren et al. (1985)	APG II (2003)
Liliales (contd.)	LILIIFLORAE/ Asparagales (contd.)	Liliales (contd.)
Smilacaceae	Hyacinthaceae	Philesiaceae
Luzuriaga	*Muscari*	*Philesia*
Petermannia	Hypoxidaceae	Rhipogonaceae
Philesia	*Curculigo*	*Rhipogonum*
Rhipogonum	Ixioliriaceae	Smilacaceae
Smilax	*Ixiolirion*	*Smilax*
Stemonaceae	Luzuriagaceae	
Stemona	*Luzuriaga*	Pandanales (5 families)
Velloziaceae	Philesiaceae	Cyclanthaceae
Talbotia	*Philesia*	*Carludovica*
Xanthorrhoeaceae	Phormiaceae	Pandanaceae
Dasypogon	*Phormium*	*Pandanus*
Kingia	*Xeronema*	Stemonaceae
Lomandra	Tecophilaeaceae	*Stemona*
Xanthorrhoea	*Lanaria*	Velloziaceae
Orchidales (4 families)	Xanthorrhoeaceae	*Talbotia*
Burmanniaceae	*Xanthorrhoea*	
Burmannia	Burmanniales (3 families)	COMMELINIDS
Orchidaceae	Burmanniaceae	Unplaced at ordinal level
Amerorchis	*Burmannia*	Dasypogonaceae
Coelogyne	Dioscoreales (7 families)	*Dasypogon*
Cypripedium	Dioscoreaceae	*Kingia*
	Dioscorea	
ZINGIBERIDAE	Petermanniaceae	Arecales
Bromeliales	*Petermannia*	Arecaceae
Bromeliaceae	Smilacaceae	*Roystonea*
Ananas	*Rhipogonum*	
	Smilax	Commelinales (5 families)
Zingiberales (8 families)	Stemonaceae	Haemodoraceae
Musaceae	*Stemona*	*Xiphidium*
Ensete	Trilliaceae	Philydraceae
Strelitziaceae	*Trillium*	*Philydrum*
Strelitzia	Liliales (10 families, incl. orchids)	Pontederiaceae
	Alstroemeriaceae	*Hydrothrix*
	Alstroemeria	
Placement in monocots not noted:	Calochortaceae	Poales (18 families)
Lanaria	*Calochortus*	Bromeliaceae
Japonolirion	Colchicaceae	*Ananas*
	Wurmbea	Cyperaceae
	Iridaceae	*Cyperus*
	Iris	Ecdeiocoleaceae
	Sisyrinchium	*Ecdeiocolea*
	Liliaceae	Flagellaricaceae
	Lilium	*Flagellaria*
	Medeola	Mayacaceae
	Uvulariaceae	*Mayaca*
	Prosartes	Poaceae
	Tricyrtis	*Oryza*
	Tripladenia	*Triticum*
		Zea
	Orchids (rankless in Liliales; 3 families)	Restionaceae
	Cypripediaceae	*Elegia*
	Cypripedium	Sparganiaceae
	Orchidaceae	*Sparganium*
	Amerorchis	Typhaceae
	Coleogyne	*Typha*

Dahlgren et al. (1985)	APG II (2003)
LILIIFLORAE (contd.)	Poales (contd.)
Melanthiales	Xyridaceae
Campynemataceae	*Xyris*
Campynema	
Melanthiaceae	Zingiberales (8 families)
Anticlea	Musaceae
Narthecium	*Ensete*
Petrosavia	Strelitziaceae
Tofieldia	*Strelitzia*
PANDANIFLORAE	
Pandanales	
Pandanaceae	
Pandanus	
ZINGIBERIFLORAE	
Zingiberales (8 families)	
Musaceae	
Ensete	
Strelitziaceae	
Strelitzia	
Placement in monocots not noted:	
Japonolirion	

[1] Taxon authorities at ordinal level and below are listed in Brummitt (1992) and APG (2003); taxon counts in APG II include all "optional" (bracketed) families.

[2] Generic placement follows Cronquist (1988) and Brummitt (1992). Three families of Liliales *sensu* Cronquist (1988) were not sampled here (Aloeaceae, Hanguanaceae and Taccaceae).

[3] Two families of Liliales *sensu* Dahlgren et al. (1985) were not sampled here (Geosiridaceae, Apostasiaceae).

[4] Generic placement follows Kubitzki (1998a,b), except for Asparagales. For most Liliales we follow Rudall et al. (2000), apart from Rhipogonaceae, Philesiaceae and Smilacaceae, where we follow Conran and Clifford (1985). Kubitzki (1998a) [the former two families are recognized as part of Smilacaceae by Rudall et al. (2000)]. Only one family Liliales *sensu* APG II (2003) was not sampled here (Corsiaceae).

Table 4.1. GenBank accession numbers and vouchers for exemplar monocot taxa. Underlined sequences are new to this study and were generated by various authors in the Graham laboratory. I produced all new Liliales sequences, *Agapanthus*, *Doryanthes* and *Pandanus*, except where noted). Sequences in plain type were generated by Graham et al. (2000) and McPherson et al. (in press); see these for details on individual sequences. Sequences in bold were generated by other lab groups using different source material. The three grasses (Poaceae) included here (*Oryza sativa* L., *Triticum aestivum* L. and *Zea mays* L.) are from whole plastid genomes deposited on GenBank (accessions NC001320, NC002762 and NC001666.2, respectively).

	Gene or region									
Taxon [Voucher (herbarium)]	*atp*B [1]	*ndh*F [2]	*psb*B, T, N, & *psb*H [3]	*psb*D & *psb*C [4]	*psb*E, F, L & *psb*J [5]	*rbc*L	*rpl2* [6]	3'-*rps*12, *rps*7, *ndh*B, *trn*L(CAA) [7]	*atp*B- *rbc*L spacer region	*trnL*(UAA)-*trnF*(GAA)
ACORALES										
Acorus calamus L. [see Graham and Olmstead, 2000a]	**AJ235381**	AY007647	AF123843	AF123813	AF123828	**D28865**	AF123785	AF123771	--	--
ALISMATALES										
Sagittaria latifolia Willd. [SCH Barrett s.n. (TRT)]	AF239788	AY007657	AY007469	AF239789	AY007484	**L08767**	AY007497	AF238074	--	--
Scheuchzeria palustris L. [M Waterway & SW Graham 97-60 (ALTA)]	AY147594	AF547007	AY147500	AY147639	AY147547	**U03728**	AY147686	AY147451	--	--
Spathiphyllum wallisii Hort. [MW Chase 210 (NCU)]	AJ235606.2	AY007658	AY007471	AF239794	AY007487	**AJ235807**	AY007500	AF238077	--	--
ASPARAGALES										
Agapanthus africanus (L) Hoffmanns [020415-1 (ALTA)]	AY465542	AY465647	AY465568	AY465672	AY465595	AY465699	AY465724	AY46562	AY699127	AY699224
Alania endlicheri Kunth [DH Vitt 27706 (ALTA)]	AY147612	AY147773	AY147519	AY147658	AY147566	**Y14982**	AY147705	AY147471	AY147737	**Fay et al. (2000)** [8]

Taxon [Voucher (herbarium)]	Gene or region *atp*B [1]	*ndh*F [2]	*psb*B, T, N, & *psb*H [3]	*psb*D & *psb*C [4]	*psb*E, F, L & *psb*J [5]	*rbc*L	*rpl*2 [6]	3'-*rps*12, *rps*7, *ndh*B, *trn*L(CAA) [7]	*atp*B- *rbc*L spacer region	*trnL*(UAA)-*trnF*(GAA)
ASPARAGALES (contd.)										
Allium textile A Nelson & J F MacBride [MA McPherson 990704-79 (ALTA)] and	AY147628	AF547000	AY147536	AY147675	AY147583	AY149372	AY147721	AY147489	AY147753	--
A. altaicum Pall.	--	--	--	--	--	--	--	--	--	**Fay et al. (2000)** [8]
Amerorchis rotundifolia (Banks) Hultén [MA McPherson 010610-1 (ALTA)] and	AY147623	AY147783	AY147531	AY147670	AY147578	AY149368	AY147716	AY147484	AY147748	--
Orchis militaris L.	--	--	--	--	--	--	--	--	--	**AY014586**
Aphyllanthes officinalis L. [MA McPherson 010819-2 (ALTA)]	AY147629	AY147787	AY147537	AY147676	AY147584	AY149373	AY147722	AY147490	AY147754	**Fay et al. (2000)** [8]
Asphodelus albus Willd. [L Harder 1-000430 (ALTA)] and	AY147613	AY147774	AY147520	AY147659	AY147567	AY149360	AY147706	AY147472	AY147738	--
A. aestivus Brot.	.--	--	--	--	--	--	--	--	--	**AAE290257/ AAE290291**
Astelia alpina R. Br. [MW Chase 1103 (NCU)]	AY147614	AY147775	AY147521	AY147660	AY147568	**Z77261**	AY147707	AY147473	AY147739	**Fay et al. (2000)** [8]
Blandfordia punicea (Labill.) Sweet [MW Chase 519 (NCU)] and	AY147615	AY147776	AY147522	AY147661	AY147569	**Z73694**	AY147708	AY147474	AY147740	--
B. nobilis Smith	--	--	--	--	--	--	--	--	--	**AJ232441/ AJ232564**

Taxon [Voucher (herbarium)]	Gene or region *atp*B [1]	*ndh*F [2]	*psb*B, T, N, & *psb*H [3]	*psb*D & *psb*C [4]	*psb*E, F, L & *psb*J [5]	*rbc*L	*rpl*2 [6]	3'-*rps*12, *rps*7, *ndh*B, *trn*L(CAA) [7]	*atp*B- *rbc*L spacer region	*trnL*(UAA)- *trnF*(GAA)
ASPARAGALES (contd.)										
Chlorophytum comosum (Thunb.) Jacques [MA McPherson 000321-1 (ALTA)] and	AY147631	AY147789	AY147539	AY147678	AY147586	AY149375	AY147724	AY147492	--	--
Anthericium ramosum L.	--	--	--	--	--	--	--	--	--	**AJ232445/AJ232568**
Coelogyne cristata Lindl. [MA McPherson 010921-1 (ALTA) and	AY147616	AY147777	AY147523	AY147662	AY147570	AY149361	AY147709	AY147475	--	--
C. macdonaldii F Muell.	--	--	--	--	--	--	--	--	--	**AF463396/ AF463381**
Curculigo capitata (Lour.) Kuntze [MW Chase 205 (NCU)] and	AY147617	AY147778	AY147524	AY147663	AY147571	AY149362	AY147710	AY147476	AY147742	--
Hypoxis villosa L. f.	--	--	--	--	--	--	--	--	--	**X74579**
Cyanastrum cordifolium Oliver [Graham & Barrett 2 (TRT)] and	**AF168902**	U79228	AY147525	AY147664	AY147572	U41572	AY147711	AY147477	AY147743	--
Kabuyea hostifolia (Engl.) Brummit	--	--	--	--	--	--	--	--	--	**AJ290312/ AJ290278**
Cypripedium passerinum Richardson [MA McPherson 010722-6 (ALTA)]	AY147618	AY147779	AY147526	AY147665	AY147573	AY149363	AY147712	AY147478/ AY147479	--	--
Doryanthes palmeri W. Hill ex Benth [Chase 2837 (K)]	AY465543	AY465648	AY465569	AY465673	AY465596	AY465700	AY465725	AY465624	AY699128	AY699160/AY699161
Hemerocallis [MW Chase 3833 (K)] and	AY147619	AY147780	AY147527	AY147666	AY147574	AY149364	AY147713	AY147480	AY147744	--
Dianella ensifolia (L.) DC	--	--	--	--	--	--	--	--	--	**AB095605/ AJ290307**

Taxon [Voucher (herbarium)]	Gene or region *atp*B [1]	*ndh*F [2]	*psb*B, T, N, & *psb*H [3]	*psb*D & *psb*C [4]	*psb*E, F, L & *psb*J [5]	*rbc*L	*rpl*2 [6]	3'-*rps*12, *rps*7, *ndh*B, *trn*L(CAA) [7]	*atp*B- *rbc*L spacer region	*trnL*(UAA)-*trnF*(GAA)
ASPARAGALES (contd.)										
Iris missouriensis Nutt. [MA McPherson 000707-5a-7 (ALTA)] and	AY147620	AF547003	AY147528	AY147667	AY147575	AY149365	AY147714	AY147481	AY147745	--
I. unguicularis Poir.	--	--	--	--	--	--	--	--	--	**AJ409609**
Ixiolirion tataricum (Pallas) Herbert [MW Chase 489 (K)]	AY147621	AY147781	AY147529	AY147668	AY147576	AY149366	--	AY147482	AY147746	**AJ290280/AJ290314**
Lanaria lanata (L.) Druce [MW Chase 458 (NCU)]	AY147622	AY147782	AY147530	AY147669	AY147577	AY149367	AY147715	AY147483	AY147747	--
Lomandra longifolia Labill. [DH Vitt 27411 (ALTA)] and	AY147632	AF547004	AY147540	AY147679	AY147587	**L05039.2**	AY147725	AY147493	AY147757	--
Thysanotus spiniger Brittan	--	--	--	--	--	--	--	--	--	**Fay et al. (2000)** [8]
Maianthemum racemosa (L.) Link [MA McPherson 990704-97 (ALTA)] and	AY147633	AF547005	AY147541	AY147680	AY147588	AY149376	AY147726	AY147494	AY147758	--
M. bifolium (L.) FW Schmidt	--	--	--	--	--	--	--	--	--	**AJ441175**
Muilla maritima S Watson [JC Pires 98-028 (WIS)]	AY147634	AY147790	AY147542	AY147681	AY147589	AY149377	AY147727	AY147495	AY147759	**Fay et al (2003)** [8]
Muscari comosum (L.) Miller [L Harder 000419-1 (ALTA)]	AY147635	AF547006	AY147543	AY147682	AY147590	AY149378	AY147728	AY147496	AY147760	**AJ232546/ AJ232669**

Taxon [Voucher (herbarium)]	Gene or region *atp*B [1]	*ndh*F [2]	*psb*B, T, N, & *psb*H [3]	*psb*D & *psb*C [4]	*psb*E, F, L & *psb*J [5]	*rbc*L	*rpl*2 [6]	3'-*rps*12, *rps*7, *ndh*B, *trn*L(CAA) [7]	*atp*B- *rbc*L spacer region	*trnL*(UAA)- *trnF*(GAA)
ASPARAGALES (contd.)										
Narcissus elegans (Haw.) Spach [SCH Barrett 1434 (TRT)]	AY147636	U79216.2	AY147544	AY147683	AY147591	AY149379	AY147729	AY147497	AY147761	**AY357142**
Phormium tenax JR Forst. & G Forst. [MA McPherson 000612-3 (ALTA)]	AY147624	AY147784	AY147532	AY147671	AY147579	**Z69232**	AY147717	AY147485	AY147749	**TBA**
Sisyrinchium montanum Greene [MA McPherson 990704-71 (ALTA)] and	AY147625	AF547008	AY147533	AY147672	AY147580	AY149369	AY147718	AY147486	AY147750	--
S. micranthum Cav.	--	--	--	--	--	--	--	--	--	**AJ290307**
Xanthorrhoea resinosa Pers. [MW Chase 192 (NCU)]	AY147626	AY147785	AY147534	AY147673	AY147581	AY149370	AY147719	AY147487	AY147751	**Fay et al (2000)** [8]
Xeronema callistemon W. R. B. Oliv. [MW Chase 653 (K)]	AY147627	AY147786	AY147535	AY147674	AY147582	AY149371	AY147720	AY147488	AY147752	**Fay et al. (2000)** [8]
Yucca glauca Nutt. [Voucherless field collection (*SWG 00121* DNA)] and	AY147637	AF547014	AY147545	AY147684	AY147592	AY149380	AY147730	AY147498	AY147762	--
Agave celsii Hoot	--	--	--	--	--	--	--	--	--	**AF508509**
COMMELINALES										
Ensete ventricosum (Welw.) Cheesman [Kress 96-5372 (US)]	**AF168910**	AY147769	AY147510	AY147649	AY147557	AY149354	AY147696	AY147461	AY147733	--

Taxon [Voucher (herbarium)]	Gene or region *atp*B [1]	*ndh*F [2]	*psb*B, T, N, & *psb*H [3]	*psb*D & *psb*C [4]	*psb*E, F, L & *psb*J [5]	*rbc*L	*rpl*2 [6]	3'-*rps*12, *rps*7, *ndh*B, *trn*L(CAA) [7]	*atp*B- *rbc*L spacer region	*trnL*(UAA)-*trnF*(GAA)
COMMELINALES (contd.)										
Philydrum lanuginosum Banks and Sol. ex Gaertn. [Graham & Barrett 1 (TRT)]	AY147607	U41622	AY147514	AY147653	AY147561	U41596	AY147700	AY147465	AY147734	--
Roystonea princeps (Becc.) Burret [E Santiago #J-4 (UPR)]	AY147608	AY147772	AY147515	AY147654	AY147562	AY149357	AY147701	AY147466	AY147735	--
DIOSCOREALES										
Burmannia capitata Mart. [R Neyland 958 (MCN)]	AY147596	--	AY147502	AY147641	AY147549	AY149347	AY147688	AY147453	--	--
Dioscorea bulbifera L. [see Graham and Olmstead, 2000a]	AF187059	AY007652	AF123849	AF123819	AF123834	**D28327**	AF123791	AF123777	--	--
Narthecium ossifragum Hudson [Rothwell & Stockey 59 (ALTA]	AY147597	AY147763	AY147503	AY147642	AY147550	**AJ286560**	AY147689	AY147454	--	--
LILIALES										
Alstroemeria aurea Graham [MJC 157 Silwood Park (BERKS)]	AY465546	AY465651	AY465572	AY465676	AY465599	AY465703	AY465728	AY465627	AY699131	AY699225
Anticlea elegans (Pursh) Rydberg [MA McPherson 990704-68 (ALTA)]	AY147600	AY147765	AY147506	AY147645	AY147553	AY149351	AY147692	AY147457	AY699130	AY699168/AY699169

Taxon [Voucher (herbarium)]	Gene or region *atp*B [1]	*ndh*F [2]	*psb*B, T, N, & *psb*H [3]	*psb*D & *psb*C [4]	*psb*E, F, L & *psb*J [5]	*rbc*L	*rpl2* [6]	3'-*rps*12, *rps*7, *ndh*B, *trn*L(CAA) [7]	*atp*B- *rbc*L spacer region	*trnL*(UAA)- *trnF*(GAA)
LILIALES (contd.)										
Calochortus apiculatus Baker [JZ26 (ALTA)]	AY465547	AY465652	AY465573	AY465677	AY465600	AY465704	AY465729	AY465628	AY699132	**Fay et al. (2000)**
Campynema lineare Labill [Walsh 3488 (MEL)]	**AJ417573**	**AF276013**	AY465574	AY465678	AY465601	**Z77264**	AY465730	AY465629	AY3699133	**Fay et al. (2000)**
Lilium superbum L. [MW Chase 112 (NCU)]	AY116649	AY007655	AY007465	AF239783	AY007480	**L12682**	AY007493	AF238070	AY699129	**Fay et al. (2000)**
Luzuriaga radicans Ruiz & Pav. [Chase 499 (K)]	AY465548	AY465653	AY465575/ AY465742	AY465679	AY465602	AY465705	AY465731	AY465630	AY699134	AY699162/AY699163
Medeola virginiana L. [TL Eades May 18 2001 (ALTA)]	AY465549	AY465654	AY465576	AY465680	AY465603	AY465706	AY465732	AY465631	AY699135	AY699226
Petermannia cirrosa F Muell. [S Frederiksen et al. s. n. 4 Oct. 1998 (C)	AY465558	AY465662	AY465585/ AY465743	AY465689	AY465612	AY465714	AY465741	AY465640	AY699144	--
Philesia buxifolia Lam. [Chase 545 (K)]	AY465551	AY465656	AY465578/ AY465744	AY465682	AY465605	AY465707	AY465734	AY465633	AY699137	AY699227
Prosartes trachycarpa S. Watson [E MacDonald 180 (ALTA)]	AY465552	AY465657	AY465579	AY465683	AY465606	AY465708	AY465735	AY465634	AY699138	AY699228

Taxon [Voucher (herbarium)]	Gene or region *atp*B [1]	*ndh*F [2]	*psb*B, T, N, & *psb*H [3]	*psb*D & *psb*C [4]	*psb*E, F, L & *psb*J [5]	*rbc*L	*rpl*2 [6]	3'-*rps*12, *rps*7, *ndh*B, *trn*L(CAA) [7]	*atp*B-*rbc*L spacer region	*trnL*(UAA)-*trnF*(GAA)
LILIALES (contd.)										
Rhipogonum elseyanum F. Muell [Chase 187 (NCU)]	AY465553	AY465658	AY465580/ AY465745	AY465684	AY465607	AY465709	AY465736	AY465635	AY699139	AY699164/AY699165
Smilax rotundifolia L. [Uhl 92-07 (BH)]	AY465554	AY465659	AY465581/ AY465746	AY465685	AY465608	AY465710	AY465737	AY465636	AY699140	AY699170/AY699171
Tricyrtis sp. Wall. [M. Waterway, ALTA]	AY465555	AY465660	AY465582	AY465686	AY465609	AY465711	AY465738	AY465637	AY699141	AY699229
Trillium grandiflorum (Michx) Salisb [TL Eades June 1 2001 (ALTA)]	AY465556	AY465661	AY465583	AY465687	AY465610	AY465712	AY465739	AY465638	AY699142	AY699166/AY699167
Tripladenia cunninghamii D Don [RC Coveny 16692 & AJ Whalen, (K)]	AY465550	AY465655	AY465577	AY465681	AY465604	**Z77267**	AY465733	AY465632	AY699136	--
Wurmbea pygmaea (Endl.) Benth. [AL Case 77 (PERTH)]	AY465557	AF547012	AY465585	AY465688	AY465611	AY465713	AY465740	AY465639	AY699143	A Case, unpublished data.

Taxon [Voucher (herbarium)]	Gene or region *atp*B [1]	*ndh*F [2]	*psb*B, T, N, & *psb*H [3]	*psb*D & *psb*C [4]	*psb*E, F, L & *psb*J [5]	*rbc*L	*rpl*2 [6]	3'-*rps*12, *rps*7, *ndh*B, *trn*L(CAA) [7]	*atp*B- *rbc*L spacer region	*trnL*(UAA)-*trnF*(GAA)
PANDANALES										
Carludovica drudei Mast. [73:574 (BH)]	AY465545	AY465650	AY465571	AY465675	AY465598	AY465702	AY465727	AY465626	--	--
Pandanus copelandii Merr	AY465544	AY465649	AY465570	AY465674	AY465597	AY465701	AY465726	AY465625	--	--
Stemona tuberosa Lour. [Rothwell & Stockey 46 (ALTA)]	AY147599	AF547009	AY147505	AY147644	AY147552	AY149350	AY147691	AY147456	--	--
Talbotia elegans Balf. [Rothwell & Stockey 48 (ALTA)]	AY147609	AF547011	AY147516	AY147655	AY147563	AY149358	AY147702	AY147467	--	--
POALES										
Ananas comosum (L.) Merr. [HS Rai 1003 (ALTA)]	AY147601	AY147766	AY147507	AY147646	AY147554	**L19977**	AY147693	AY147458	AY147731	--
Cyperus papyrus L. [Alan Yen 174]	AY465534	AY465642	AY465560	AY465664	AY465587	**Y12966**	AY465717	AY465615	--	--
Ecdeiocolea monostachya F Muell. [Hopper 8531 (KPBG)]	AY465535	AY438617	AY465561	AY465665	AY465588	AY465692	AY465718	AY465616	--	--

Taxon [Voucher (herbarium)]	Gene or region *atp*B [1]	*ndh*F [2]	*psb*B, T, N, & *psb*H [3]	*psb*D & *psb*C [4]	*psb*E, F, L & *psb*J [5]	*rbc*L	*rpl*2 [6]	3'-*rps*12, *rps*7, *ndh*B, *trn*L(CAA) [7]	*atp*B-*rbc*L spacer region	*trnL*(UAA)-*trnF*(GAA)
POALES (contd.)										
Elegia fenestra Pillans [New York Botanical Garden 1697/95 (NY)]	AY465536	AY547016	AY465562	AY465666	AY465589	AY465693	AY465719	AY465617	--	--
Flagellaria indica L. [Bailey Hortorium 77:394 (BH)]	AY465537	AY465643	AY465563	AY465667	AY465590	AY465694	AY465720	AY465618	--	--
Sparganium eurycarpum Engelm [Hansen s.n., June 1993 (BH)]	AY465539	AY465645	AY465565	AY465669	AY465592	AY465696	AY465721	AY465620	--	--
Typha latifolia L. [MA McPherson 010819-3 (ALTA)] and	AY147610	--	--	--	--	**M91634**	AY147703	AY147469	AY147736	--
T. angustifolia L. [SW Graham 1040 (ALTA)]	--	U79230	AY147517	AY147656	AY147564	--	--	AY147468	--	--
Xyris jupicai Rich [D Goldman 1766 (BH)]	AY465541	AF547017	AY465567	AY465671	AY465594	AY465698	AY465723	AY465622	--	--
ZINGIBERALES										
Strelitzia reginae Aiton [H. O'Brien, ALTA]	AY465540	AY465646	AY465566	AY465670	AY465593	AY465697	AY465722	AY465621	--	--

Taxon [Voucher (herbarium)]	Gene or region *atp*B [1]	*ndh*F [2]	*psb*B, T, N, & *psb*H [3]	*psb*D & *psb*C [4]	*psb*E, F, L & *psb*J [5]	*rbc*L	*rpl*2 [6]	3'-*rps*12, *rps*7, *ndh*B, *trn*L(CAA) [7]	*atp*B- *rbc*L spacer region	*trnL*(UAA)-*trnF*(GAA)
UNCERTAIN ORDINAL PLACEMENT										
Dasypogonaceae										
Dasypogon hookeri J. R. Drumm. [MW Chase 430 (NCU) & Conran et al. 917 (PERTH)]	AY147603	AY147768	AY147509	AY147648	AY147556	AY149353	AY147695	AY147460	AY147732	--
Kingia australis R. Br. [Conran et al. 922 (PERTH)]	AY465538	AY465644	AY465564	AY465668	AY465591	AY465695	--	AY465619	--	--
Petrosaviaceae										
Petrosavia Beccari sp. [MW Chase 1933 (K)] and	--	--	--	--	--	AY465690	AY465715	AY465613	--	--
Petrosavia stellaris Beccari	AF209649	--	--	--	--	--	--	--	--	--

[1] Taxa with portions of missing *atp*B: *Kinga* (5' end is 185 bp short), *Tripladenia* (5' end is 160 bp short).

[2] Partial sequences of the *ndh*F gene were obtained several taxa. These are:*Agapanthus, Alstroemeria, Calochortus, Doryanthes, Medeola, Luzuriaga, Ripogonum, Pandanuus, Philesia, Smilax, Tricyrtis*, and *Trillium* (~1500 bp of the 5' end of the gene were sequenced) and *Lilium, Prosartes, Strelitzia* and *Tripladenia* (~1300 bp of the 5' region of the gene were sequenced). Additionally, *Cyperus ndh*F is missing 202 bp from the 5' end.

[3] Taxa with missing sequence for the *psb*B-*psb*T-*psb*N-*psb*H region include: *Calochortus* (104 bp missing from *psb*B 5' end); *Luzuriaga* (approximately 50 bp missing from the *psb*T-*psb*N intergenic spacer); *Philesia* (246 bp missing from 3' *psb*B, all of *psb*T is missing, the *psb*B-*psb*T intergenic spacer is missing, and a large portion of the *psbT-psb*N intergenic spacer is missing), *Rhipogonum* (16 bp missing from 3' *psb*T, 6 bp missing from 3' *psb*N, and no *psb*T-*psb*N intergenic spacer); *Smilax* (a total of ~110 bp are missing, including *psb*T, and portions of the *psb*B-*psb*T and *psb*T-*psb*N intergenic spacers); and *Petermannia* (only 1101 bp of 3' *psb*B was sequenced, *psb*T and portions of the *psb*B-*psb*T and *psb*T-*psb*T intergenic spacers are also missing).

[4] Taxa with portions of missing sequence of *psb*D-*psb*C are: *Ecdeiocolea* and *Elegia* (5' end of *psb*D is 54 bp short), *Medeola* (5' end of *psb*D is 107 bp short), *Petermannia* (only 212 bp of the 3' end of *psb*D was sequenced).

[5] Taxa with portions of missing sequence of the *psb*E-*psb*F-*psb*L-*psb*J region are: *Ecdeiocolia* (5' end of *psb*E is 60 bp short) and *Petermannia* (5' end of *psb*E is 62 bp short).

[6] *Strelitzia* is missing 111 bp from the 5' end of *rpl*2 first exon.

[7] Taxa with portions of missing sequence of the *rps-ndh*B-*trn*L region are: *Medeola* (59 bp missing from 5' end of *rps*12 first exon), *Trillium* (62 bp missing from 5' end of *rps*12 first exon), and *Petrosavia* (*ndh*B intron and second exon missing, only 319 bp of *ndh*B first exon included). The following taxa are missing the *trn*L-*ndh*B intergenic spacer: *Anticlea, Campynema, Cyperus, Lilium, Petermannia, Strelitzia*, and *Wurmbea*.

[8] Sequences were presented by Fay et al. (2000) but are not yet available on GenBank.

Table 4.3. Likelihood ratio test (LRT) for various substitution models. Likelihood scores and parameter estimates are based on the MP topology presented in Fig. 1. The significance value for rejection of the null hypothesis was adjusted using a Bonferroni correction and set to 0.01. In all cases, $P < 0.01$.

Substitution Model [1]	**-ln likelihood**	**Comparison**	**-2 ln Λ** [2]	**d.f.**
JC69	170461.40	-----	-----	
F81	169192.55	JC69 vs. F81	2537.70	3
HKY85	163492.12	F81 vs. HKY85	11400.86	1
GTR	162993.51	HKY85 vs. GTR	997.22	5
GTR + Γ	149479.03	GTR vs. GTR + Γ	27028.96	1

[1] Abbreviations: JC69 = Jukes-Cantor (1969); F81 = Felsenstein (1981); HKY85 = Hasegawa et al. (1985); GTR = General Time-Reversible (Lanave et al., 1984; Tavaré, 1986; Barry and Hartigan, 1987; Rodríguez et al., 1990). Γ = Gamma.

[2] Likelihood ratio test statistic.

Figure legends.

Fig. 4.1. One of two most-parsimonious trees inferred for Liliales and related monocots, based on a large plastid data set (*atp*B, *ndh*B, *ndh*F, ten photosystem genes, *rbc*L, *rps*7, 3'-*rps*12, and various introns and other noncoding regions; see text). Tree length = 29,309 steps, consistency index = 0.404, retention index = 0.483. The arrow indicates a clade not seen in both shortest trees. Parsimony-based bootstrap values are indicated above branches, posterior probability estimates of clade support from a Bayesian phylogenetic analysis are indicated in italics below branches. Posterior probabilities were not estimated for several clades (indicated with "n/a", not applicable), because the Bayesian analysis included fewer taxa than the parsimony analysis (excluded taxa are indicated with asterisks). The order Liliales is highlighted, and all remaining orders are indicated except for those in the commelinid monocots (see Table 4.1.)

Fig. 4.2. Reduced phylogram representing one of two most-parsimonious trees inferred for Liliales and related monocots, based on a large plastid data set (*atp*B, *ndh*B, *ndh*F, ten photosystem genes, *rbc*L, *rps*7, 3'-*rps*12, and various introns and other noncoding regions; see text). Individual taxon names outside Liliales have been removed for clarity, but follow the same order as Fig. 4.1. Family names in Liliales follow APGII (2003), except that Petermanniaceae is recognized as a separate family. Branch lengths were computed using ACCTRAN (accelerated transformation). Letter codes indicate branches of interest in Liliales and relatives (see Fig. 4.3 and text for further details).

Fig. 4.3. Profiles of clade support from analyses of various data partitions based on a large plastid data set, focussing on clades in and around the order Liliales. Four of five sets of values are from a maximum-parsimony bootstrap analysis of a 79-taxon data set. Various symbols (see inset box) represent: (i) all data combined; (ii) codon positions 1 and 2 combined; (iii) codon position 3 combined; (iv) noncoding data combined). One set of values (symbol indicated in inset box) represents Bayesian posterior probability estimates, derived from analysis of a 58-taxon data set, for all of the plastid data combined. All partitions/ analyses yielded non-zero values for all clades. Branch labels correspond to those shown in Fig. 4.2. Branches are ranked according to their value in the parsimony-based bootstrap analysis (for all the plastid data combined).

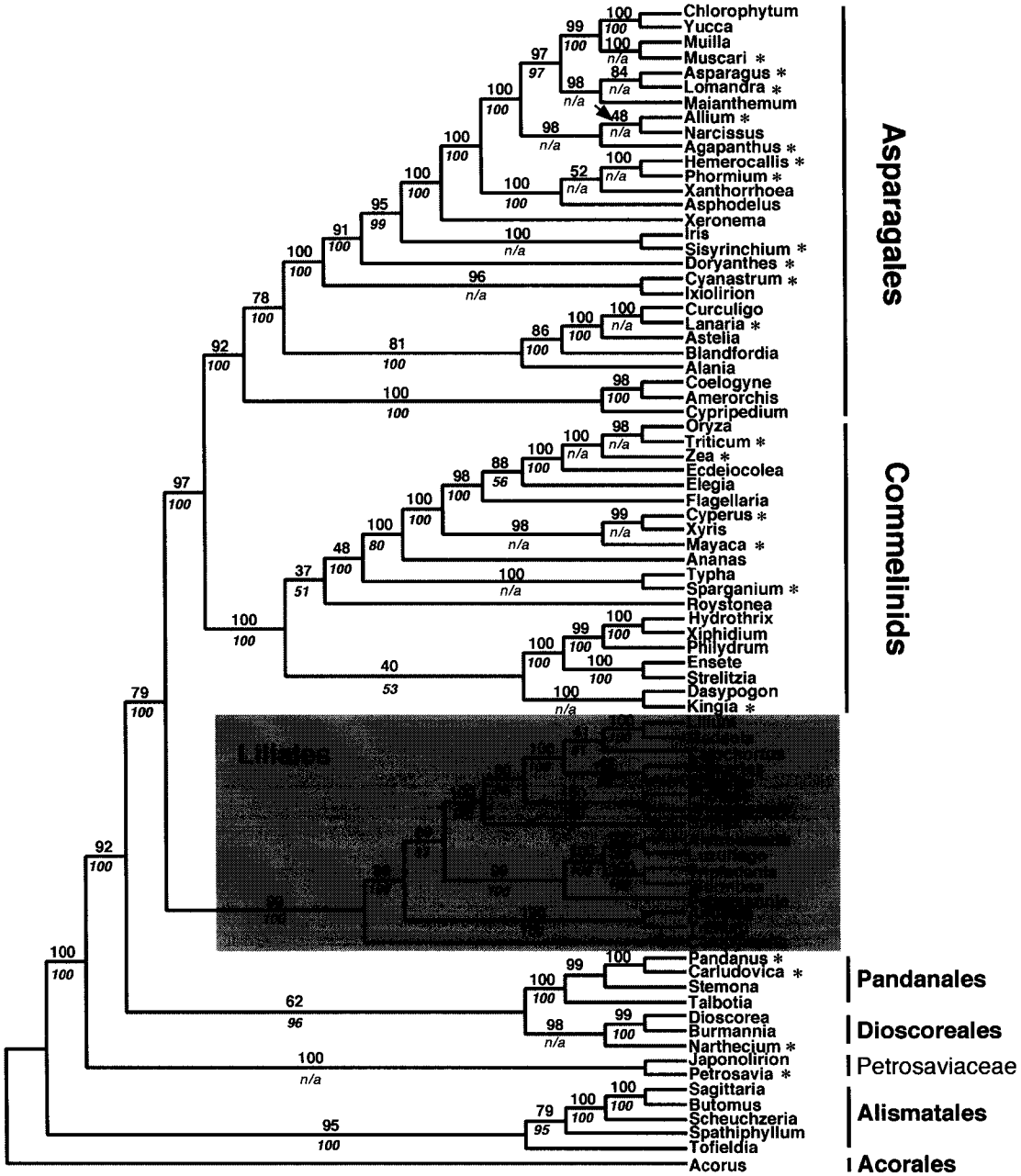


Fig.41

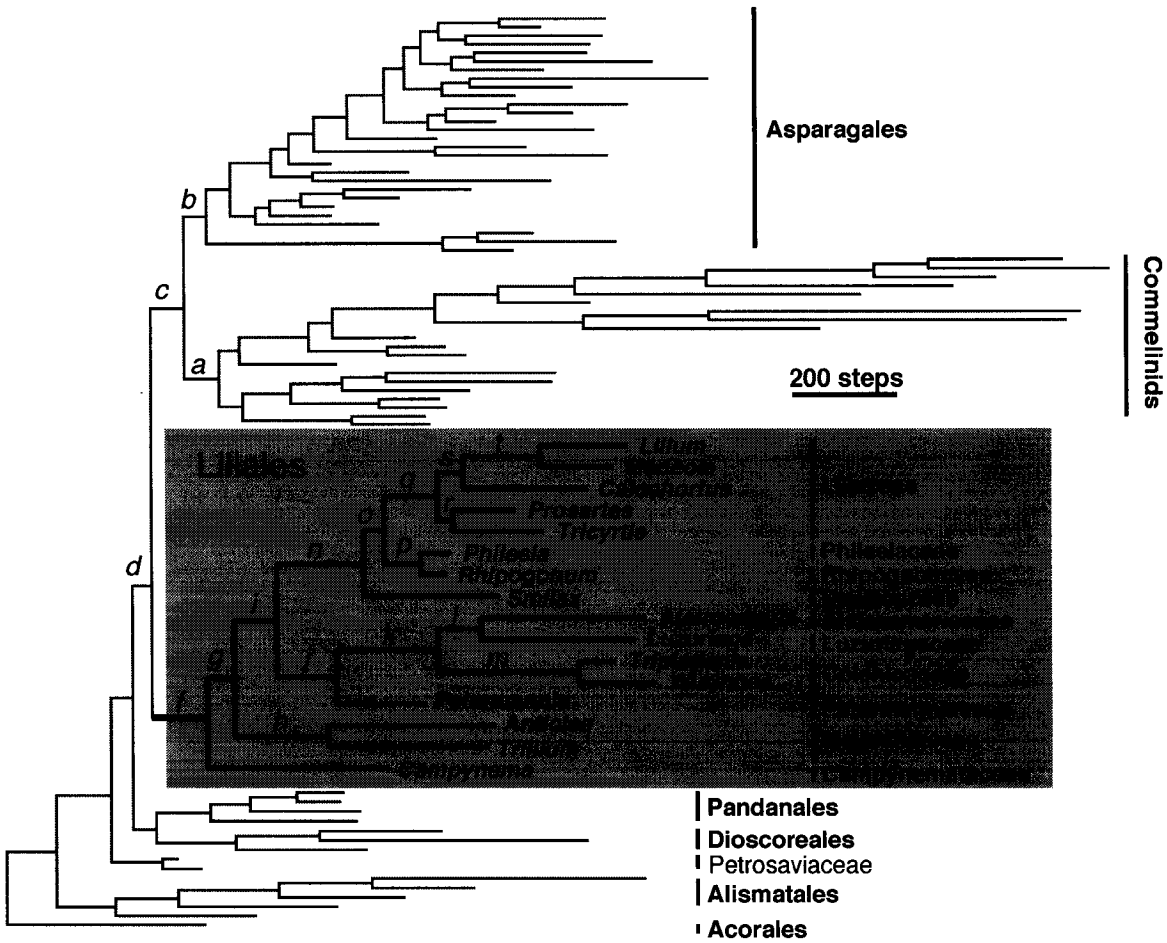


Fig.42

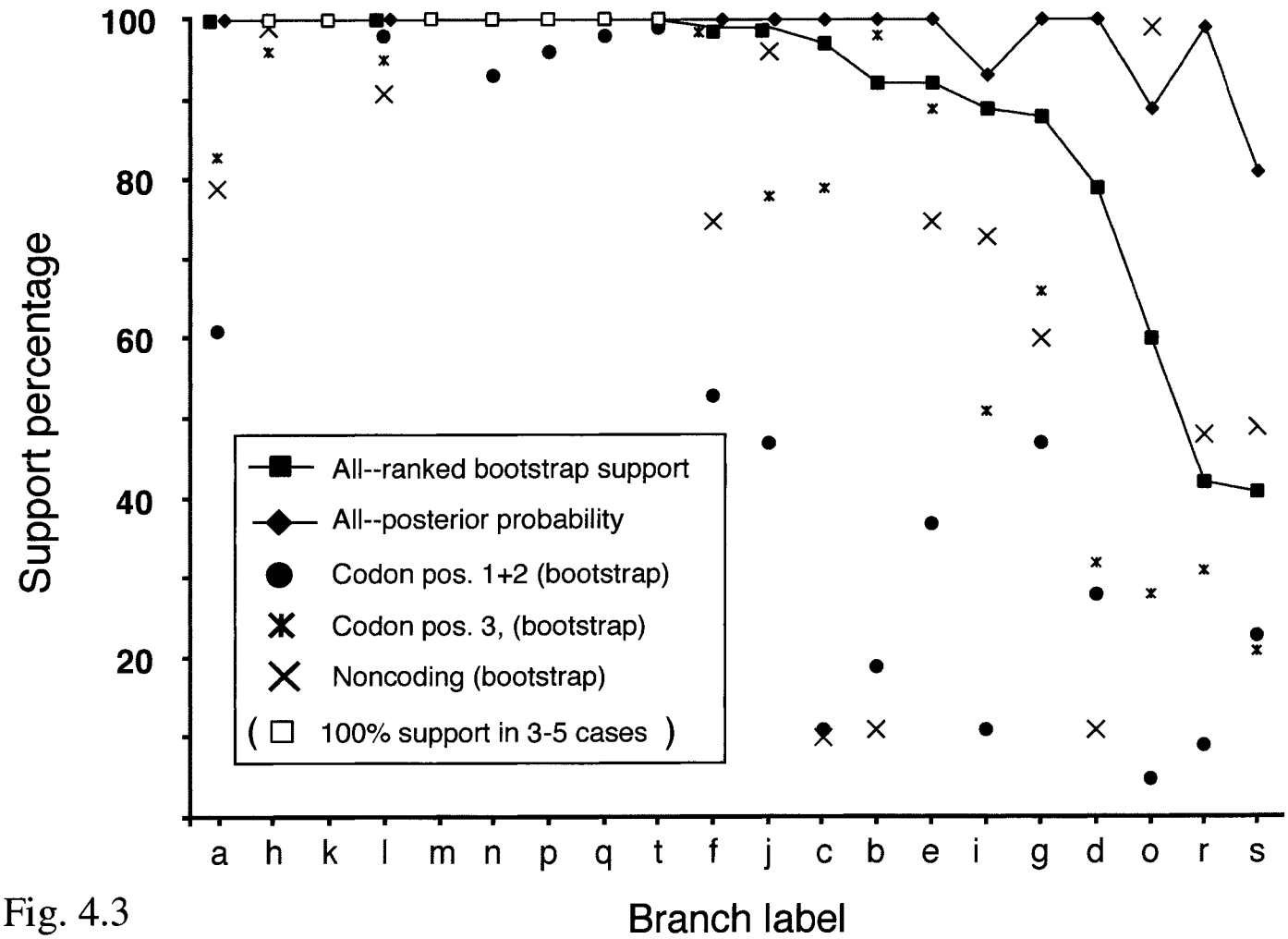


Fig. 4.3

~Chapter Five~

Conclusion

Plant systematists have been unable to resolve satisfactorily many of the deepest aspects of phylogenetic relationship in the vascular plants. Most recent attempts to determine their relationships have used plastid genomic or nuclear ribosomal RNA sequences as a data source. Nuclear protein-coding genes have been used to address deep vascular-plant phylogenetic relationships relatively rarely, and so this genome has been a poorly exploited source of phylogenetic data. Many nuclear genes are present as multiple copies, which has complicated their use as phylogenetic markers. However, the nuclear RNA polymerase II subunit gene *RPB2* has been demonstrated to be single- or double-copy in most vascular-plant groups (Denton et al., 1998; Oxelman and Bremer, 2000; Oxelman et al., in press). I aimed to contribute to a better understanding of deep vascular-plant phylogeny by developing new primers for the comparative study of *RPB2*. I used these to sequence 13-14 exons of *RPB2* for representatives of the major vascular-plant groups.

Within the vascular plants, the position of Gnetales has been particularly controversial. In the pre-molecular era, the position of Psilotaceae was also elusive. Many molecular studies (e.g. Wolf et al., 1997; Pryer et al., 2001) have indicated that Psilotaceae are the sister group of Ophioglossaceae, a result that has not been suggested by morphological data. In agreement with many previous studies (Goremykin et al., 1996; Chaw et al., 1997, 2000; Winter et al., 1999; Bowe et al., 2000; Nickerson and Drouin, 2004), maximum likelihood and Bayesian analyses of the *RPB2* data indicate that Gnetales are more closely related to conifers (specifically Pinaceae) than to angiosperms. Bayesian and

maximum parsimony analyses also strongly support the idea that Psilotaceae are the sister group of Ophioglossaceae.

I performed a series of simulation analyses to determine the error rate of parsimony and likelihood with regards to the positions of Gnetales and Psilotaceae in vascular-plant phylogeny. Phylogenetic trees were constructed that constrain several major hypotheses concerning the placement of each group, and multiple additional data sets for each were simulated using the constrained tree topology, branch lengths and molecular-evolution parameters estimated from the original data set. Both parsimony and likelihood had a high probability of recovering the "correct" (model) tree in simulation analyses that focus on Gnetales placement in seed-plant phylogeny, except when the model tree is consistent with the anthophyte hypothesis (with Gnetales as the sister group of the angiosperms). In the anthophyte simulations, parsimony tended to place Gnetales as the sister group of all remaining seed plants, the topology found by Rydin et al. (2002) and Rai et al. (2003), but not by parsimony analysis of the *RPB2* data here (which instead find Gnetales as the sister group of all other gymnosperms). The anthophyte simulation results imply that, if the anthophyte tree were the "true" tree, parsimony analyses of *RPB2* data would tend not recover it. Likelihood however, had a far lower type 1 error rate than parsimony for the anthophyte hypothesis, suggesting that this method of phylogenetic inference, and presumably other model-based methods such as Bayesian analysis, may be more reliable than parsimony for inferring Gnetales placement using *RPB2* data.

Sanderson et al. (2000) obtained comparable results with simulation analyses of two plastid photosystem genes. Since similar patterns of bias are present in both nuclear and plastid data sets, results from other molecular phylogenetic studies on deep seed-plant relationships should also be viewed with some caution. Sanderson et al. (2000) also found that first- and second-codon

positions (combined) and third codon positions supported different seed-plant topologies. This was not the case with the Bayesian analyses of the *RPB2* data set presented here – both codon position partitions support a sister-group relationship between Pinaceae and Gnetales. However, in Chapter 3, where I investigated seed-plant and cycad relationships using a large plastid data set alone (and in combination with two nuclear loci, *RPB2* and 26S rDNA), separate parsimony analyses of the two plastid codon position partitions (first two vs. third) resulted in strongly incongruent topologies with regards to the position of Gnetales.

I also attempted to determine if the position of Psilotaceae seen here and in other molecular studies was a result of systematic error using this simulation approach. When the observed topology (Ophioglossaceae as the sister group of Psilotaceae) was used to simulate the data sets, the correct tree was recovered 100% of the time. However, the type 1 error rate was quite high for all other hypotheses, and type 2 error rates were also high for most hypotheses examined. Often, both parsimony and likelihood placed Ophioglossaceae as the sister group of Psilotaceae even if that was not the model tree used to simulate the data sets, suggesting that the Ophioglossaceae-Psilotaceae result from *RPB2* data should be treated with caution. However, other "wrong" hypotheses (non-model trees) were also frequently recovered in these simulations. All recent molecular studies have found the Ophioglossaceae-Psilotaceae clade. It would be valuable to perform similar simulation analyses for other plastid, nuclear and mitochondrial genes to determine the amount of bias in these genomes concerning Psilotaceae placement in vascular-plant phylogeny.

Higher-order relationships within two major vascular-plant groups, cycads and Liliales, were examined in Chapters 3 and 4, respectively. Cycad relationships were examined using a large data set composed of both plastid and nuclear genes, including the *RPB2* sequences presented in Chapter 2. Parsimony

analysis of concatenated genes and separate analysis of the first two vs. third codon positions in plastid protein-coding genes place cycads as the sister group of *Ginkgo*, with moderate to strong support. Analysis of transitions or transversions alone found a different placement, but with poor support. Rai et al. (2003) found cycads and *Ginkgo* to be sister taxa, but with only weak support. Bayesian analysis of *RPB2* alone (Chapter 2) also find a cycad-*Ginkgo* relationship, with strong support, although analysis of the first two codon positions alone produced a moderately conflicting result (with *Ginkgo* as the sister group of conifers and Gnetales).

Relationships within cycads are generally well-resolved by the data presented in Chapter 3, and there is only one strongly supported conflict between subsets of the data, concerning the precise relationship between *Encephalartos*, *Lepidozamia* and *Macrozamia*. The plastid data and the nuclear locus 26S rDNA agree that these three taxa form a clade, but disagree strongly about their exact relationship to each other. *Cycas* and *Dioon* are, respectively, the successive sister groups of all other (extant) cycads with strong bootstrap and Bayesian support. *Stangeria* and *Bowenia* are not closely related to each other, indicating that the family Stangeriaceae is not monophyletic. My analysis supports the suggestion by Hill et al. (2003) that only two families of cycads should be recognized: Cycadaceae (for *Cycas*) and Zamiaceae (for all other genera).

I also used a large plastid data set to examine the phylogenetic placement of the order Liliales among the other major groups of monocots, and to infer higher-level relationships within this order. Parsimony and Bayesian analyses indicate that Liliales are the sister group of a large clade composed of Asparagales (the order that now includes orchids, irises, onions and daffodils) and commelinids (a large clade that includes grasses, sedges, gingers and palms). This relationship has been recovered in several recent studies of monocot

phylogeny, but with only poor support. It is moderately well-supported by bootstrap and Bayesian support values here.

The circumscription of Liliales varied widely in the pre-molecular era. Cronquist (1988) placed many taxa in a single family, Liliaceae, including many that are now housed in Asparagales and other orders of monocots. He recognized that the family was most likely polyphyletic, but could not find a satisfactory way to classify the lilioid monocots. Dahlgren et al. (1985) used morphological evidence to place many Liliaceae in Asparagales or other families in Liliales, and their classification is broadly congruent with that of APG II (2003), based largely on molecular data. Even so, the precise placement of higher-level taxa within Liliales have remained unclear in recent studies. In my study (Chapter 4), I was able to infer most aspects of interfamilial relationships within Liliales with strong support. The small Australasian family Campynemataceae is the sister group of the remainder of the order, and Melanthiaceae are the sister group of the remaining families. The results here are congruent with a circumscription of Melanthiaceae by APG II (2003) that incorporates Trilliaceae, as I find *Trillium* and *Anticlea* (= *Zigadenus*) to be very closely related. Two remaining large clades are well-supported here: one consisting of Alstroemeriaceae, Colchicaceae, Luzuriagaceae and *Petermannia*, and a second comprised of Liliaceae, Philesiaceae, Rhipogonaceae and Smilacaceae. *Petermannia cirrosa*, which was recently placed in Colchicaceae based on a misidentified sample (Vinnersten and Reeves, 2003), is clearly not part of that family, supporting its recognition as its own family, Petermanniaceae.

The use of molecular data to examine phylogenetic relationships among vascular plants at a variety of taxonomic levels has become a highly successful tool for plant systematists. Here, I used very large molecular data sets to examine relationships within cycads and Liliales. Phylogenetic inferences in these two

groups are generally well-resolved and supported by the genomic regions I surveyed, and there was little incongruence among major partitions of the data sets in these two studies. Deeper vascular-plant relationships inferred using *RPB2* conifers and ferns are largely congruent with those found in other studies (Hasebe et al., 1995; Stefanovic et al., 1998), supporting the value of *RPB2* for more detailed studies within each of these groups. However, simulation analyses demonstrate that molecular data can sometimes be strongly misleading, as maximum parsimony analysis (often), and likelihood analyses (sometimes), did not recover the correct model tree in various simulations studies, or recovered the incorrect model tree, in some cases the one actually observed using real data. My thesis presents some major new results on deep vascular-plant phylogeny and more detailed studies of two major groups, the cycads and lilies. The primers I developed for analysis of *RPB2* should be useful for other workers interested in resolving relationships within the major vascular-plant lineages. My simulation analyses also demonstrate that caution is appropriate for interpreting the results of all current studies of deep vascular-plant phylogeny.

References

Angiosperm Phylogeny Group. 1998. An ordinal classification for the families of flowering plants. *Annals of the Missouri Botanical Garden* **85**: 531–553.

Angiosperm Phylogeny Group II. 2003. An update of the angiosperm phylogeny group classification for the orders and families of flowering plants: APG II. *Botanical Journal of the Linnean Society* **141**: 399–436.

Bowe, L. M., Coat, G., and dePamphilis, C. W. 2000. Phylogeny of seed plant based on all three genomic compartments: Extant gymnosperms are monophyletic and Gnetales' closest relatives are conifers. *Proceedings of the National Academy of Sciences USA* **97:** 4092–4097.

Chaw, S., Zharkikh, A., Sung, H., Lau, T., and Li, W. 1997. Molecular phylogeny of extant gymnosperms and seed plant evolution: Analysis of 18S rRNA sequences. *Molecular Biology and Evolution* **14**: 56–68.

Chaw, S-M., Parkinson, C. L., Cheng, Y., Vincent, T. M., and Palmer, J. D. 2000. Seed plant phylogeny inferred from all three plant genomes: Monophyly of extant gymnosperms and origin of Gnetales from conifers. *Proceedings of the National Academy of Sciences USA* **97:** 4086–4091.

Cronquist, A. 1988. *The Evolution and Classification of Flowering Plants,* 2nd edition. The New York Botanical Garden, Bronx, New York.

Dahlgren, R. M. T., Clifford, H. T., & Yeo, P. F. 1985. *The Families of Monocotyledons: Structure, Evolution and Taxonomy* Springer-Verlag, Berlin.

Denton, A. L., McConaughy, B. L., and Hall, B. D. 1998. Usefulness of RNA polymerase II coding sequences for estimation of green plant phylogeny. *Molecular Biology and Evolution* **15**: 1082–1085.

Goremykin, V., Bobrova, B., Pahnke, J., Troitsky, A., Antonov, A., and Martin, W. 1996. Noncoding sequences for the slowly evolving chloroplast inverted repeat do not support gnetalean affinities of angiosperms. *Molecular Biology and Evolution* **13**: 383–396.

Hasebe, M., Wolf, P. G., Pryer, K. M., Ueda, K., Ito, M., Sano, R., Gastony, G. J., Yokoyama, J., Manhart, J. R., Murakami, N., Crane, E. H., Haufler, C. H., and Hauk, W. D. 1995. Fern phylogeny based on *rbc*L nucleotide sequences. *American Fern Journal* **85**: 134-181.

Hill, K. D., Chase, M. W., Stevenson, D. W., Hills, H. G., and Schutzman, B. 2003. The families and genera of cycads: A molecular phylogenetic analysis of Cycadophyta based on nuclear and plastid DNA sequences. *International Journal of Plant Sciences* **164**: 933-948.

Nickerson, J., and Drouin, G. 2004. The sequence of the largest subunit of RNA polymerase II is a useful marker for inferring seed plant phylogeny. *Molecular Phylogenetics and Evolution* **31**: 403–415.

Oxelman, B., and Bremer, B. 2000. Discovery of paralogous nuclear gene sequences coding for the second-largest subunit of RNA polymerase II (*RPB*II) and their phylogenetic utility in Gentianales of the asterids. *Molecular Biology and Evolution* **17**: 1131–1145.

Oxelman B., Yoshikawa, N., McConaughy, B. L., Luo, J., Denton, A. L., and Hall, B. D. In press. RPB2 gene phylogeny in flowering plants, with particular emphasis on asterids. *Molecular Phylogenetics and Evolution.*

Pryer, K. M., Schneider, H., Smith, A. R., Cranfill, R., Wolf, P. G., Hunt, J. S., and Sipes, S. D. 2001. Horsetails and ferns are a monophyletic group and the closest living relatives to seed plants. *Nature* **409**: 618–622.

Rai, H. S., O'Brien, H. E., Reeves, P. A., Olmstead, R. G., and Graham, S. W. 2003. Inference of higher-order relationships in the cycads from a large chloroplast data set. *Molecular Phylogenetics and Evolution* **29**: 350–359.

Rydin, C., Kallersjo, M., and Friis, E. M. 2002. Seed plant relationships and the systematic position of Gnetales based on nuclear and chloroplast DNA: conflicting data, rooting problems and the monophyly of conifers. *International Journal of Plant Science* **163**: 197–214.

Sanderson, M. J., Wojciechowski, M. F., Hu, J. M., Sher Khan, T., and Brady, S. G. 2000. Error, bias and long-branch attraction in data for two chloroplast photosystem genes in seed plants. *Molecular Biology and Evolution* **17**: 782–797.

Stefanovic, S., Jager, M., Deutsch, J., Broutin, J., and Masselot, M. 1998. Phylogenetic relationships of conifers inferred from partial 28S rRNA gene sequences. *American Journal of Botany* **85**: 688–697

Vinnersten, A., and Reeves, G. 2003. Phylogenetic relationships within Colchicaceae. *American Journal of Botany* **90**: 1455–1462.

Winter, K.-U., Becker, A., Munster, T., Kim, J. T., Saedler, H., and Theissen, G. 1999. MADS-box genes reveal that gnetophytes are more closely related to conifers than to flowering plants. *Proceedings of the National Academy of Sciences USA* **96**: 7342–7347.

Wolf, P. G. 1997. Evaluation of *atpB* nucleotide sequences for phylogenetic studies of ferns and other pteridophytes. *American Journal of Botany* **84**: 1429–1440.