Université d'Ottawa • University of Ottawa

Université d'Ottawa · University of Ottawa

FACULTÉ DES ÉTUDES SUPÉRIEURES
ET POSTDOCTORALES

FACULTY OF GRADUATE AND
POSTDOCTORAL STUDIES

XIA, Junnan
AUTEUR DE LA THÈSE - AUTHOR OF THESIS

M.Sc. (Biology)
GRADE - DEGREE

Department of Biology
FACULTÉ, ÉCOLE, DÉPARTEMENT - FACULTY, SCHOOL, DEPARTMENT

TITRE DE LA THÈSE - TITLE OF THE THESIS

The Largest Subunit of RNA Polymerase II (rpb1) as a
Phylogenetic Marker of Seed Plant Species

Guy Drouin
DIRECTEUR DE LA THÈSE - THESIS SUPERVISOR

EXAMINATEURS DE LA THÈSE - THESIS EXAMINERS

L. Gillespie

J. Vierula

X. Xia

J.-M. De Koninck, Ph.D.

LE DOYEN DE LA FACULTÉ DES ÉTUDES
SUPÉRIEURES ET POSTDOCTORALES

SIGNATURE

DEAN OF THE FACULTY OF GRADUATE
AND POSTDOCTORAL STUDIES

The largest subunit of RNA polymerase II (*rpb1*) as a phylogenetic marker of seed plant species

JUNNAN XIA

Thesis submitted to the
Faculty of Graduate and Postdoctoral Studies
University of Ottawa
in partial fulfillment of the requirements for the
M.Sc. degree in the
Ottawa-Carleton Institute of Biology

Library and
Archives Canada

Published Heritage
Branch

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque et
Archives Canada

Direction du
Patrimoine de l'édition

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence
ISBN: 0-494-01647-7
Our file Notre référence
ISBN: 0-494-01647-7

NOTICE:
The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

AVIS:
L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

Canada

Acknowledgements

I would like to extend my sincere gratitude and appreciation to many people who made this master's thesis possible.

First and foremost, I would express great thanks to my supervisor Dr. Guy Drouin. This master's thesis would not have been possible without his guidance, encouragement and consideration. His broad and profound knowledge, his warm smile and his patient instruction have given me a great help. I would also like to thank my committee members, Dr. John Vierula and Dr. Lynn Gillespie for their valuable suggestions.

I am deeply indebted to Jennifer Nickerson and James Nickerson, my former lab mates and good friends. You are always there and warm my heart. Furthermore, without the step by step lab training Jennifer gave me at the beginning of this project, I could not have finished my lab work so quickly.

I would also like to acknowledge with much appreciation the help of my lab mates past and present, Hongyan Zhou, Mehrdad Hajibabaei, Robert Morris and David-Maxime Benovoy. Your friendly help and encouragement were important to me.

I would like to thank Mehrdad Hajibabaei for providing *Taxus* clones of DG fragments, and Jennifer Nickerson, Yin-Long Qiu (University of Massachusetts), Richard Olmstead (Washington University), Alan Whittemore (US National Arboretum), Chris Graham (Royal Botanical Gardens, Hamilton, Ontario), Lynn Gillespie (Canadian Museum of Nature, Gatineau), Hugette Allard (University of Ottawa, greenhouse) and Ed Bruggink (Carleton University, greenhouse) for providing plant tissues.

I would like to extend my thanks to the Ontario Graduate Scholarship (OGS) for their financial support.

I cannot end my acknowledgement without thanking my family, on whose constant encouragement and love I have relied throughout my lonely time in Canada. I wish to give a very special thank you and all my love to my deeply loved husband Bing Gao and son Xiashuang. Without your every evening oversea calls, I doubt that this thesis would ever have been written.

Abstract

The phylogeny of land plants, especially angiosperms, has been perplexing scientists for more than 125 years. The aim of this study was to help elucidate unanswered phylogenetic issues using *rpb1* genes from twenty-six land plant species (8 gymnosperm species, 17 angiosperm species and *Psilotum*, a fernally, as the outgroup). The sixteen *rpb1* genes which were sequenced in this study had very similar lengths and contained no base compositional bias. Synonymous substitutions of *rpb1* sequences were saturated when compared to the outgroup. The third codon positions of these genes contained misleading phylogenetic information. The topology of trees based on first and second codon positions were in line with that of protein trees. Both angiosperms and gymnosperms were monophyletic. *Amborella* was found at the base of the angiosperm tree, followed by *Nymphae*, then *Illicium*. These data rejected the anthophyte hypothesis, weakly supported the gnepine hypothesis, but did not resolve the interrelationship among eumagnoliids and eudicots. *rpb1* genes were combined with the 18S ribosomal RNA gene and chloroplast *atpB* and *rbcL* genes to obtain more robust phylogenies. This combined data set produced a topology similar to that of the first and second positions of *rpb1* genes except that they better resolved the interrelationship among eumagnoliids and eudicots. We conclude that our *rpb1* sequences evolve too slowly to provide enough phylogenetic information to fully resolve the phylogeny of seed plants and that other gene sequences will need to be added to these data sets to obtain a well-resolved phylogenetic tree of seed plants.

Table of Contents

Acknowledgements … .. i

Abstract …iii

Table of Contents …iv

Introduction …1

Materials and Methods … 18

Results … 34

Discussion … 52

Conclusion …62

References … 65

Appendix … ..78

Introduction

Phylogenies are regarded as the central underpinning of research in much of biology by most systematists and evolutionary biologists. Soltis and Soltis (2000) stated that everything makes a lot more sense in the light of phylogeny. Plant evolution has affected the functioning of the planet. Morphological innovations of plants have profoundly changed the terrestrial biosphere. Plant innovations have affected global geochemistry, hydrology, atmospheric composition and the evolution of other biota. Molecular studies have revealed that land plants arose from their green algal ancestor and were phylogenetically placed within the Charophyta, a sister group to the Chlorophyta (Karol *et al.*, 2001).

Extant seed plants include angiosperms and four groups of gymnosperms: cycads, conifers, *Ginkgo*, and gnetales. They differ from all other living land plants by several characters, especially reproduction via seeds. Among all plants, the angiosperms play an indispensable role. The angiosperms dominate the terrestrial biota with approximately 440 families and nearly 250,000 species. The rich species diversity of angiosperms relative to other seed plants is due to their rapid life cycle relative to gymnospermous seed plants, the close relationships with anthophilous insects and phytophagous insects, and dispersal of fruits (Crepet, 1996). Angiosperms are the primary sources of foods and drugs. Understanding the phylogenetic relations of angiosperms is very important and absolutely necessary. First, it can facilitate comparative studies of plant development, metabolism, reproduction, pathology and genomics, and these comparative studies also can be used to infer and construct the evolutionary history of a broad range of biological characters (Floyd *et al.*, 1999). Second, it would allow us to understand species distributions and their ecological implications better and provide more efficient phylogenetic context-guided searches for natural drugs. Finally, it can allow more informed decision to be made about biodiversity conservation issues by permitting comparisons for the explicit uniqueness of taxa in situations involving difficult choices (Crepet, 2000).

However, the phylogeny of seed plants especially angiosperms has been perplexing scientists for more than 125 years, since Charles Darwin called the sudden appearance of angiosperms in the fossil record as an "abominable mystery". There are three major obstacles that hinder reconstructing the phylogeny among extant seed plants (Friedman and Floyd, 2001). First, the fossil record failed to provide strong evidence on which angiosperm lineages are truly most ancient. Different lineages including monocots (Crane *et al.*, 1995), Platanaceae (Friis *et al.*, 1994), Ceratophyllaceae (Dilcher, 1989), Nelumbonaceae (Upchurch *et al.*, 1994), Laurales (Upchurch *et al.*, 1994), Winteraceae (Walker *et al.*, 1983), and Chloranthaceae (Friis *et al.*, 1986) all appeared suddenly in the fossil record in the Cretaceous. There is also a lack of fossil leaves and pollen for a period of about 100 million years between 250 and 150 MYA, when all the gymnosperms were present and before the angiosperms appeared. Furthermore, many of the early fossils show a mix of features that define modern groups, making them difficult to interpret. Second, some lineages are extinct and are found only in the fossil record. This complicates molecular phylogenies because many extant lineages are represented by long branches. Third, relatively low numbers of morphological characters are available for phylogenetic reconstruction and these characters typically provide only weak or ambiguous support for the phylogenetic analysis of deep angiosperm relationships.

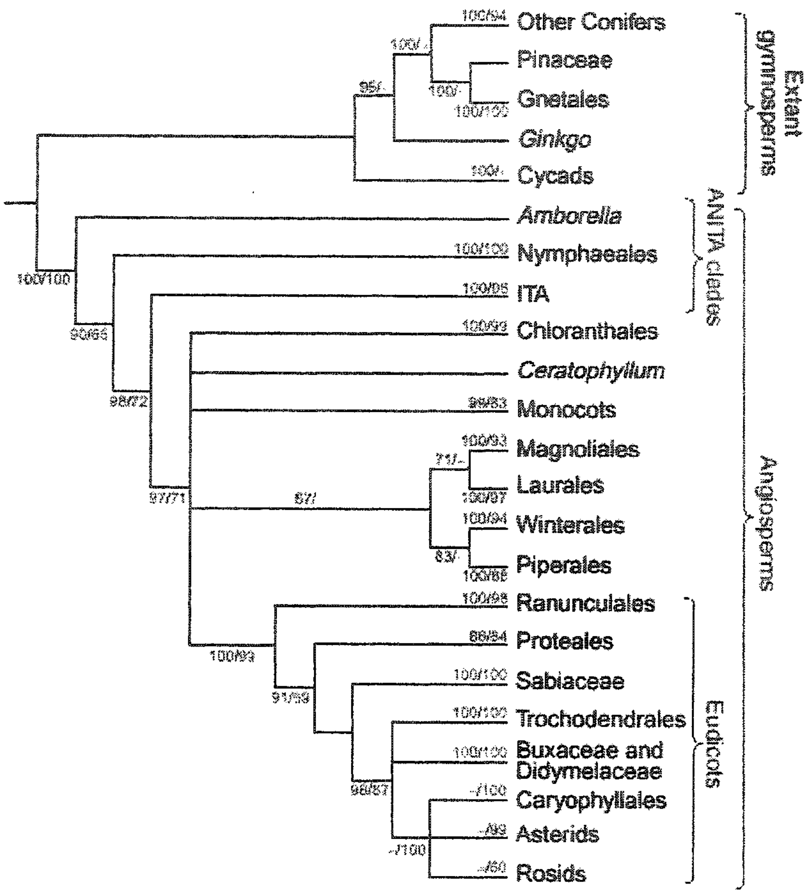
Basal angiosperms

At least six different hypotheses for the origin of extant angiosperms have been proposed (Loconte 1996). The *Magnoliales*, the *Winteraceae*, the *Calycanthales,* the *Chloranthaceae,* the *Ceratophyllaceae* or the *Casuarinaceae* were all suggested as the most primitive angiosperms. With the development of molecular biology, molecular data have been introduced to the phylogenetic studies of angiosperms. The genes used include the mitochondrial *atp1* and *matR* genes, the plastid *atpB* and *rbcL* genes, and the nuclear 18S rDNA genes. They encode products involved in metabolism, carbohydrate synthesis and information processing. Before 1999, molecular methods proved to be of relatively little help to unravel the origin and phylogeny of angiosperms because they were based on single genes. Single genes are insufficient to infer the phylogeny of angiosperms because of the evolutionary rate heterogeneity among lineages of the particular gene used and the weak phylogenetic signal and information they contain. For example, the *rbcL* data set reported by Chase *et al.* (1993) contained 499 sequences, and the 18S ribosomal RNA (rRNA) data set of Soltis *et al.* (1997) contained 223 sequences. Although the phylogenetic trees derived from these two data sets were generally congruent (e.g., they both showed that monocots are derived from dicots), they also showed some noteworthy differences. For example, the Chase *et al.* (1993) study supported *Ceratophyllum*, a paleoherb, as the most basal angiosperm whereas the Soltis *et al.* (1997) study supported the ranalean theory, placing woody magnoliids at the base of the angiosperm tree.

Furthermore, the study of Chaw *et al.* (1997), which like the Soltis *et al.* (1997) study, also used 18S rRNA sequences, supported water lilies as being the most primitive angiosperm. These three studies therefore supported three different and incompatible hypotheses.

The year 1999 has witnessed a breakthrough in land plant phylogeny. Multiple genes began to be used to infer the phylogenetic relationships of land plants and they all gave largely congruent results, because multiple genes of different functions can reduce homoplasy generated by gene-, function- and genome-specific molecular evolutionary phenomena such as rate heterogeneity, GC content bias, and protein functional constraints. Those multiple genes came from the mitochondrial, chloroplastic and nuclear genomes. Mathews and Donoghue (1999) sampled and analyzed two paralogous phytochrome genes, *phyA* and *phyC*, which are duplicated only among angiosperms, from 26 species which were thought to be the most basal angiosperms. The sequences of 17 chloroplast genes (~13.4 kb) were sampled from each of 3 gymnosperms and 18 putative basal angiosperm lineages by Graham and Olmstead (2000), to identify the first-branching lineages of angiosperms and assess the effects of greatly increased character sampling and to determine the phylogenetic usefulness of 14 previously untested chloroplast genes. Soltis *et al.* (1999) increased sampling number by using 560 species with 4733 nucleotides per species. Chloroplast *rbcL* and *atpB* genes and nuclear 18S rDNA genes were also sampled. Qiu *et al.* (1999) used five genes from three genomes which included chloroplast *rbcL* and *atpB* genes, nuclear 18S rDNA genes, and mitochondrial *matR* and *atp1* genes. These genes provided 8733 aligned nucleotides and were sampled from 105 major gymnosperms and potential primitive angiosperms lineages. Parkinson *et al.* (1999) also used five genes from three genomes: *atpB* genes (chloroplast), 18S rDNA genes (nuclear), *mtSSU, cox1* and *rps2* genes (mitochondria) where 6564 nucleotides were sampled from 51 gymnosperms and angiosperms. These studies all strongly supported species of the "ANITA clade" (*Amborella, Nymphaeales, Illicales, Trimeniaceae,* and *Austrobaileyaceae*) as being the first lineages at the base of the angiosperm tree. *Amborella* was found at the base of the angiosperm tree, followed by *Nymphaeales* and then ITA members (reviewed in Kuzoff and Gasser, 2000; Figure 1). These studies therefore ruled out many of the previous hypotheses regarding the nature of the earliest angiosperms.

Amborella is an evergreen woody shrub in New Caledonia. It lacks vessels. According to the characters of *Amborella,* the common ancestor of all extant angiosperms probably had a woody habit, vessel-less wood, unilacunar leaf nodes with two traces, leaves with chloranthoid teeth along their margins and no ethereal oils. Flowers of this common ancestor probably had an undifferentiated perianth arranged in more than two cycles or series, perianth appendages that were unfused above the base and anthers that shed pollen towards the center of the flower. Carpels in these flowers were urnshaped (ascidiate), were not attached to one another (apocarpous) and had margins that did not fuse completely but were closed at maturity by secretions (reviewed in Kuzoff and Gasser, 2000). However, in 2000, Barkman *et al.* reported another different result from genes of all three genomic compartments and suggested that *Nymphaea*, a water lily, shares the first branch with *Amborella*. Nymphaeales and *Amborella* both seem to lack



Trends in Plant Science

Figure 1. Simplified supertree based on five multigene phylogenies for seed plants. Support values for clades in the supertree are based on parsimony bootstrap (before the slash) and jackknife (after the slash) analyses from two source trees ('-' indicates <50% support) (Taken from Kuzoff and Gasser, 2000).

vessel elements homologous to those seen in practically all other angiosperms. They also both lack the ethereal oils that are present in most other basal angiosperm lineages. It implies that an angiosperm ancestor might share characters with *Nymphaea* as well as with *Amborella.* This is an important contrast because *Nymphaea* has vessels and bisexual flowers while *Amborella* has unisexual flowers.

A new basal angiosperm family, Archaefructaceae, was recently discovered in the fossils. This family is believed to be between 125 and 145 million years old (Sun *et al.*, 2002). It is represented only by two species, *Archaefructus liaoningensis* and *Archaefructus sinensis*. The fossils of flowers, seeds, and fruits, were found in northeastern China. These fossils suggest that the first flowering plants may have been submerged aquatic plants similar to the Nymphaeales (water lily family) because *Archaefructus* came out as the sister group to all living angiosperms, even closer to the common ancestor than the woody *Amborella* (Stokstad, 2002).

Position of Gnetales

Extant gymnosperms include conifers, cycads, gnetales and *Ginkgo*. The phylogenetic position of the gnetales is still the subject of debates. First, from morphological data, gnetales have traditionally been considered to be most closely related to angiosperms since Haeckel first proposed it in 1894 (Chaw *et al.*, 1997). More recently, Crane (1985) applied cladistic methodology to evaluate relationships of flowering plants and gymnosperms and his analysis suggested that gnetales were indeed the closest living relatives of flowering plants. Several studies from Doyle and Donoghue

(1986, 1992, and 1993) also gave the same result. Gnetales have only three genera, *Gnetum*, *Ephedra*, and *Welwitschia*. *Gnetum* is a tropical rainforest tree or vine, *Ephedra* is a desert shrub with small, pointed leaves, and *Welwitschia,* with only one pair of strap-shaped leaves, grows in the deserts of southwest Africa. These three genera are divergent, but they all resemble most angiosperms in having the flower-like appearance of reproductive units, special water-conducting vessels in the secondary wood, and a kind of double-fertilization (Doyle, 1996). Therefore taxonomists grouped them as well as angiosperms, Bennettitales and *Pentoxylon* as anthophytes (Doyle and Donoghue, 1986). This anthophyte hypothesis is supported by morphological data and leads to a sister-group relationship between gnetales and angiosperms. Bennetitales reproductive units have flat structures surrounding male organs with ovules in the middle. This arrangement resembles the order of parts in flowers (sepals and/or petals surrounding stamens with carpels in the middle). Gnetales plants are functionally either male or female. Their reproductive units have flat structures (bracts) surrounding either male structures or an ovule, but in *Welwitschia*, the male structures surround a sterile ovule. The similarity in overall organization suggested that this pattern evolved only once and was homologous in the three groups. (Male and female structures are borne separately in Pentoxylales.) If it is true, this would mean a major feature of flowering plants evolved millions of years before the flowering plants themselves and was shared (although in modified form) by the living gnetales.

Some molecular analyses are consistent with the anthophyte hypothesis (Hamby and Zimmer, 1992; Stefanovic *et al.,* 1998; Rydin *et al*, 2002). Stefanovic *et al.* (1998)

used a relatively short length of nuclear large subunit rDNA (638 bp) and their analysis weakly support an anthophyte clade. Rydin *et al.* (2002) investigated the systematic position of gnetales and other seed plant groups using molecular data from 119 land plant species. More than 100 new sequences of *rbcL*, *atpB*, 26S, and 18S ribosomal DNA were analyzed together with those available in GenBank. Their study showed that a sister relationship between gnetales and angiosperms could not be fully ruled out but a close relationship between gnetales and the conifer family Pinaceae was rejected.

Other studies (Hickey and Taylor, 1996; Nixon *et al.*, 1994) suggested that gnetales were the direct ancestors of flowering plants instead of gnetales being a sister group of angiosperms. Crane (1985) and Nixon *et al.* (1994) supported a gnetales/angiosperm alliance according to morphological, anatomical and histological studies on structural features. In the Nixon *et al.*'s (1994) study, gnetales are paraphyletic, and *Gnetum* and *Welwitschia* form a sister group to angiosperm. Friedman's studies (1990, 1994) showed that gnetales undergo a kind of double fertilization like angiosperms, but that they don't form triploid endosperm.

On the other hand, many recent molecular studies have shown that gnetales are most closely related to conifers. These molecular studies therefore put an end to the anthophyte hypothesis. A study by Chaw *et al.* (1997), using nuclear 18S rRNA sequences, showed that angiosperms and gymnosperms are two separate monophyletic lineages and that none of the extant gymnosperms is the ancestor of flowering plants. Their results also suggested that the gnetales are a sister group to the conifers. Studies of

MADS-box genes (Winter *et al.*, 1999) and chloroplast genes (Hansen *et al.*, 1999) also placed the gnetales as sister to the conifers. In 2000, based on the sequences of mitochondrial small subunit rRNA genes, nuclear small subunit rRNA genes and chloroplast *rbcL* genes, Chaw *et al.* (2000) proposed a new idea, the gnepines hypothesis, where the gnetales were suggested to have arisen from within the conifers as the sister group to Pinaceae. The remaining conifers, including Araucariaceae, Podocarpaceae, Taxodiaceae, Taxaceae, Cupressaceae, and allied taxa, constitute a "Conifer II" clade. The gnetales may thus be viewed as extremely divergent conifers, and this suggests that many morphological similarities between angiosperms and gnetales arose independently. The gnepines hypothesis has been supported by two other recent analyses based on molecular sequence data from multiple genes (Qiu *et al.*, 1999; Bowe *et al.*, 2000). Qiu *et al.* (1999) analyzed five mitochondrial, plastid and nuclear genes. Bowe *et al.*'s (2000) study based on two mitochondrial genes, *cox*1 and *atp*A, unambiguously rejected the anthophyte hypothesis and favored a close relationship between gnetales and conifers. Parsimony- and likelihood-based analyses of plastid *rbcL* and nuclear 18S rDNA alone and with *cox*1 and *atpA* also strongly support a gnetales-conifer grouping. Except for nuclear rDNA, three of four genes and combined three-genome analyses also suggested or strongly supported that gnetales are sister to Pinaceae. Parsimony and Maximum likelihood analyses of first and second codon positions of conserved chloroplast genes, *psaA* and *psbB*, also placed angiosperms and gymnosperms as sister clades and gnetales as sister to Pinaceae (Magallón and Sanderson, 2002). The recent *rpb1* phylogenetic analyses in our lab (Nickerson and Drouin, in press) also disagreed with the anthophyte hypothesis. However, this study could not address the gnepines hypothesis because pine was the only conifer present in this data set.

If the gnepines hypothesis is correct, the morphological and ultrastructural traits that have been regarded as uniquely shared by gnetales and angiosperms, such as flower-like reproductive structures, lignin chemistry, a tunica in the apical meristem, pollen with granular exine, reduction of the megaspore wall, and vessel-like conducting elements, were most likely derived separately in the two groups (Bowe *et al.*, 2000). The process of double fertilization in gnetales almost certainly arose separately from, rather than being homologous with, the classical double fertilization of angiosperms (discussed in Chaw *et al.*, 2000). The gnepine hypothesis not only conflicts with morphological, anatomical, and histological data, but it also conflicts with evidence from the fossil record. The relationship between gnetophytes, angiosperms, and conifers therefore remains an open question.

Although the initial branch order in the angiosperms now appears well established with strong support, and the phylogenetic tree shown in Figure 1 is now generally accepted, many details still need to be resolved. For example, the root of the angiosperm tree remains elusive; the relationships among the six subclades of eumagnoliids (Chloranthales, monocots, Piperales, Laurales, Winterales and Magnoliales) need to be clarified. The placement of Ceratophyllaceae also remains a major question. The gnepine hypothesis also needs more support from more molecular data.

Recent empirical studies have demonstrated that adding more taxa and more characters lead to more robust phylogenies (Soltis and Soltis, 2000). Additional taxa can break up long branches that would otherwise attract each other and can make the problem easier to solve. The combined data sets showed great improvements in computer run times compared to the separate data sets and also have higher internal support for clades (Soltis and Soltis, 2000). I therefore sequenced *rpb1* genes (genes of the largest subunits of RNA polymerase II) and combined them with other available gene sequences (*atpB*, *rbcL* and 18S genes) to address these phylogenetic issues.

RNA polymerase

A key step in gene expression is the transcription of DNA sequences. RNA polymerase is the enzyme responsible for transcription and is conserved in all living organisms from *Escherichia coli* to human. Bacteria and archaebacteria contain only one type of RNA polymerase while eukaryotes contain three types, referred to as RNA polymerases I (or A), II (or B), and III (or C). *Escherichia coli* RNA polymerase contains four different subunits of 155 kDa (β'), 151 kDa (β), 37 kDa (α), and 6 kDa (ω). Only the first three subunits are required for polymerase activity and are considered as the core enzyme. Two copies of the α subunit are present in the core enzyme, resulting in a β' β $(\alpha)_2$ subunit structure. Subunits related to the *E.coli* β', β, and α subunits are found in most multimeric RNA polymerases, including the eukaryotic RNA polymerases I, II, and III. The β' and β subunits share at least eight and nine blocks of amino acid sequence similarity with their counterparts in archaeal and eukaryotic nuclear RNA polymerases (Thuriaux and Sentenac, 1992). The α subunit shares two blocks of limited amino acid sequence similarity with two different subunits found in archaeal RNA polymerase and eukaryotic RNA polymerases I, II, and III. The bacterial subunit β' is homologous to subunit A in archaeal RNA polymerases, and to subunit RPA1, RPB1 and RPC1 in eukaryotes. The bacterial subunit β is homologous to subunit B in archaeal RNA polymerases, and to subunit RPA2, RPB2 and RPC2 in eukaryotes. The bacterial subunit α^{I} is homologous to subunit D in archaeal RNA polymerases, to subunit RPA5 and RPC5 in eukaryotic RNA polymerases I and III, and to subunit RPB3 in eukaryotic RNA polymerase II. The bacterial subunit α^{II} is homologous to subunit L in archaeal RNA polymerases, to subunit RPA9 and RPC9 in eukaryotic RNA polymerases I and III, and to the subunit RPB11 in eukaryotic RNA polymerase II. The bacterial subunit ω is homologous to subunit K in archaeal RNA polymerase, and to subunit RPA6, RPB6, and RPC6 in eukaryotic RNA polymerases I, II and III. Bacterial RNA polymerase contains only these conserved subunits. Archaeal and eukaryotic RNA polymerase contain these conserved subunits as well as additional subunits (reviewed by Ebright, 2000).

In eukaryotes, RNA polymerase I makes pre-rRNA, RNA polymerase II synthesizes all messenger RNA and some snRNA, and RNA polymerase III makes pre-tRNA, pre-5sRNA, other small RNA and some snRNA. RNA polymerase II is the central enzyme of gene expression. The intricate regulation of RNA polymerase II transcription underlies cell growth and differentiation. The best characterized form of the enzyme, that from the yeast *Saccharomyces cerevisiae*, comprises 12 different polypeptides, with a total mass of about 0.5 megadaltons (MD). The human enzyme is very similar to the yeast enzyme. Not only do human genes show a high degree of sequence conservation with those of yeast, but at least 10 human subunits of RNA Polymerase II genes can be substituted for their counterparts in yeast (Woychik, 1998). Plant RNA polymerase II, like RNA polymerase II in animals and lower eukaryotes, contains four subunits related to the core RNA polymerase of prokaryotic organisms. In addition to these four conserved subunits, plant RNA polymerase II contains eight additional polypeptides that form a core enzyme. RNA polymerase II is the core of the transcription machinery. On its own, it can unwind the DNA double helix, polymerize RNA, and proofread the nascent transcript. In the presence of additional proteins, it assembles even larger initiation and elongation complexes, capable of promoter recognition and response to regulatory signals. A regulated initiation complex comprises RNA polymerase II, five general transcription factors, and a multiprotein mediator (Conaway and Conaway, 1997). This mediator contains some 60 proteins, with a total mass of 3.5 MD. In transcription elongation complexes, the mediator and some of the general transcription factors are replaced by the SII (TFIIS), elongator, other elongation factors, and RNA processing proteins (Otero *et al.*, 1999).

The largest subunit of RNA polymerase II (*rpb1*)

The largest subunit of RNA polymerase has eight highly conserved domains, termed domains A through H. These domains were defined based on their similarity to bacterial and archaeal homologs (Figure 2; Hall, 2001). Nearly 40% of the amino acid residues of the largest subunit of RNA polymerase II (RPB1) in eukaryotes are identical and a large fraction of the remaining residues are conserved (Young, 1991). The RPB1 has one more

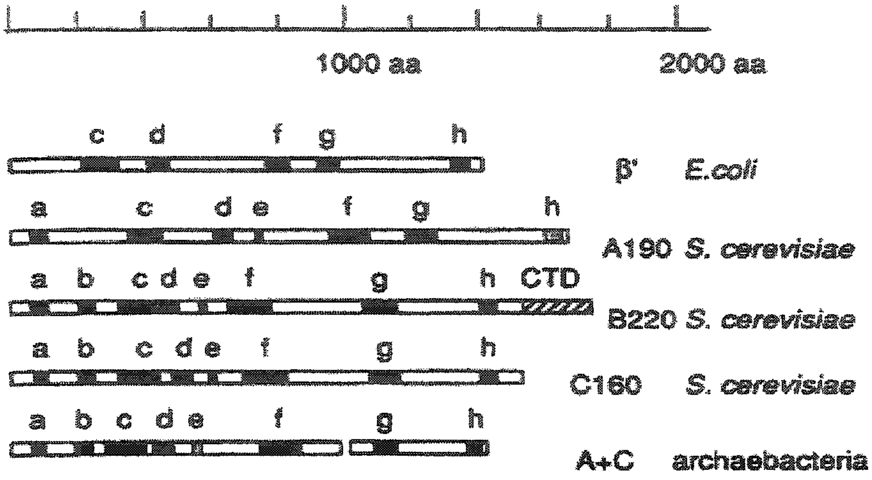


Figure 2. The conserved domains in the largest subunits of RNA polymerase. A190 is the product of the gene *rpa1*. B220 is the product of the gene *rpb1*. C160 is the product of the gene *rpc1*. CTD domain is unique to *rpb1*. Note the conserved regions are marked by heavy black lines (Taken from Hall, 2001).

unique conserved C-terminal (CTD) domain, called I, which is not shared with its prokaryotic homlogues. This domain contains different numbers of repeats of the peptides Try-Ser-Pro-Thr-Ser-Pro-Ser in various species: 26 in yeast, 44 in Drosophila and 54 in mouse (Nawrath *et al.*, 1990). This domain is essential for cell viability, involved in initiation at promoters and is highly phosphorylated in a substantial portion of the RNA polymerase II molecules in the cell (Young, 1991). Domain A and domain C are thought to play a role in DNA template binding and nascent RNA chain binding, respectively. Domain A has zinc-binding motifs and conditional mutations in this domain primarily affect the assembly or the structure of the enzyme (Treich *et al.*, 1991). Domain C makes contact with the heteroduplex product of RNA synthesis (Jokerst *et al.*, 1989). Domain D is believed to be part of the DNA binding site (Nawrath *et al.*, 1990). Domain F and domain G represent the central and most conserved part of the largest subunits of eukaryotic RNA polymerase II. Domain F may be involved in chain elongation, since this domain contains a α-amanitin-binding site that specifically blocks elongation (Nawrath *et al.*, 1990).

Previous analyses suggested that the gene duplications that gave rise to the RNA polymerase family occurred before the divergence of the major eukaryotic groups (Iwabe *et al.*, 1991). RNA polymerase genes, especially the genes encoding the two largest subunits of RNA polymerase I, II, and III, have been used to elucidate many phylogenetic questions. These genes were first introduced as a phylogenetic marker by Zillig and colleagues (Zillig *et al.*, 1989; Pühler *et al.*, 1989). They have proven valuable for evolutionary research at the broadest taxonomic levels, including molecular phylogenetic studies of protists (Stiller and Hall, 1997,1998; Hall, 2001), ascomycetes (Liu *et al.*, 1999), basidiomycetes (Matheny *et al.* 2002), microsporidians (Hirt *et al.,* 1999), metazoans (Sidow and Thomas, 1994), plants (Denton *et al.*, 1998, Oxelman and Bremer, 2000) and arthropods (Shultz and Regier, 2000). The recent results of our lab have also shown that RNA polymerase II is an excellent phylogenetic marker to study the phylogeny of land plants (Nickerson and Drouin, in press).

As phylogenetic markers, *rpb1* genes have favorable attributes (Hall, 2001; Nickerson and Drouin, in press). First, they are ubiquitous, so it is possible to obtain the orthologous genes from different species. Second, they are single copy genes in most species (although paralogous copies have been found in several plants and trypanosomes). Third, they are conveniently amplified from conserved PCR primers. Fourth, they are large enough to provide extensive sequence data. Fifth, the sequences are easily alignable. An *rpb1* gene is about 4000 nucleotides long and the region between domain A and domain G is about 3000 nucleotides long.

This study aimed at studying three uncertain evolutionary relationships among seed plants: 1) the position of gnetales, 2) the relationships among eumagnoliids, and 3) the relationships among recent angiosperms (Caryophyllales, asterids and rosids). Maximum likelihood and maximum parsimony analyses were performed using *rpb1* genes from 26 species either by themselves or combined with *atpB*, *rbcL* and 18S rDNA gene sequences.

Materials and Methods

Plant Material:

Twenty-six plant species were analyzed in this study (Table 1). I used 8 gymnosperm species, 17 angiosperm species and *Psilotum*, a fern ally, which was the outgroup. The *rpb1* sequences for 8 species were generated from former studies in our lab (Nickerson and Drouin, in press) and the *rpb1* sequences for *Arabidopsis* and *Oryza* were obtained from GenBank. The other 16 new *rpb1* sequences were generated following the methods described below. *Ephedra* was collected from the Carleton University greenhouse. *Papaver* was collected from Agriculture Canada (Ottawa). *Beta* was provided by Jennifer Nickerson. The leaves of *Drimys* were provided by Yin-Long Qiu (University of Massachusetts). The leaves of *Illicium*, *Saruma* and *Liriodendron* were from Richard Olmstead (Washington University), Alan Whittemore (US National Arboretum) and Chris Graham (Royal Botanical Gardens, Hamilton, Ontario), respectively. *Thuja* and *Taxus* were collected in Gatineau with the help of Lynn Gillespie (Canadian Museum of Nature, Gatineau). *Ginkgo* was collected on Osgoode street near the University of Ottawa. The remaining plant species were bought from local stores, and then planted in the greenhouse of the University of Ottawa. All tissue samples were stored at -80°C until used.

The sequences of *atpB*, *rbcL* and 18S rDNA gene were retrieved from GenBank. In this study, I attempted to use four kinds of genes from the same species. However, for some species, the sequences were not available. In these cases, the sequences from the

Table 1 The classification of twenty-six species used in this study based on the NCBI taxonomy database.

Species	**Common name**	**Order and Family**	**Corresponding clade in the supertree**
Amborella trichopoda	*Amborella*	Amborellales; Amborellaceae	*Amborella*
Nymphaea odorata	Water lily	Nymphaeales; Nymphaeaceae	Nymphaeales
*Illicium parviflorum**	*Illicum*	Illiciales; Illiciaceae	ITA
*Ceratophyllum demersum**	*Ceratophyllum*	Ceratophyllales; Ceratophyllaceae	*Ceratophyllum*
*Asparagus spp**	*Asparagus*	Asparagales; Asparagaceae	Monocots
Oryza sativa	Rice	Poales; Poaceae	Monocots
Zea mays	Maize	Poales; Poaceae	Monocots
Magnolia soulangeana	*Magnolia*	Magnoliales; Magnoliaceae	Magnoliales
*Liriodendron tulipifera**	*Liriodendron*	Magnoliales; Magnoliaceae	Magnoliales
*Persea Americana**	Avocado	Laurales; Lauraceae	Laurales
*Drimys winteri**	*Drimys*	Winterales; Winteraceae	Winterales
*Saruma henryi**	*Saruma*	Piperales; Aristolochiaceae	Piperales
*Papaver orientalis**	*Papaver*	Ranunculales; Papaveraceae	Ranunculales
*Beta vulgaris**	Beet	Caryophyllales; Amaranthaceae	Caryophyllales
*Nicotiana tabacum**	*Nicotiana*	Solanales; Solanaceae	Asterids
*Pisum sativum**	Pea	Fabales; Fabaceae	Rosids
Arabidopsis thaliana	*Arabidopsis*	Brassicales; Brassicaceae	Rosids

Table 1 The classification of twenty-six species used in this study based on the NCBI taxonomy database.

Species	**Common name**	**Order and Family**	**Corresponding clade in the supertree**
*Thuja occidentalis**	White Cedar	Coniferales; Cupressaceae	Other conifer
*Taxus canadensis**	*Taxus*	Coniferales; Taxaceae	Other conifer
*Podocarpus macrophyllus**	*Podocarpus*	Coniferales; Podocarpaceae	Other conifer
Pinus nigra	Pine	Coniferales; Pinaceae	Pinaceae
*Ephedra viridis**	*Ephedra*	Ephedrales; Ephedraceae	Gnetales
Welwitschia mirabilis	*Welwitschia*	Welwitschiales; Welwitschiaceae	Gnetales
*Gingko biloba**	*Gingko*	Ginkgoales; Gingkoaceae	*Gingko*
Zamia muricata	*Zamia*	Cycadales; Zamiaceae	Cycads
Psilotum nudum	Whisk fern	Psilotales; Psilotaceae	outgroup

(continued).

Note: * indicates the *rpb1* genes sequenced in this study.

species in the same genera, or family were used. Except for *Beta*, the 18S rDNA, *atpB* and *rbcL* sequences data came from species in the same order. The a*tpB* gene of *Welwitschia mirabilis* was replaced by that of *Gnetum gnemon*. GenBank accession numbers of the *rpb1*, 18S rDNA, *atpB* and *rbcL* sequences are listed in Table 2.

RNA Isolation:

Three methods were employed to isolate RNA from leaf or needle tissues (Table 3). In order to minimize RNAs contamination, all procedures were performed while wearing gloves and tips and tubes were autoclaved and baked (200°C for 3hrs). The Qiagen RNeasy Plant Mini Kit was used to extract RNA from 6 plant species, *Liriodendron, Asparagus, Papaver, Beta, Nicotiana, Pisum*, according to the protocol provided by the manufacturer (RNeasy mini handbook, third edition June 2001, p 76-p 77). 100 mg tissue was processed and buffer RLT, which contained guanidine isothiocyanate (GITC), was chosen as the lysis buffer due to the greater cell disruption and denaturation properties of GITC. β-Mercaptoethanol (β-ME) was added to buffer RLT before use. The RNA was DNase treated according to the Qiagen RNase-Free DNase Set protocol (instead of continuing with the buffer RW1 step; RNeasy mini handbook, third edition June 2001, p 99-p 100). The elution step was performed twice to elute the RNA in 30μl RNase-free water.

RNAs of five other species (*Drimys*, *Persea*, *Saruma*, *Ceratophyllum*, and *Gingko*), from which the Qiagen RNeasy Plant Mini Kit failed to extract RNAs, were

Table 2 GenBank accession numbers of the *rpb1*, 18S rDNA, *atpB* and *rbcL* sequences

Species	***rpb1***	**18S**	***atpB***	***rbcL***
Amborella trichopoda	AF519541	U42497	AJ235389	L12628
Nymphaea odorata	AF519540	AF096696 *Nymphaea sp.*^{G}	AJ235544	M77034
Illicium parviflorum	this study	L75832	U86385	L12652
Ceratophyllum demersum	this study	U42517	AJ235430	D89473
Asparagus spp	this study	AF069205 *Asparagus falcatus*^{G}	AJ235400 *Asparagus officinalis*^{G}	AB029849 *Asparagus cochinchinensis*^{G}
Oryza sativa	AC084218	X00755	X15901	X15901
Zea mays	AF519538	U42796	X86563	X86563
Magnolia soulangeana	AF519539	AF206956	AJ235526	AF206791
Liriodendron tulipifera	this study	AF206954	AJ235522	X54346
Persea americanum	this study	U52031 *Sassafras albidum*^{F}	AF209668 *Sassafras albidum*^{F}	X54347
Drimys winteri	this study	U42823	AF093425	L01905
Saruma henryi	this study	AF207013	AJ235595	L12664
Papaver orientalis	this study	L75836 *Hypecoum imberbe*^{F}	U86394	L08764
Beta vulgaris	this study	AF207027 *Stellaria media*^{O}	AF209680 *Stellaria media*^{O}	AF206823 *Stellaria media*^{O}
Nicotiana tabacum	this study	AJ236016	NC001879	Z00044

Table 2 GenBank accession numbers of the *rpb1*, 18S rDNA, *atpB* and *rbcL* sequences (Continued)

Species	***rpb1***	**18S**	***atpB***	***rbcL***
Pisum sativum	this study	U43011	X03852	X03853
Arabidopsis thaliana	AL031986	X16077	AP000423	AP000423
Thuja occidentalis	this study	D38243 *Juniperus chinensis*^F	AJ235534 *Metasequoia glyptostroboides* ^F	AF127428 *Thuja plicata*^G
Taxus canadensis	this study	D16445 *Taxus mairei*^G	AJ235619 *Taxus baccata*^G	L12580 *Amentotaxus argotaenia*^F
Podocarpus macrophyllus	this study	D38473 *Podocarpus costalis*^G	AJ235567 *Podocarpus milanjianus*^G	AF307931 *Podocarpus totara*^G
Pinus nigra	AF519536	D38245 *Pinus elliottii*^G	D17510	D17510
Ephedra viridis	this study	U42492 *Ephedra californica*^G	AY056533 *Ephedra californica*^G	AY056569 *Ephedra californica*^G
Welwitschia mirabilis	AF519537	AF207059	AF187060 *Gnetum gnemon*^C	AJ235814
Gingko biloba	this study	D16448	AJ235481	D10733
Zamia muricata	AF519534	M20017	AF188845 *Zamia furfuracea* ^G	L12683 *Zamia inermis* ^G
Psilotum nudum	AF519542	X81963	AP004638	AP004638

Note: G indicates that it is in the same genus; F indicates that it is in the same family; O indicates that it is in the same order; C indicates that it is in the same class.

Table 3. The A to G regions of *rpb1* for sixteen species sequenced in this study. Sequence length, RNA isolation methods and number of clones for species are listed. Three methods were employed to isolate RNA from these 16 species. The *rpb1* sequences of *Drimys* and *Pisum* were PCR amplified as one single fragment from domain A to domain G. For all the other 14 species, the *rpb1* sequences were PCR amplified as two overlapping fragments, AD and DG.

Species	Length (bases)	RNA isolation method	Clone number		
			AG fragment	AD fragment	DG fragment
Illicium	3039	Modified method	--------	3	4
Drimys	3030	Bahloul and Burkard protocol	3	--------	--------
Liriodendron	3030	Qiagen kit	--------	4	3
Persea	3027	Bahloul and Burkard protocol	--------	2	2
Saruma	3027	Bahloul and Burkard protocol	--------	3	4
Ceratophyllum	3030	Bahloul and Burkard protocol	--------	2	3
Asparagus	3027	Qiagen kit	--------	5	1
Papaver	3030	Qiagen kit	--------	3	3
Beta	3036	Qiagen kit	--------	4	3
Nicotiana	3030	Qiagen kit	--------	4	3
Pisum	3030	Qiagen kit	3	--------	--------
Gingko	3039	Bahloul and Burkard protocol	--------	4	4
Taxus	3039	Modified method	--------	4	3
Ephedra	3039	Modified method	--------	4	4
Podocarpus	3039	Modified method	--------	4	2
Thuja	3039	Modified method	--------	4	2

successfully isolated following the Bahloul and Burkard (1993) protocol. In this protocol, the extraction buffer consists of 100 mM sodium acetate, 50 mM EDTA (pH 8.0), 500 mM sodium chloride, 2% PVP (MW 20,000). The pH of this buffer was adjusted to 5.5 and SDS was then added to 1.4%. Cysteine powder was also added to 10 mM just before use. 10mL extraction buffer was used to extract RNA from 1 g tissue.

For the remaining 5 species (*Illicium, Taxus, Ephedra, Podocarpus,* and *Thuja),* the Qiagen RNeasy Plant Mini Kit and the Bahloul and Burkard (1993) protocol did not allow to obtain RNAs of good quality and yield. We therefore modified the Bahloul and Burkard (1993) protocol. The main modification was that the ethanol precipitation of the RNA was replaced by a "pre-precipitation" of the contaminants (polysaccharides and secondary metabolites) using a dilute ethanol solution and was followed by an isopropanol precipitation of the RNA. This modified method is as follows:

1) 1 gram of frozen tissue was ground in a mortar and pestle to a fine powder in liquid nitrogen. The powder was transferred to 10 ml of extraction buffer and homogenized gently. It was then incubated in a water bath at 65 °C for 10 min with occasional shaking. The extraction buffer was the same as that in the Bahloul and Burkard (1993) protocol.

2) The supernatant was collected by centrifugation at 10,000 g at 4 °C for 10 min, 1/3 volume of 5M potassium acetate (pH 4.8) was added, and this solution was mixed gently and incubated on ice for 30 min.

3) The solution was centrifuged twice at 10,000 g at 4 °C for 10 minutes and the supernatant was collected.

4) 1/30 volume 3M sodium acetate was added followed by the addition of ethanol to a final concentration of 10% v/v. This solution was incubated on ice for 10 minutes.

5) The supernatant was collected by centrifugation at 5,000 g at 4 °C for 20 minutes, to which 1/9 v/v 3M sodium acetate and isopropanol to a final concentration of 33% v/v were added. After incubation at -20 °C for 2 hours, centrifugation at 10,000 g for 30 minutes at 4 °C was performed.

6) The pellet was dissolved in 400 µl water, and 400 µl 4M lithium chloride was then added to precipitate the RNA overnight at 4°C.

7) The RNA pellet was collected by centrifugation at 10,000 g at 4°C for 30 min and washed with 0.5 ml 70% ethanol, air dried until all ethanol was evaporated, and re-dissolved in 50 µL of RNase-free water.

RNAs which were isolated using the Bahloul and Burkard (1993) protocol and the modified method were DNase treated by the Ambion DNA-free kit according to the directions of the manufacturer. 1 unit of DNAse was used to each RNA sample at 37 °C for 30 min. The RNA purity and yield were measured by spectrophotometry and gel electrophoresis.

cDNA synthesis:

DNase treated RNA samples were heated at 70 °C for 10 minutes to destroy possible secondary structure, and then placed on ice for 5 minutes. The single-stranded cDNA was reverse transcribed from total RNA by using the Boehringer Mannheim's 1st Strand cDNA Synthesis Kit for RT-PCR (AMV) following the instructions of the manufacturer. The random primer p(dN)6, rather than the Oligo-p(dT)15 primer, was used because former studies in our lab showed that the former primer could always generate the expected PCR fragments. Each reaction was in a 20 µl volume and the amount of RNA was less than 1 µg. At the end of the reactions, all cDNA samples from the same species were pooled together. Negative controls without AMV reverse transcriptase were used to confirm that RNA samples were not contaminated with DNA.

PCR amplification of *rpb1* gene:

Four degenerate primers were used to amplify the AG, AD and DG fragments of *rpb1* genes. Primer AD and Primer GF were used for amplifying AG fragments of approximately 3.1 kb. Primers AD and DA were used for amplifying AD fragments of approximately 1.3 kb. Primers D and GF were used for amplifying DG fragments of approximately 1.8 kb. Primer AD was designed to the ECPGHFG motif in domain A (5' GAI TGY CCI GGI CAY TTY GG 3', forward). Primer DA was designed to the MMVPKCIV motif in domain D (5' AC IAT RCA YTT IGG IAC CAT CAT 3', reverse). Primer D was designed to the PYNADFDGDEMN motif also in domain D (5' CCI TAY AAY GCI GAY ITY GAY GGI GAY GAR ATG AA 3', forward). Primer GF was designed to the GEPATQMT motif in domain G (5' GT CAT YTG IGT IGC IGG YTC TCC 3', reverse).

Boehringer Mannheim's Expand High Fidelity PCR System was used to amplify AG fragments from the cDNA samples for *Drimys* and *Pisum* following the manufacturer's instructions. This system is designed to amplify most efficiently DNA fragments up to 5 kb with high yield, high fidelity and high specificity because of its powerful polymerase mixture of Taq DNA polymerase and Pwo DNA polymerase (which has 3'-5' exonuclease activity). The reaction volume was 50 µl and the concentrations of the various reagents were added according to the recommendation of the manufacturer. The PCR reaction was run under the following condition: 1 cycle of denaturation at 94°C for 2 min, 10 cycles of denaturation at 94°C for 15 sec, annealing at 45°C for 30 sec and elongation at 68°C for 3 min, 20 cycles of denaturation at 94°C for 15 sec, annealing at 50°C for 30 sec and elongation at 68°C for 3 min with a cycle elongation of 5 sec for each cycle, and 1 cycle of final elongation at 72°C for 7 min.

Taq polymerase (Amersham Pharmacia Biotech) was used to amplify the AD and DG fragments from the remaining species because Boehringer Mannheim's Expand High Fidelity PCR System failed to generate AG fragments. The PCR experiment was composed of a no template negative control, a no AMV cDNA negative control with template, and three reactions with an $MgCl_2$ concentration of 1.0 mM, 1.5 mM and 2 mM, respectively. For negative controls, 1.5 mM $MgCl_2$ was used. In addition to $MgCl_2$, 5 µl 10X buffer without $MgCl_2$, 4 µl (10 mM) dNTPs, 1 µl (100 µM) upstream primer, 1 µl (100 µM) downstream primer, 0.5 µl (5000 U/ml) Taq polymerase, 10 µl cDNA, and sdH_2O were added up to a final volume of 50 µl. For AD fragments, primers AD and DA were used. For DG fragments, primers D and GF were used. The PCR reaction was run under the following condition: 1 cycle of denaturation at 94°C for 5 min, 35 cycles of denaturation at 94°C for 1 min, annealing at 55°C for 1 min and elongation at 72°C for 3 min, and 1 cycle of final elongation at 72°C for 10 min. The PCR products were observed by electrophoresis in 0.7% agarose gels with 1X TBE buffer.

Mehrdad Hajibabaei, a Ph. D. student in our lab, kindly provided *Taxus* clones of DG fragments.

Gel fragment DNA purification:

The PCR products were run on 0.7% agarose gels in 1X TAE. The desired fragments were cut and purified using the UltraClean DNA Purification Kit (MOBIO Laboratories) following the protocol provided by the manufacturer. For each fragment, 10 μl of ULTRA BIND was used to bind the DNA which was eluted in 20 μl of water.

TA Cloning and plasmid purification:

The purified gel DNA fragments were cloned using Invitrogen's TOPO TA Cloning kit (vector pCR 2.1). For AG fragments amplified with the Expand High Fidelity PCR System, 3'A overhangs had to be added to each of the gel purified fragments before cloning because these PCR products were a mixture of products with 3'A overhangs and blunt ends. According to the direction of Invitrogen's TOPO TA Cloning instruction manual, 2 μl of 10X PCR buffer, 0.5 μl (10 μmol) dATP and 0.5 μl (5000 U/mL) Taq polymerase were added to 17 μl of gel-purified DNA fragments. The reaction was performed at 72°C for 10 minutes.

The TOPO cloning was performed following the instruction of the manufacturer. The TOPO cloning reaction was composed of 4 µl PCR product, 1 µl salt solution and 1µl vector. 4 µl of this reaction was transformed into competent cells. The cell culture was plated on Luria-Bertani (LB) agar plates containing 50 µg/ml ampicillin and 40 µl of 40 mg/ml X-gal, and incubated at 37 °C overnight.

The white colonies were picked and PCR colony screened. For AG fragments, the Accurase enzyme kit was used because it works better than Taq polymerase with longer fragments. In 12.5 µl PCR reactions, 1.25 µl 10X PCR buffer, 0.5 µl 5mM dNTPs, 2.5µl 5 µM primer AD, 2.5µl 5 µM primer GF, 0.75 µl 25mM $Mg(OAC)_2$, 0.0625 µl (1.25 U) Accurase enzyme and 4.94 µl sterile water were added. The PCR reactions were performed under the following conditions: 1 cycle of denaturation at 94°C for 2 min, 10 cycles of denaturation at 94°C for 20 sec, annealing at 50°C for 30 sec and elongation at for 68°C for 2 min, followed by 20 cycles of denaturation at 94°C for 20 sec, annealing at 53°C for 30 sec and elongation at for 68°C for 2 min, and 1 cycle of final elongation at 68°C for 7min. For AD and DG fragments, Taq polymerase was used. In 12.5 µl PCR reactions, 1.25 µl 10X PCR buffer with NH_4SO_4, 0.5 µl 5mM dNTPs, 2.5µl 5 µM primer AD (or D), 2.5µl 5 µM primer DA (or GF), 1 µl 25mM $MgCl_2$, 0.1 µl (5000 U/ml) Taq polymerase and 4.65 µl sterile water were added. The PCR condition was the same as the Accurase reaction except that the elongation temperature was changed to 72°C.

The positive colonies verified by PCR colony screens were then cultured overnight in 5 ml of LB containing ampicillin at 37°C. In 5 ml overnight culture, 1.3 ml of culture was added into 200 µl of 75% glycerol for long-term storage. The remaining 3.7 ml culture were used to purify plasmid DNA using the Qiagen QIAprep Spin Miniprep Kit according to the directions of the manufacturer. EcoR1 restriction digests and agarose gel electrophoresis were used to identify the sizes of the clones.

ABI 310 automated Sequencing:

The ABI Prism Version 3.0 Big Dye Terminator cycle sequencing ready reaction kit (Perkin-Elmer Applied Biosystems) was used for sequencing. In each reaction, 1 µl 5 µM primer, 3 µl of terminator, 0.5 µg plasmid and sterile water were added to 10 µl. For cycle sequencing, 25 cycles of the following conditions were used: denaturation at 96°C for 10 sec, annealing at 50°C for 5 sec and elongation at 60°C for 4 min. Individual cycle sequencing products were cleaned to remove the excess dye terminators using MultiScreen 96-well filtration plates (Milllipore) according to the instructions of the manufacturer. The purified products were dried out and resuspended in 15 µl of template suppression reagent and then run on an ABI 310 Genetic Analyzer. M13 forward primer and reverse primer were used as initial sequencing primers. In order to obtain the full-length sequence of both DNA strands, internal primers were designed using the Primer 3 software (http://www-genome.wi.mit.edu/cgi-bin/primer3_www.cgi). The melting temperatures (Tm) of the internal sequencing primers varied from 59°C to 63°C and GC contents varied from 45% to 55%.

Sequence analyses:

All putative *rpb1* sequences were confirmed using BLASTX searches on the NCBI (the National Center for Biotechnology Information) website (http://www.ncbi.nlm.nih.gov/BLAST/). The *rpb1* sequences were assembled into contigs and consensus sequences were exported after the sequences corresponding to the primers used in the PCR amplification were deleted at the beginning and the end of each *rpb1* sequence using the Sequencher (version 4.0.5) software. The DNA sequences were aligned using ClustalW (Thompson et al. 1994) and default parameters. Sequences were then translated into proteins using the SequEdit sequence editor (SequEdit Version 0.912, Drouin *et al.*, 1999). The DNA and protein sequences were exported in FASTA, PHYLIP and MEGA format. The first and second codon positions data matrix and third codon positions data matrix were created using MEGA version 2.0 (Kumar et al. 2001). The DNA and protein sequences of *atpB*, *rbcL* were extracted from GenBank (Table 2). The protein sequences were aligned using ClustalW and the DNA sequences were aligned using the protein alignment as a guide with the program align2aa.pl. The 18S rDNA sequences were retrieved from GenBank and aligned using ClustalW.

The GC content for *rpb1* genes from each of the 26 species was calculated by the CODONS version 1.4 program (Lloyd & Sharp, 1992). The RRTree program version 1.1.7 (Robinson *et al.*, 1998) was used to examine if the rate of nonsynonymous nucleotide substitutions per site evolved at significantly different rates between lineages. Evolutionary distances were calculated using the MEGA2 program (Kumar et al. 2001). We used the method of Pamilo-Bianchi- Li (Pamilo and Bianchi 1993) to calculate the number of synonymous and non-synonymous substitutions in *rpb1, atpB, and rbcL* genes.

In reconstructing the phylogenetic trees, the protein and DNA sequences with different codon positions were analyzed using TREE-PUZZLE 5.0 (Strimmer and Von Haeseler, 1996) and PAUP 4.0b 10 (Swofford, 1998). *Psilotum* was used as the outgroup in all phylogenetic analyses. TREE-PUZZLE 5.0 was used to reconstruct phylogenetic trees by maximum likelihood. For maximum likelihood analysis of DNA sequences, the HKY substitution model and a mixed rate heterogeneity model with 1 invariable rate and 8 gamma distributed rates were used. For protein sequences, the JTT substitution model and a mixed rate heterogeneity model with 1 invariable rate and 8 gamma distributed rates were used. PAUP 4.0 was used to perform the maximum parsimony analyses. The most parsimonious trees were obtained using heuristic searches, 1000 replicates of RANDOM taxon additions, equal weights and tree bisection-reconnection (TBR) branch swapping. Bootstrap values were also obtained using this program. The phylogenetic trees were edited using TreeView version 1.2 (Page, 1996).

Results

Alignment, GC-content and evolutionary rates of *rpb1* sequences

Sixteen *rpb1* genes were sequenced in this study. The protein sequence alignment (Appendix A) consisted of 1017 amino acids and the nucleotide sequence alignment (Appendix B) consisted of 3051 nucleotides. The length of the A-G region was well conserved: there was only a 12 bp difference between the shortest and longest sequences (Table 3). The *rpb1* sequences were PCR amplified as two overlapping fragments, AD and DG, in fourteen species except for *Drimys* and *Pisum* whose *rpb1* sequences were PCR amplified as a complete AG fragment. There was an overlap of 48 nucleotides (16 amino acids) between primer DA and primer D. The two overlapping regions were identical in eight species. In the sequences where differences were observed in the overlap region, most nucleotide differences occurred in third codon positions and only 4 differences were in the first codon positions (Table 4). These nucleotide differences might represent allelic variation.

The GC-content of our 16 *rpb1* sequences varied from 41% to 44%. The GC-content at the first, second and third codon position varied from 51% to 53%, from 35% to 38%, and from 38% to 43%, respectively. These GC-content are similar to those observed in the *rpb1* sequences analyzed by Nickerson and Drouin (in press).

Relative rate tests were performed to determine if there was significant evolutionary rate difference between angiosperm and gymnosperm sequences. All

Table 4. Nucleotide differences between the overlapping regions of the sixteen *rpb1* genes sequenced in this study. The *rpb1* sequences of *Drimys* and *Pisum* were PCR amplified as a single fragment from domain A to domain G. The other 14 *rpb1* sequences were PCR amplified as two overlapping fragments, AD and DG. This overlapping region was 48 nucleotides long.

Species	total nucleotide differences	first codon position	second codon position	third codon position
Illicium	0	0	0	0
Drimys	n.a.	n.a	n.a.	n.a.
Liriodendron	0	0	0	0
Persea	0	0	0	0
Saruma	0	0	0	0
Ceratophyllum	6	0	0	6
Asparagus	0	0	0	0
Papaver	2	0	0	2
Beta	7	2	0	5
Nicotiana	0	0	0	0
Pisum	n.a.	n.a.	n.a.	n.a.
Gingko	0	0	0	0
Taxus	1	0	0	1
Ephedra	11	2	0	9
Podocarpus	0	0	0	0
Thuja	3	0	0	3

angiosperm species and all gymnosperm species were combined into two separate lineages. The relative rates of nonsynonymous substitutions between two lineages in the *rpb1*, *atpB*, *rbcL* sequences were not statistically different (dKa = 0.0019 ± 0.0035, $p = 0.5918$; dKa = 0.0034 ± 0.0042, $p = 0.4261$; dKa = 0.0042 ± 0.0029, $p = 0.1506$, respectively). The relative rates of synonymous substitutions between two lineages in the *atpB*, and *rbcL* genes were also not statistically different (dKs = -0.0965 ± 0.0585, $p = 0.099$; dKs = 0.0887 ± 0.0522, $p = 0.0897$, respectively). Relative rates of the synonymous substitutions of *rpb1* sequences could not be calculated because these substitutions were saturated. The relative rates of substitutions of 18S sequences of these two lineages was also not significant (dK = 0.0098 ± 0.0050, $p = 0.0508$).

The average number of substitutions in *rpb1, atpB*, and *rbcL* genes was also calculated. For synonymous substitution, *rpb1* showed the highest number (1.408 ± 0.038), followed by *atpB* (0.478±0.030) and *rbcL* (0.463±0.025). For non-synonymous substitutions, *rpb1* still showed the highest number (0.067 ± 0.004), followed by *rbcL* (0.026±0.004) and *atpB* (0.022±0.004).

Phylogenetic analyses of *rpb1* sequences:

Twenty-six plant species were used in all phylogenetic analyses, including 8 gymnosperms, 17 angiosperms and *Psilotum*, a fernally, as the outgroup. In the results and discussion presented here, clades with bootstrap values or support values of 70% or more are regarded as "well supported" (Hillis and Bull, 1993).

The maximum parsimony PAUP analyses based on the *rpb1* amino acid sequences and DNA sequences showed different topologies (Figure 3). The parsimony analysis based on the amino acid sequences of *rpb1* genes from regions A to G (1017 amino acids,) shows distinct angiosperm and gymnosperm clades and *Amborella* is the sister group to all other angiosperms with 88% bootstrap support and followed by *Nymphaea* (82%) and *Illicium* (67%) (Figure 3, left). The gymnosperm clade is also well supported with 89% bootstrap support and conifers and gnetales form a clade sister to the *Zamia* and *Gingko* clade. The parsimony analysis based on DNA sequences of *rpb1* genes (3051 nucleotides) is unable to resolve the angiosperm clade (Figure 3, right). It places *Nymphaea* out of the angiosperm clade and at the bottom of the gymnosperms and remaining angiosperms. For the gymnosperm clade, this tree puts the *Zamia* and *Ginkgo* clade as the sister group to conifers, and they form a sister group to gnetales.

Maximum likelihood TREE-PUZZLE analyses were also performed using amino acid sequences and DNA sequences (Figure 4). The PUZZLE tree and the PAUP tree based on analyses of *rpb1* amino acid sequences share the same topology. The maximum likelihood analysis based on DNA sequences shows a different topology from the corresponding maximum parsimony analysis. Although it shows distinct angiosperm and gymnosperm clades, it does not resolve the position of most angiosperms except for *Amborella*. For the gymnosperm clade, this tree shows the *Zamia* and *Ginkgo* clade as being the sister group of conifers, and this group is a sister group to gnetales.

Different codon positions of the *rpb1* DNA sequence data set were analyzed to determine which codon position is more useful. The maximum parsimony tree resulting from the PAUP analysis based on the first and second codon position (2034 nucleotides, Figure 5, left) shows a topology similar to that observed with the parsimony analysis of *rpb1* protein sequences (Figure

3, left). The tree based on third codon positions (1017 nucleotides) shows a phylogeny where angiosperms are paraphyletic and where gymnosperms were found inside one of these paraphyletic groups.

The maximum likelihood TREE-PUZZLE tree based on the first and second codon position has the same topology as the corresponding *rpb1* protein phylogenetic analyses except for the gymnosperms. It groups the gnetales and *Pinus* as a clade with weak support (55%), and this gnepine clade is sister to the other conifers (Figure 6, left). The maximum likelihood analyses based on third codon positions shows a better phylogeny than the maximum parsimony analysis based on the same data set. Although it does not resolve most angiosperm branches, angiosperms and gymnosperms are in two distinct sister clades. The gymnosperm clade is well-resolved with gnetales as a sister to the *Zamia/Gingko* and conifers clade (Figure 6, right).

Phylogenetic analyses of *rpb1, atpB,* and *rbcL* sequences:

In order to obtain a better resolved phylogeny, two chloroplast gene sequences (*atpB* and *rbcL*) were combined with *rpb1* sequences to perform phylogenetic analyses. These two genes are extensively used in angiosperm phylogeny. The *atpB* gene codes for the beta subunit of the ATP synthase, while the *rbcL* gene codes for the large subunit of the ribulose-1,5-bisphosphate-carboxylase/oxygenase. The parsimony tree based on the amino acid sequences of *rpb1/atpB/rbcL* (2005 amino acids) shows two distinct clades of angiosperm and gymnosperm with strong bootstrap supports (100% and 98%, respectively; Figure 7, left). *Amborella* (77%) is the first branch in the angiosperm clade followed by *Nymphaea* (90%) and *Illicium* (90%). Monocots form a strongly supported clade (94%) and *Asparagus* is sister to a *Zea* and *Oryza* clade. In the gymnosperm clade, a gnepine clade is weakly supported (67%). The *Zamia* and *Gingko* clade is sister to the gnepine and other conifers clade. The parsimony tree based on DNA sequences of *rpb1/atpB/rbcL* genes (6015 nucleotides) also groups angiosperms and gymnosperms as two distinct clades but the support for the gymnosperm clade is only 68% (Figure 7, right). In the gymnosperm clade, the *Zamia* and *Ginkgo* clade is the sister group to conifers, and they form a group sister to gnetales. The first branches of the angiosperms are *Amborella*, *Nymphyaea* and *Illicium*.

In contrast with the parsimony tree based on the amino acid sequences of *rpb1/atpB/rbcL*, the maximum likelihood TREE-PUZZLE tree does not resolve the order of *Amborella*, *Nymphyaea* and *Illicium*. Gnetales are sister to conifers but not to *Pinus*. The maximum likelihood TREE-PUZZLE trees based on DNA sequences and amino acid sequences of *rpb1/atpB/rbcL* have a similar topology (Figure 8).

The maximum parsimony tree of the first and second codon positions (4010 nucleotides; Figure 9, left) shows the same topology as the *rpb1/atpB/rbcL* protein tree (Figure 7, left). The maximum parsimony tree of the third codon positions (2005 nucleotides; Figure 9, right) produced a tree where *Amborella* and *Illicium* are found far from the base of the angiosperm clade.

The maximum likelihood TREE-PUZZLE tree based on the first and second codon position of *rpb1/atpB/rbcL* (Figure 10, left) has a topology similar to the maximum parsimony tree except for the position of gnetales (Figure 9, left). In the former tree, gnetales are sister to all conifers. The maximum likelihood TREE-PUZZLE tree based on the third codon position of *rpb1/atpB/rbcL* (Figure 10, right) has a topology similar to the maximum parsimony tree based on the same data set (Figure 9, right), except that most angiosperm clades are not resolved.

Phylogenetic analyses of *rpb1, atpB, rbcL*, and 18S rDNA sequences:

In order to obtain more robust trees, 18S ribosomal RNA gene sequences were combined with *rpb1/atpB/rbcL* sequences. Because the third codon positions of *rpb1* are saturated (see above), two data sets were analyzed. The first data set consisted of the combined DNA sequences of the 18S rDNA genes and the first and second codon positions of *rpb1/atpB/rbcL* genes. The second data set consisted of the combined DNA sequences of the 18S/*atpB/rbcL* genes and the first and second codon positions of *rpb1*.These two data sets were analyzed using both of the maximum parsimony and maximum likelihood methods.

The phylogenetic trees inferred using combined data of 18S DNA sequences and the first and second positions of *rpb1/atpB/rbcL* DNA sequences are based on 5931 nucleotides. The parsimony tree is well supported with most clades having $\geq$ 50% bootstrap support (Figure 11, left). The angiosperm clade has 100% bootstrap support, *Amborella* is the sister group to all other angiosperms with 73% bootstrap support, followed by *Nymphaea* (94%) and *Illicium* (98%). Monocots, represented here by *Zea*, *Oryza*, and *Asparagus*, form a strongly supported clade (97%). Rosids, represented by *Pisum* and *Arabidopsis*, form a strong-supported clade (99%) that is a sister group to *Nicotiana*. The magnoliales clade, represented by *Magnolia* and *Liriodendron*, is also strongly supported (100%). In gymnosperms, gnetales and *Pinus* form a clade (59%) sister to other conifers. The *Zamia* and *Gingko* clade is sister to this clade.

Despite the strong support for the first branches of angiosperms, most of the other angiosperm branches are not resolved in the maximum likelihood tree based on the combined data of 18S DNA sequences and the first and second positions of *rpb1/atpB/rbcL* DNA sequences. In gymnosperms, the gnepine clade is not supported and conifers are monophyletic with strong support (100% support value; Figure 11, right).

The phylogenetic trees inferred using combined data of 18S/*atpB/rbcL* DNA sequences and the first and second positions of *rpb1* DNA sequences are based on 6920 nucleotides. In contrast to the data set made up of 18S DNA sequences and the first and second positions of *rpb1/atpB/rbcL* DNA sequences, the parsimony tree (Figure 12, left) places gnetales at the bottom of the remaining gymnosperms and angiosperms. *Amborella* is the sister group to all other angiosperms with 88% bootstrap support and followed by *Nymphaea* (97%) and *Illicium* (98%). Eudicots, which include *Pisum*, *Arabidopsis*, *Nicotiana*, *Beta*, and *Papaver*, are monophyletic (86%). Eudicots are sister to monocots with 84% bootstrap value.

The maximum likelihood tree based on 18S/*atpB/rbcL* DNA sequences and the first and second codon positions of *rpb1* DNA sequences shows two distinct clades of angiosperms and gymnosperms (Figure 12, right). Gnetales are sister to all conifers. Although this tree seems to give some resolution regarding the relations among eumagnoliids subclades, it does not resolve the order among *Amborella*, *Nymphyaea* and *Illicium*.

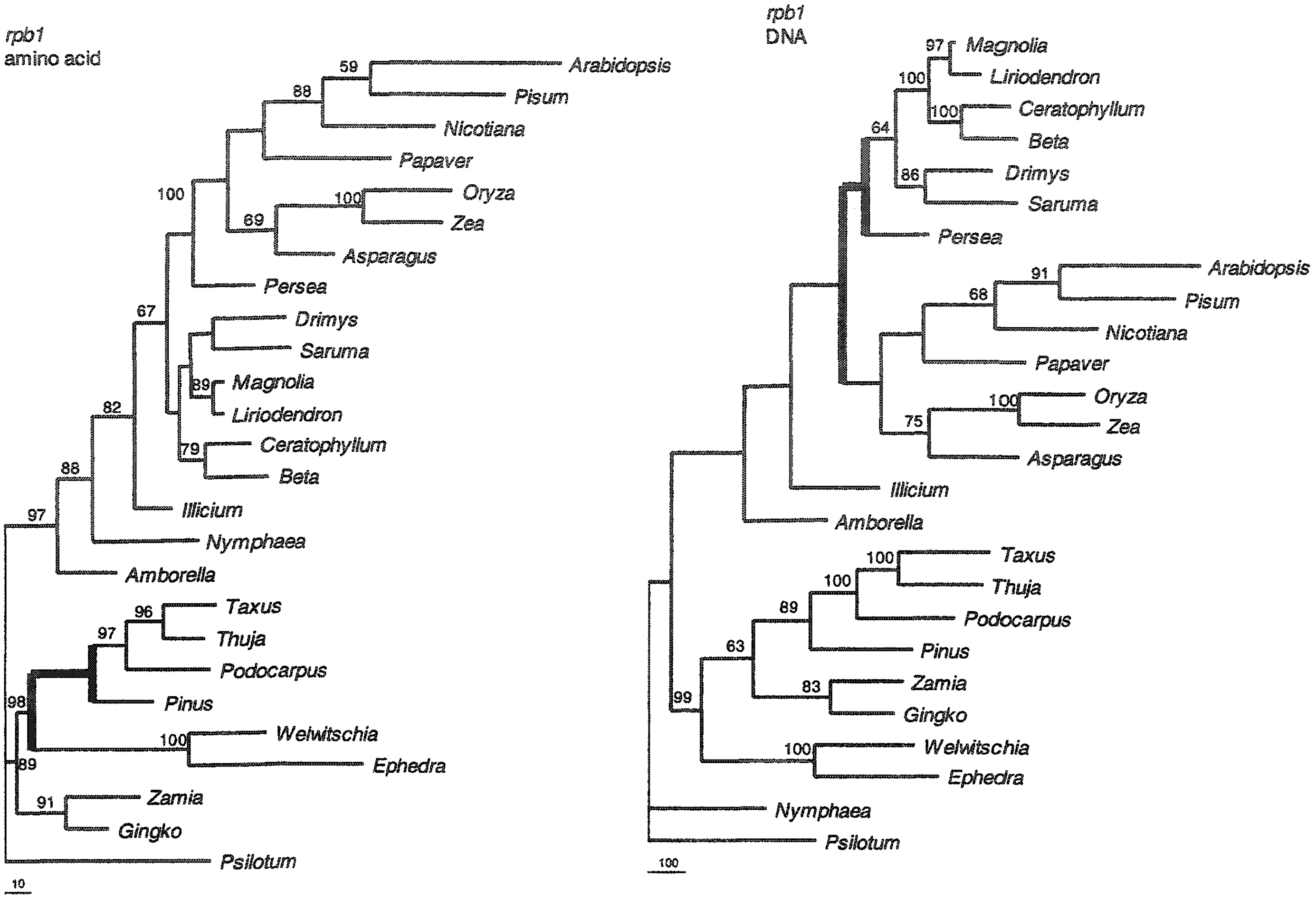


Figure 3. Maximum parsimony trees of *rpb1* amino acid sequence data (left) and DNA sequence data (right). These two trees are both one of two most parsimonious trees. Bold lines indicate branches that collapse in the strict consensus trees derived from multiple most parsimonious trees. The *rpb1* A to G region contains 1017 amino acids and 3051 nucleotides. Bootstrap values $\geq$ 50% are indicated above the branches. The scales represent the number of substitutions.

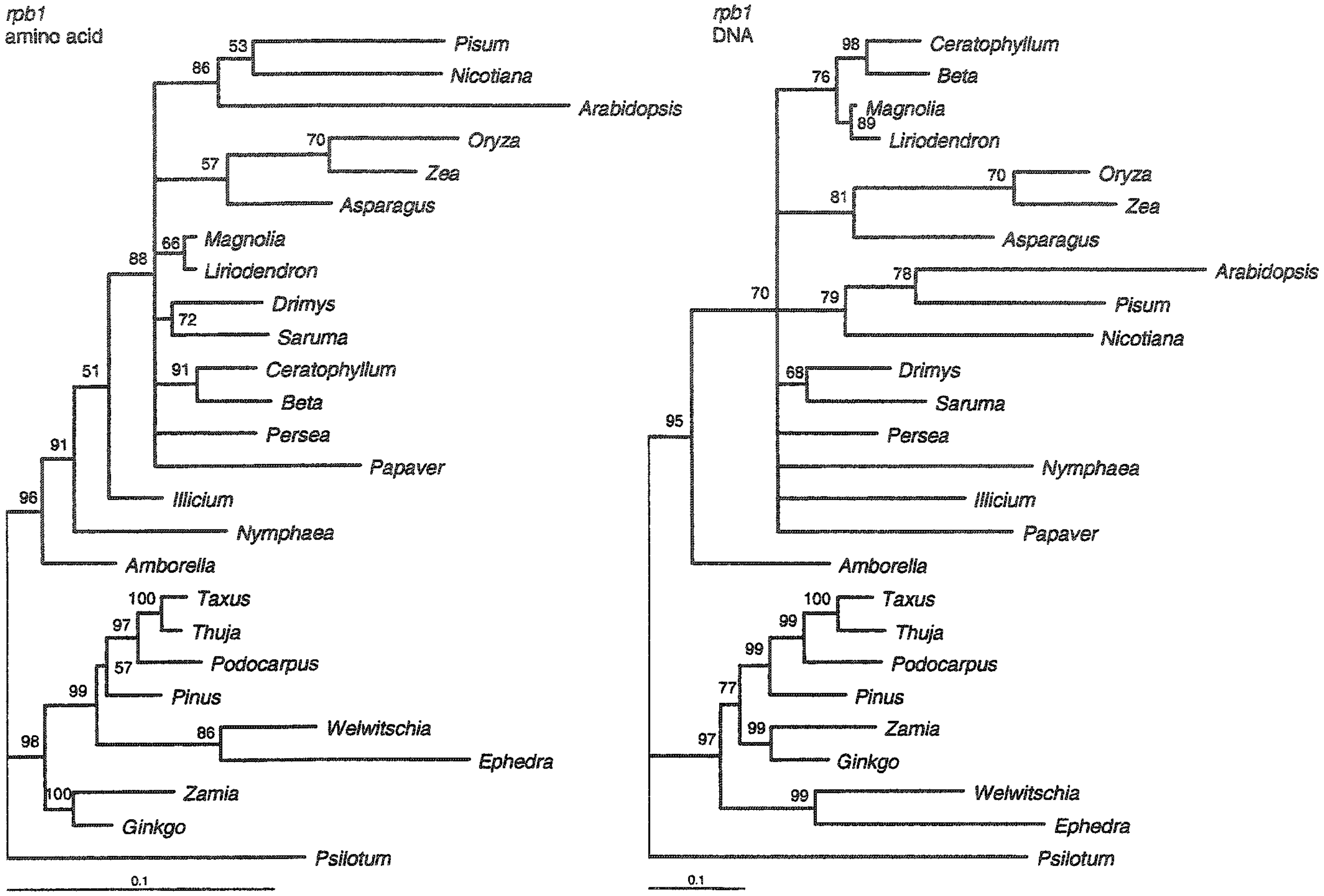


Figure 4. Maximum likelihood trees of *rpb1* amino acid sequence data (1017 amino acids, left) and DNA sequence data (3051 nucleotides, right). Support values $\geq$ 50% are indicated above the branches. The scales represent 0.1 substitution per site.

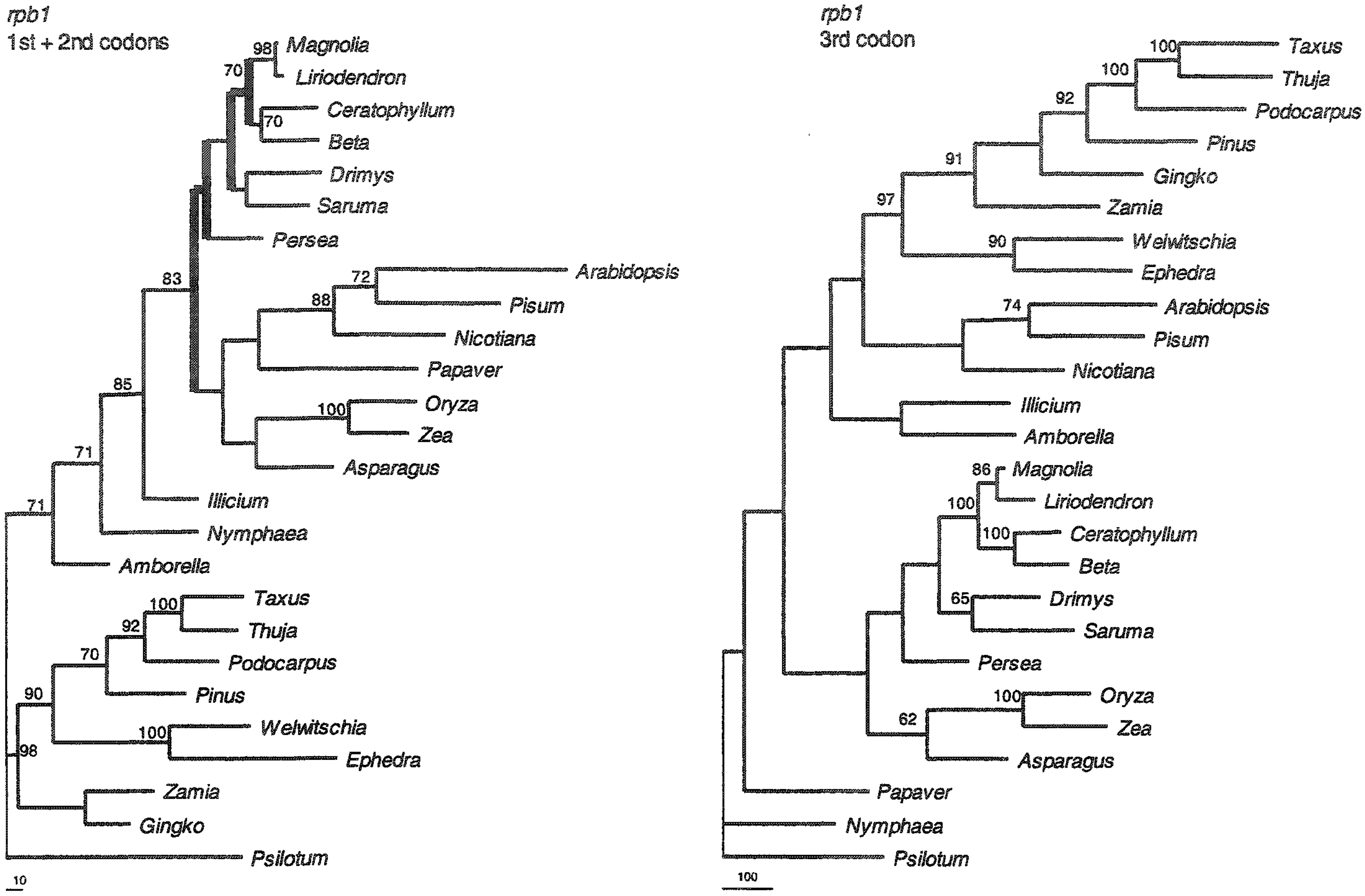


Figure 5. Maximum parsimony trees of different *rpb1* codon positions. The tree on the left is one of four best trees based on first and second codon positions (2034 nucleotides). Bold lines indicate branches that collapse in the strict consensus trees derived from multiple most parsimonious trees. The tree on the right is based on third codon positions (1017 nucleotides). Bootstrap values ≥ 50% are indicated above the branches. The scales represent the number of substitutions.

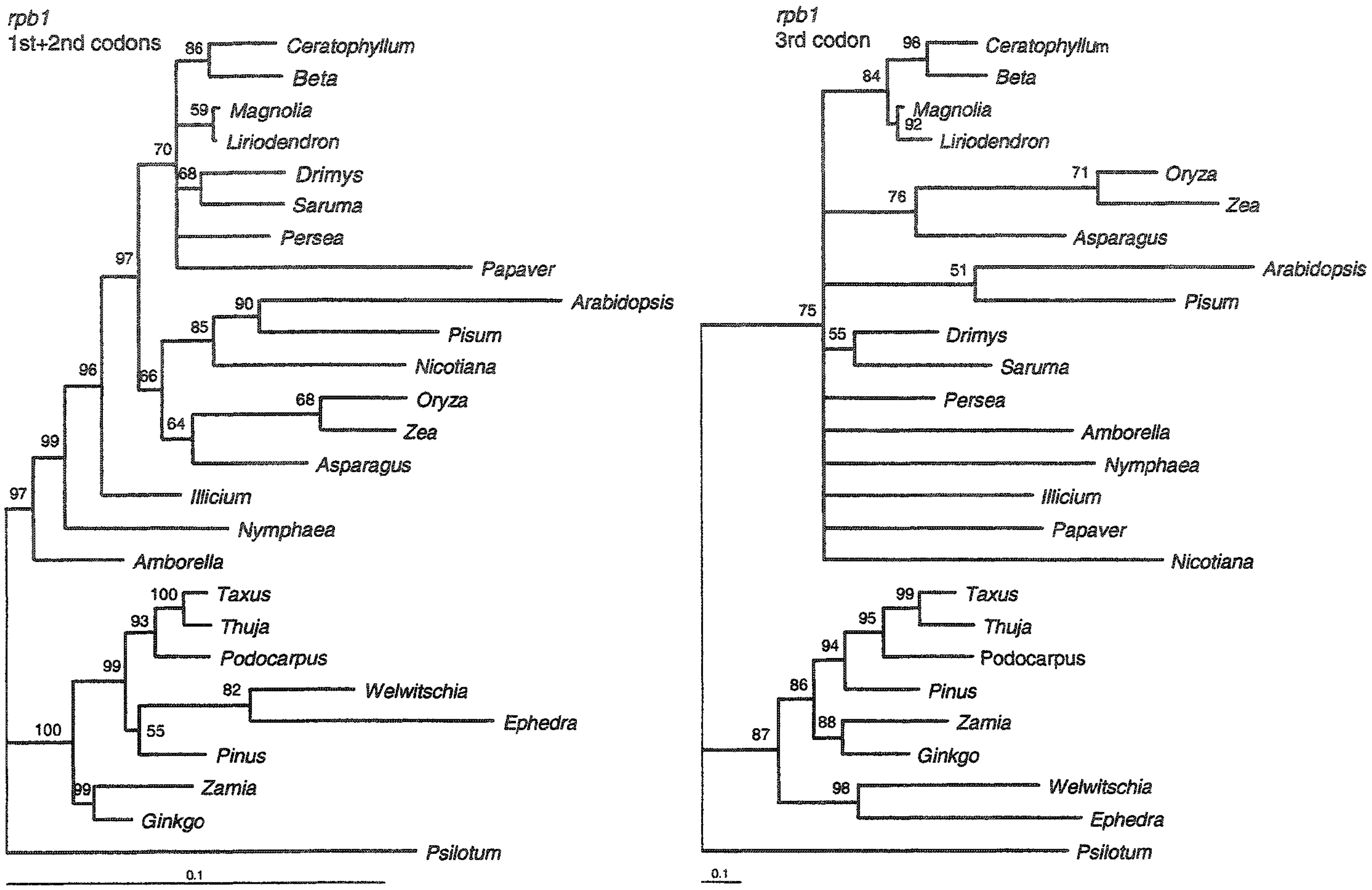


Figure 6. Maximum likelihood trees of different codon positions of *rpb1*. The tree on the left is based on first and second codon position (2034 nucleotides). The tree on the right is based on third codon positions (1017 nucleotides). Support values ≥ 50% are indicated above the branches. The scales represent 0.1 substitution per site.

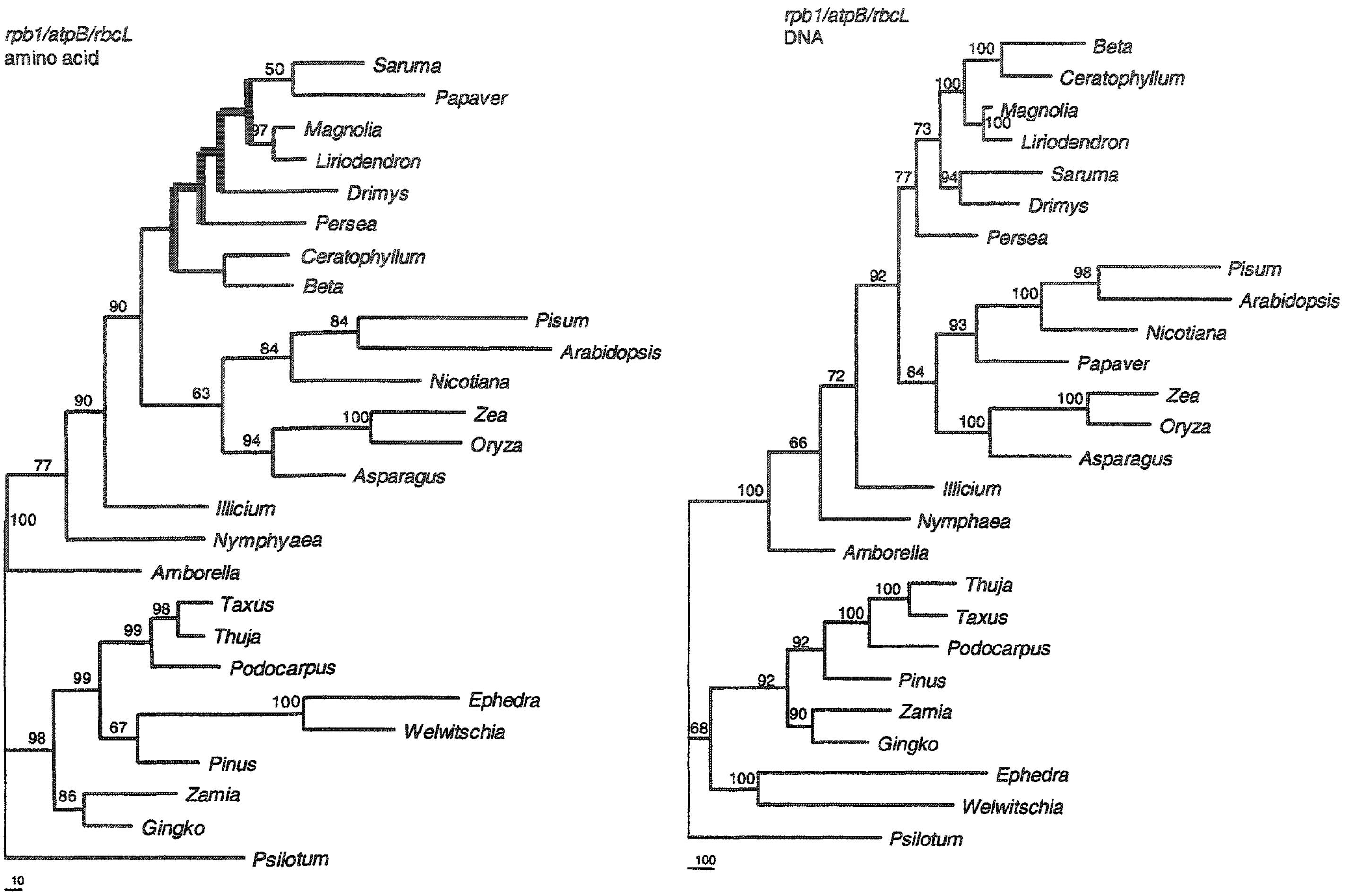


Figure 7. Maximum parsimony trees of *rpb1/atpB/rbcL* amino acid sequence data (2005 amino acids, left) and DNA sequence data (6015 nucleotides, right). The tree on the left is one of four best trees. Bold lines indicate branches that collapse in the strict consensus trees derived from multiple most parsimonious trees. Bootstrap values ≥ 50% are indicated above the branches. The scales represent the number of substitutions.

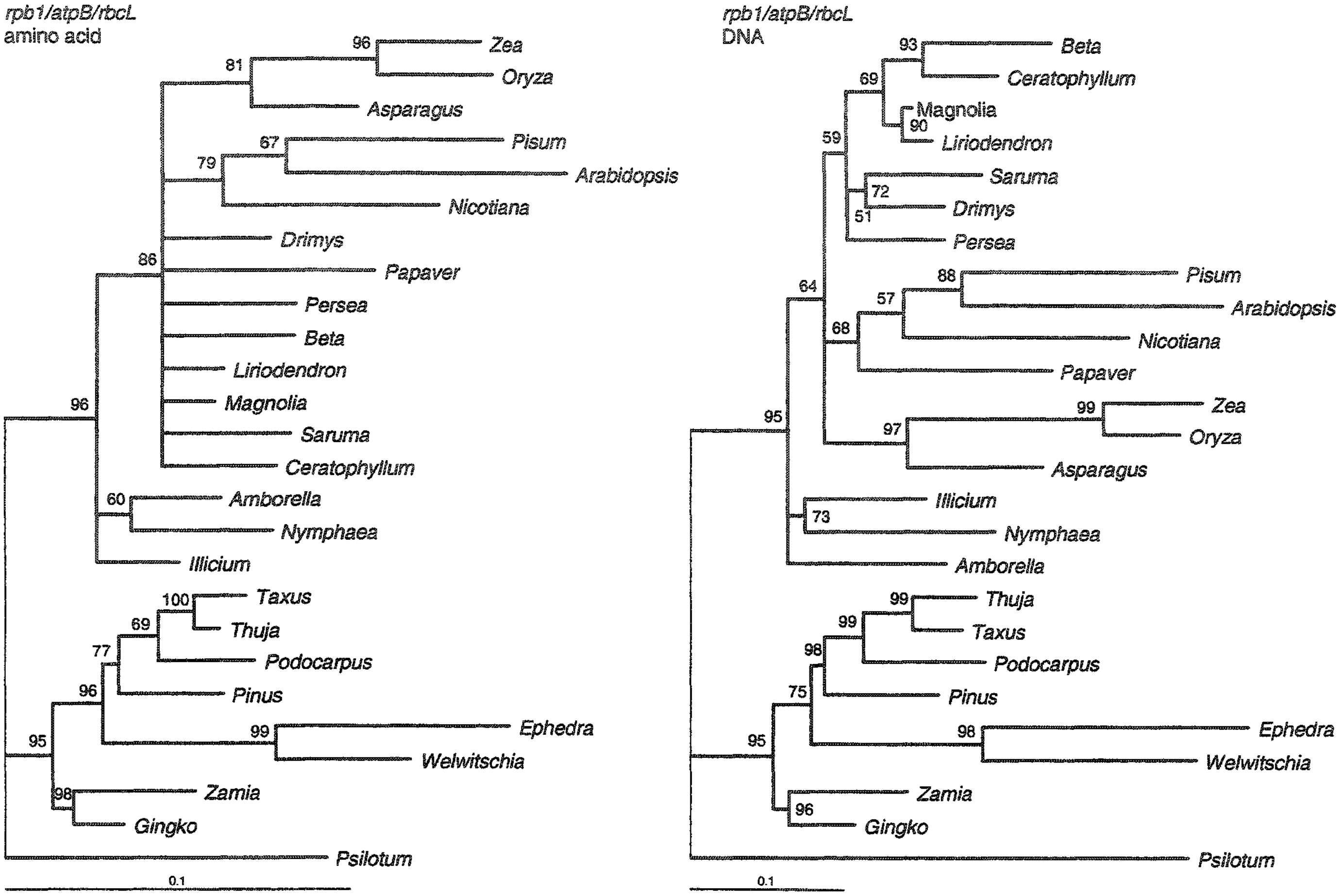


Figure 8. Maximum likelihood trees of *rpb1/atpB/rbcL* amino acid sequence data (2005 amino acids, left) and DNA sequence data (6015 nucleotides, right). Support values ≥ 50% are indicated above the branches. The scales represent 0.1 substitution per site.

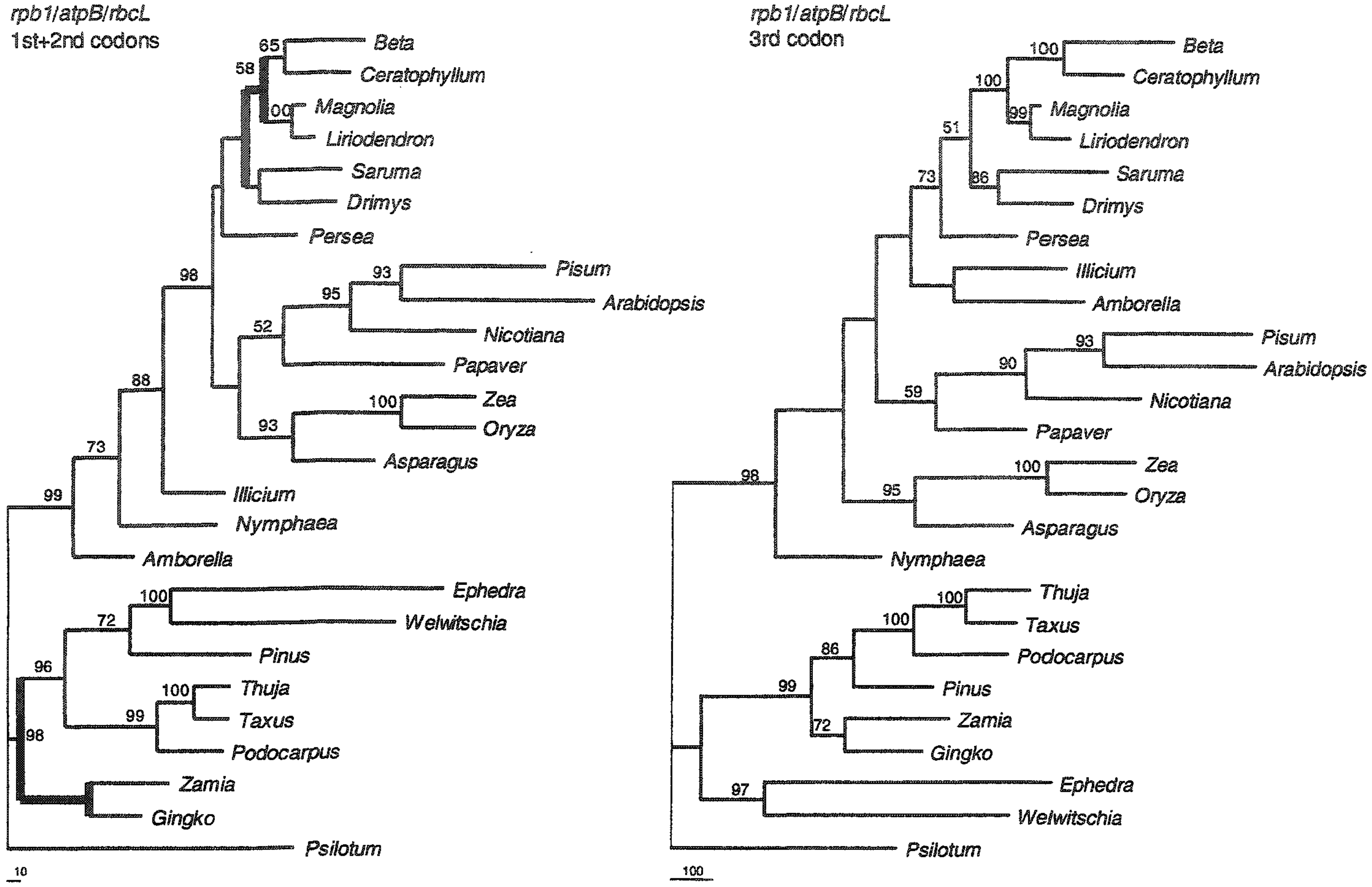


Figure 9. Maximum parsimony trees of different codon positions of *rpb1/atpB/rbcL* DNA sequence data. The tree on the left is one of four best trees based on first and second codon positions (4010 nucleotides). Bold lines indicate branches that collapse in the strict consensus trees derived from multiple most parsimonious trees. The tree on the right is based on third codon positions (2005 nucleotides). Bootstrap values ≥ 50% are indicated above the branches. The scales represent the number of substitutions.

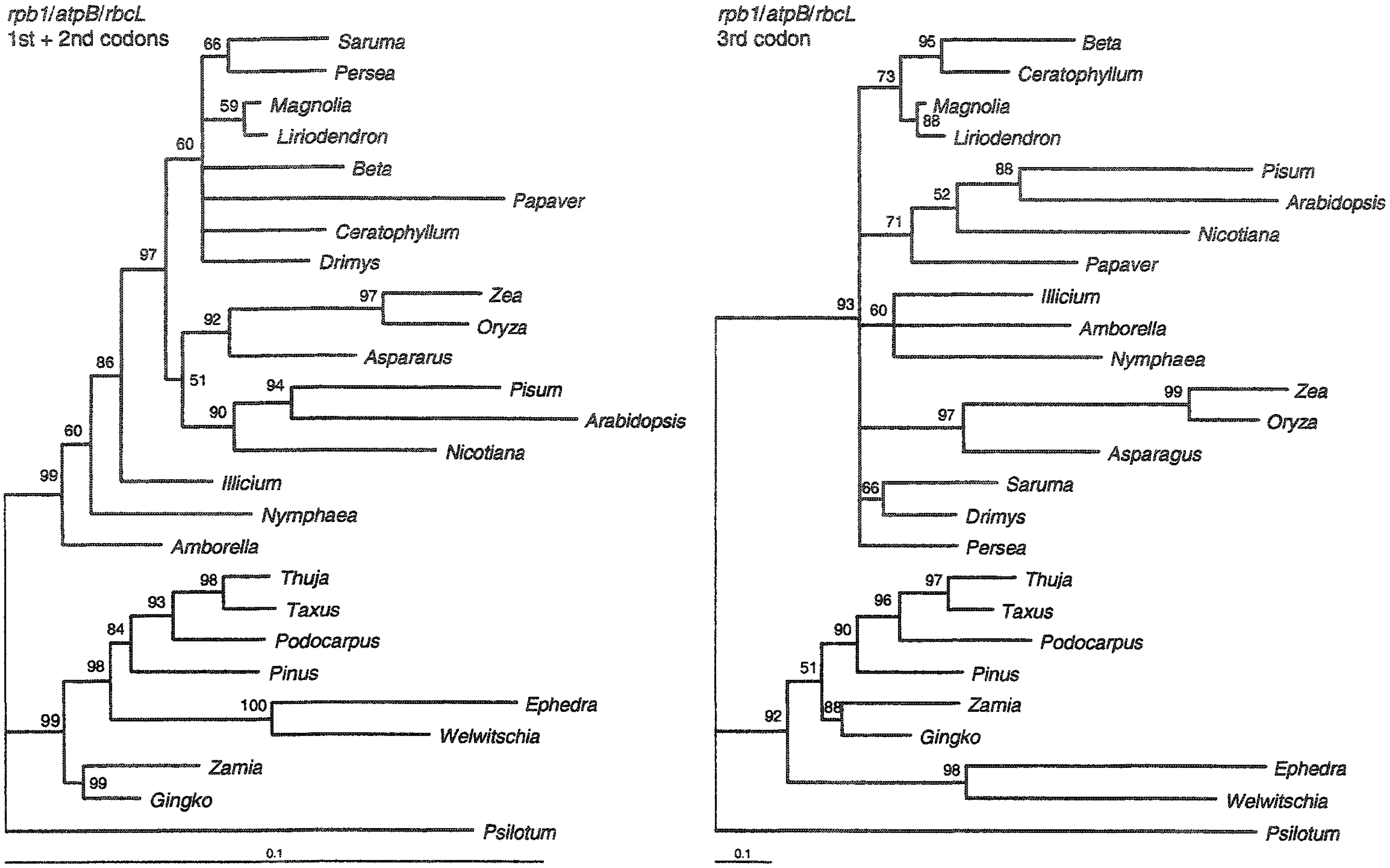


Figure 10. Maximum likelihood trees of different codon positions of *rpb1/atpB/rbcL* DNA sequence data. The tree on the left is based on first and second codon positions (4010 nucleotides). The tree on the right is based on third codon positions (2005 nucleotides). Support values ≥ 50% are indicated above the branches. The scales represent 0.1 substitution per site.

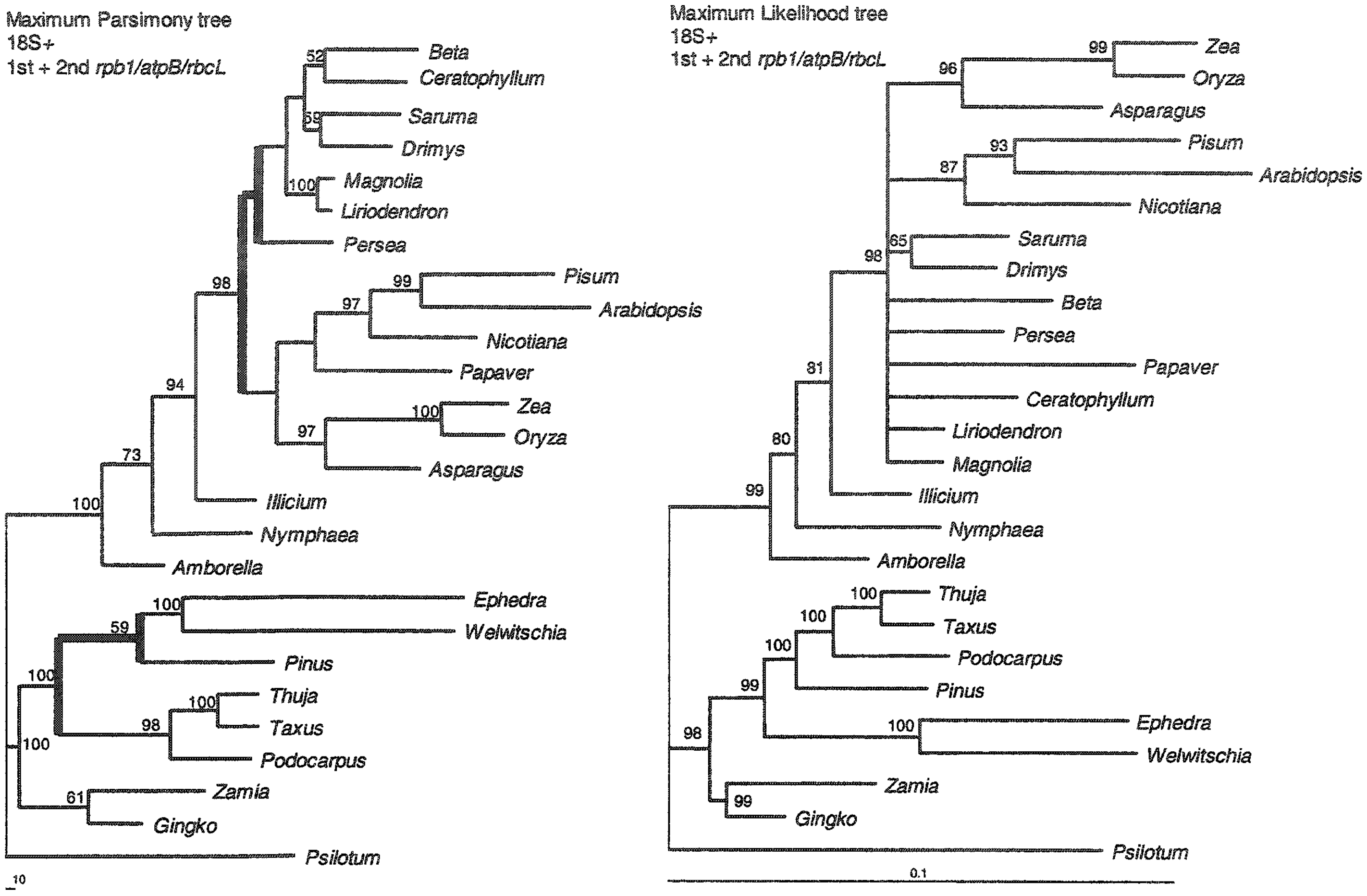


Figure 11. Phylogenetic trees inferred using 18S DNA sequences and the first and second positions of *rpb1/atpB/rbcL* DNA sequences (5931 nucleotides). The maximum parsimony tree is one of three best trees. Bold lines indicate branches that collapse in the strict consensus trees derived from multiple most parsimonious trees. Bootstrap values ≥ 50% are indicated above the branches. The scale at the bottom of this tree represents the number of substitutions. Support values ≥ 50% are indicated above the branches of the maximum likelihood tree. The scale at the bottom of this tree represents 0.1 substitution per site.

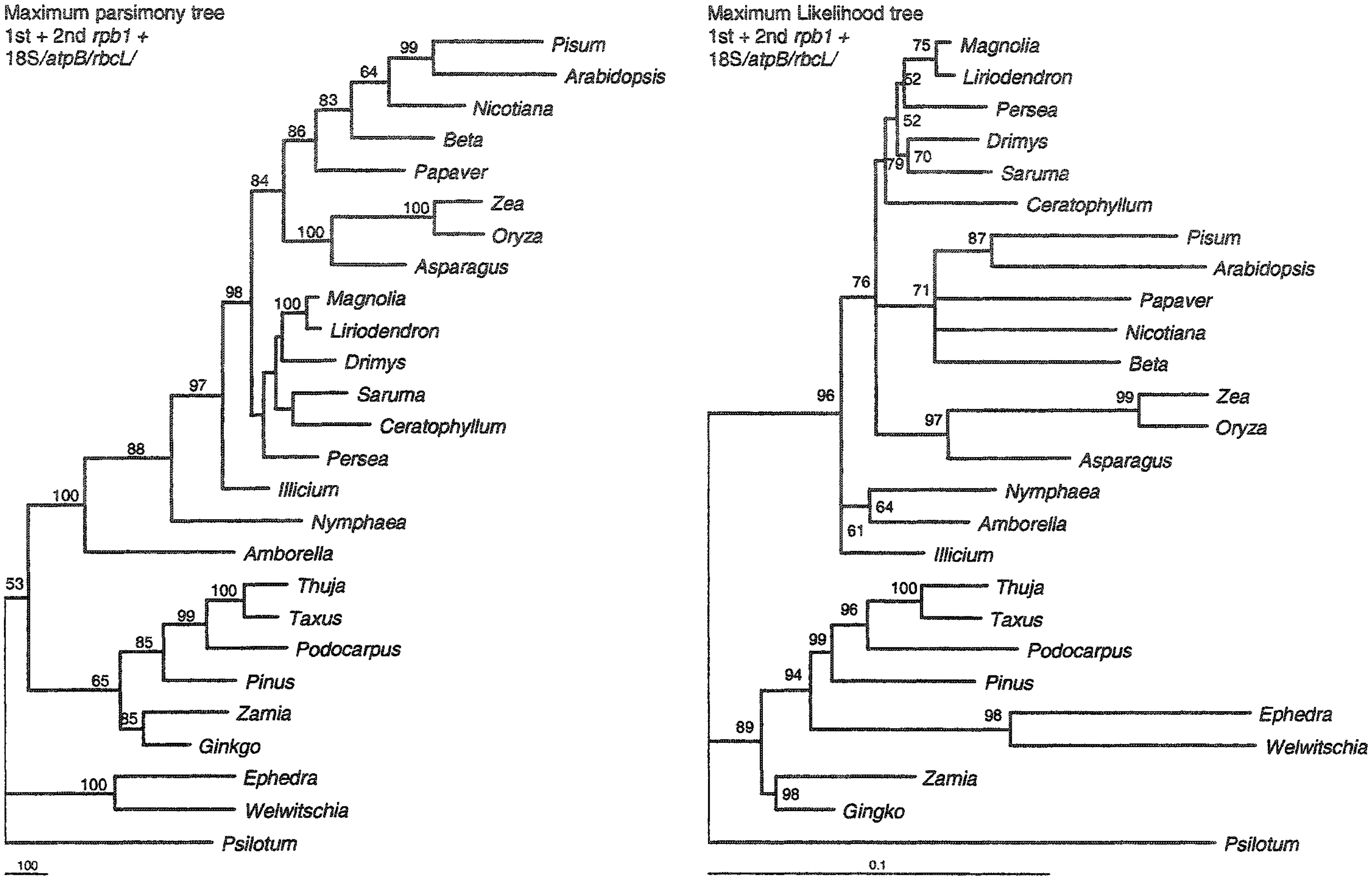


Figure 12. Phylogenetic trees inferred using 18S/*atpB*/*rbcL* DNA sequences and the first and second positions of *rpb1* genes (6920 nucleotides). Bootstrap values ≥ 50% are indicated above the branches of the maximum parsimony tree. The scale at the bottom of this tree represents the number of substitutions. Support values ≥ 50% are indicated above the branches of the maximum likelihood tree. The scale at the bottom of this tree represents 0.1 substitution per site.

Discussion

RNA isolation

Pure and intact RNA is essential for the successful production of cDNA molecules used in RT-PCR and gene cloning. However, for plant tissues, isolating high-quality RNA with good yield is often difficult due to the presence of polysaccharides and polyphenols that are released during cell disruption. These compounds form complexes with nucleic acids during tissue extraction and coprecipitate during subsequent ethanol precipitation steps (Mitra and Kootstra, 1993; Wang and Vodkin, 1994). The Qiagen RNeasy Plant Mini Kit was used to extract RNA from six plant species, but it failed to extract RNA from other plant species rich in secondary metabolites. The method developed by Bahloul and Burkard (1993) performed well with some species rich in secondary metabolites. The advantage of this method is that an extraction buffer with a low pH (5.5) was used and this low pH minimizes the ionization and subsequent oxidation of phenolics compounds which then tend to coprecipitate with nucleic acids (Guillemaut and Marechal-Dourard 1992). This method also uses polyvinylpyrrolidone (PVP) to precipitate polyphenols and cysteine as a safe and efficient anti-oxidant (Bahloul and Burkard 1993). These chemicals decrease the oxidation of polyphenols into quinones (Loomis, 1974), which would then bind and coprecipitate with nucleic acids. However, this method was still not able to produce good quality RNA for some species. Since the ethanol precipitation causes contaminants to coprecipitate with RNAs and hinders the RNA isolation, we replaced the ethanol precipitation of the RNA by a "pre-precipitation" of the contaminants (polysaccharides and secondary metabolites) using a dilute ethanol solution. This pre-precipitation was followed by an isopropanol precipitation of the RNA.

Combining data sets

In this study, because the sequences of some specific species were not available, I combined data from different taxa to provide a complete dataset for a single placeholder taxon. Because the taxa whose sequences were combined are monophyletic, the combination of data from different species into placeholder taxa, although not ideal, is unlikely to affect major results (Doyle *et al.*, 1994).

Multiple data sets are needed for estimating phylogenetic relationships. Data sets should not be combined when they strongly support conflicting tree topologies (Bull *et al.*, 1993), but Rydin et al. (2002) think that even if the incongruence between the data sets is significant, combining data is always possible and might reveal alternative phylogenetic solutions. Assessing congruence between different phylogenetic trees and data sets and ascertaining whether multiple data sets should or should not be combined into a single data matrix are essential tasks in the analysis of multiple data sets prior to phylogeny reconstruction. Some tests of incongruence are now available. However, Soltis and Soltis (2000) think that they are too coarse to be useful because they may not provide a definitive answer as to whether it is appropriate to combine data sets. Even if congruence tests reveal some low level of heterogeneity between data sets, the investigator may be justified in combining data sets (Hoot *et al.*, 1997). The best way to detect true incongruence is by examination of results (Soltis and Soltis, 2000). If hard incongruence exists, lower internal support and perhaps less resolution in groups will occur (Seelanan *et al.*, 1997). In this study, *rpb1*, *atpB*, *rbcL* and 18S rDNA genes were combined in order to obtain better resolved phylogenies.

Three methods have been proposed for handling multiple data sets in phylogenetic analyses: the combined approach, the consensus approach, and the conditional combination approach (Huelsenbeck and Bull, 1996). In the first method, all available data should be combined into a single matrix before phylogenetic analyses. It has also been referred to as the total evidence or character congruence approach, because the phylogenetic information from all characters is considered simultaneously in such analyses, and conflict between individual characters can be assessed (Larson, 1994). Kluge (1989) supports this method. In contrast, Miyamoto and Fitch (1995) support the second method that multiple data sets should be analyzed separately and the different phylogenetic estimates compared. This method seeks similarities between independent analyses for phylogenetic corroboration and assesses taxonomic congruence. Although the considerable debate regarding the advantages and limitations of the combined, consensus, and conditional combination approaches exists, many studies think that the conditional combination method is a very reasonable approach (Soltis and Soltis, 2000). This method involves combining data except in those instances in which significant heterogeneity exists between data sets and that heterogeneity appears to be attributable to different branching histories (Bull *et al.*, 1993; Lutzoni and Vilgalys, 1995).

Four genes were combined in this study. The nuclear protein coding *rpb1* gene is the central part of the transcription machinery. It is conserved from *E. coli.* to human. They have proven valuable for evolutionary research at the broadest taxonomic levels (Stiller and Hall, 1997; Hall, 2001; Liu *et al.*, 1999; Matheny *et al.* 2002). Furthermore, Nickerson and Drouin (in press) suggested that *rpb1* gene is an excellent phylogenetic marker to study the phylogeny of land plants. The plastid *atpB* and *rbcL* genes are both located in the large single-copy region of the plastid genome and are on opposite strands separated by an intergenic spacer of 600–800 base pairs (Savolainen *et al.*, 1997). The two genes code for distinct enzymatic functions. The *rbcL* gene codes for the large subunit of the ribulose-1, 5-bisphosphate-carboxylase/oxygenase, a free enzyme in the stroma, whereas the *atpB* gene codes for the beta subunit of the ATP synthase, which is bound to the thylakoid-membrane (Zurawski *et al.*, 1982). Because they are all part of the same non-recombining piece of DNA, they should have the same history. The 18S rDNA is the small subunit ribosomal DNA. Although 18S rDNA genes have been used for phylogeny reconstruction within many groups of eukaryotes, including most major groups of plants, the 18S rDNA trees are highly unstable and have relatively low numbers of supported groups compared with those based on *rbcL* and *atpB* (Soltis *et al.*, 1998; Chase and Cox, 1998; Hoot *et al.*, 1995, 1999). The basic problem of 18S rDNA sequences is that they have fewer variable positions experiencing most of the change, so they therefore contain less phylogenetic information than *rbcL* and *atpB* genes (Chase and Cox, 1998). However, 18S rDNA data can make a valuable contribution to phylogenetic study when combined with other data. For example, the combined 18S rDNA, *atpB*, and *rbcL* genes produce substantially stronger evidence of relationships than any of the separate analyses (Hoot *et al.*, 1995, 1999; Soltis *et al.*, 1997a, 1998; Chase and Cox, 1998).

It has been argued that more rapidly evolving sites (i.e., third codon positions in protein-coding chloroplast genes) should be down-weighted on grounds of saturation (Hansen *et al.*, 1999; Chaw *et al.*, 2000). Sanderson *et al.* (2000) found conflict between data partitions (first and second vs. third positions) within two plastid genes (*psa*A and *psb*B). But Rydin *et al.* (2002) analyzed four data sets (*rbc*L, *atp*B, 26S, and 18S rDNA) both separately and combined using different character coding and they found no conflict between codon positions in the plastid sequences but they found conflict between transitions and transversions within each position. They showed that first and second positions evolve more slowly in *atpB* and *rbcL*, but analyses of these positions resulted in poorly resolved phylogenies. However, third positions contained most of the phylogenetic information in the plastid data set even if they may be saturated in the conventional sense. They concluded that third codon positions are also potentially more informative (see also Källersjö *et al.* 1998, 1999). Savolainen *et al.* (2000) also supported that third codon positions and transitions of *atpB* and *rbcL* gene were the most useful characters in phylogenetic reconstructions.

Mehrdad Hajibabaei, a Ph.D. student in our lab, calculated pairwise distances in 1+2 codon positions as well third codon positions of RNA polymerases, chloroplast and mitochondrial data sets. He used three nuclear encoded RNA polymerase genes (*rpa1*, *rpb1*, and *rpc1*), four chloroplast genes (*psaA, psbB, rbcL* and *atpB*), two mitochondrial genes (*cox1* and *atpA*) and nuclear 18S rRNA. He found that the majority of the distances in the third codon position of RNA polymerases and chloroplast data sets are close to or above 1 substitution per site, but for mitochondrial genes, only a small fraction of the distances are above 1 substitution per site. He suggested that the third codon positions of nuclear RNA polymerase genes and chloroplast genes could cause misleading phylogenies, because these positions are saturated and they will lead to incorrect phylogenetic inference in highly diverged data sets (Hajibabaei, 2003).

My study also showed that there were multiple synonymous substitutions in *rpb1* genes (average Ks of 1.408 ± 0.038 substitutions per site). When distant lineages, angiosperms and gymnosperms, were combined, this high synonymous substitution rate results in saturation. Because the majority of synonymous substitutions involves third codon positions, third codon positions of *rpb1* are expected to provide little phylogenetic signal. My conclusion is the same as that of Hajibabaei (2003) that the third codon positions of *rpb1* gene could cause erroneous phylogenies (Figure 5, 6, right) and should not be used for phylogenetic reconstruction. Therefore, in my combined data sets (*rpb1*, *atpB*, *rbc*L, and 18S rDNA), only the first and second codon positions of *rpb1* were used.

Because of the arguments about the usefulness of third codon positions of *atpB* and *rbcL* genes (Savolainen *et al.*, 2000; Rydin *et al.*, 2002; Hajibabaei, 2003), I made two combined data sets. One is composed of all positions of 18S/*atpB*/*rbcL* DNA sequences and first and second codon positions of *rpb1* DNA sequences. The other one is composed of 18S sequences and first and second codon positions of *rpb1*/*atpB*/*rbcL* DNA sequences. My substitution analyses showed that although the *atpB* and *rbcL* genes had relatively low average synonymous substitution rates (0.478±0.030 and 0.463±0.025, respectively), pairwise synonymous site distances between some distant species were close to 1.0 substitution per site (result not shown).

Phylogenies inferred from *rpb1* sequences

I made *rpb1* phylogenetic trees with four data partitions: protein sequences, all codon positions, first and second codon positions and third codon positions. Protein sequences and first and second codon positions seemed more useful for phylogenetic reconstructions. Third codon positions could not help resolve angiosperm phylogeny (Figure 5, Figure 6, right), and they also brought much noise into the trees based on all codon positions (Figure 3, Figure 4, right). These results concur with the results of substitution rate analyses that third codon positions were saturated.

The topology of first and second codon positions trees were in line with that of protein trees, except that the maximum likelihood tree placed the gnetales with *Pinus* as a sister group to the other conifers (but the support value of this group was only 55%). They strongly supported that both angiosperms and gymnosperms were monophyletic. Basal angiosperms (*Amborella, Nymphaea,* and *Illicium*) were well supported and consistent with recent molecular phylogenetic studies. Those studies all strongly supported species of the "ANITA clade" (*Amborella, Nymphaeales, Illicales, Trimeniaceae,* and *Austrobaileyaceae*) as being the first lineages at the base of the angiosperm tree. *Amborella* was found at the base of the angiosperm tree, followed by *Nymphaeales* and then ITA members (Matthews and Donoghue, 1999; Graham and Olmstesd, 2000; Soltis *et al.*, 1999; Qiu *et al.*, 1999; Parkinson *et al.*, 1999). Magnoliales (represented by *Magnolia* and *Liriodendron*), Monocots (represented by *Zea*, *Oryza*, and *Asparagus*) and Rosids (represented by *Pisum* and *Arabidopsis*) are all monophyletic. Eudicots (represented by *Pisum*, *Arabidopsis*, *Nicotiana*, and *Papaver*) were monophyletic with the exception of *Beta.* Surprisingly, *Beta* was placed together with C*eratophyllum.* The interrelationships of eudicots and eumagnoliids were not resolved. In gymnosperms, conifers and gnetales formed a clade sister to the *Zamia* and *Gingko* clade. The gnetales clade (represented by *Welwitschia* and *Ephedra*) was strongly supported, but the conifer clade was not strongly supported because of the position of gnetales. The Gnepine Hypothesis was weakly supported by the maximum likelihood tree based on first and second codon positions (Figure 6). The Anthophyte Hypothesis was rejected, i.e., gnetales are not sister to angiosperms.

Comparing the first and second codon positions trees to the third codon position trees of *rpb1* genes, the former ones evolve too slowly to provide enough information, while the latter evolve so fast that they provide misleading information. Therefore, although the nonsynonymous sites of *rpb1* genes evolve 2.5 times faster than those of the chloroplast *atpB* and *rbc*L genes (Ka = 0.067 ± 0.004, 0.022 ± 0.004, 0.026 ± 0.004, respectively), they still don't have enough information sites to resolve the phylogeny of seed plants.

Phylogenies inferred from *rpb1, atpB, rbcL,* and 18S rDNA sequences

Larger data sets were formed by combining *rpb1* genes with 18S ribosomal RNA genes and chloroplast *atpB and rbcL* genes. The combined data of 18S sequences and the first and second positions of *rpb1/atpB/rbcL* DNA sequences contains 5931 nucleotides, while 18S/*atpB/rbcL* DNA sequences and the first and second positions of *rpb1* DNA sequences contains 6920 nucleotides. The combined data sets produced topologies similar to the corresponding analysis of the first and second codon positions of *rpb1* genes and, in some cases, showed more resolved nodes and stronger support values.

The first branches of angiosperm are well or strongly supported in all trees except for the maximum likelihood tree based on 18S/*atpB/rbcL* DNA sequences and the first and second codon positions of *rpb1* DNA sequences. *Amborella* is the sister group to all other angiosperms followed by *Nymphaea* and *Illicium*. The maximum likelihood tree based on 18S/*atpB/rbcL* DNA sequences and the first and second codon positions of *rpb1* DNA sequences does not resolve the order among *Amborella*, *Nymphyaea* and *Illicium*. The *rbcL* sequences are likely responsible for this result, because *rpb1* (Nickerson and Drouin, in press), 18S (Soltis et al., 1997) and *atpB* alone (Savolainen et al., 2000) gave a grade of *Amborella, Nymphaea* and *Illicium,* but in some analyses based on *rbcL* (Qiu *et al*., 1993), *rbcL/atpB* (Savolainen *et al*.,

2000) and *rbcL*/nonDNA charaters (Nandi *et al.*, 1998), *Amborella, Nymphaea* and *Illicium* formed a clade.

As one of the basal angiosperm candidates, *Ceratophyllaceae* was indicated to be the sister to all other angiosperms in some early analyses (Chase *et al.*, 1993). In recent studies, the position of Ceratophyllaceae is still uncertain. As an isolated genus, *Ceratophyllum* is sister to either all other angiosperms, the eudicots, or the monocots (reviewed by Soltis *et al.*, 2000). In my study, *Ceratophyllum* is sister to either *Beta*, *Saruma*, or eumagnoliids. Because support values were low (52%, or < 50%), I conclude that none of these results are reliable.

Eumagnoliids include Chloranthales, monocots, Piperales, Laurales, Winterales and Magnoliales. As the placeholder of Chloranthales, *Chloranthus* was not included in my study because I was not able to clone the DG fragment of this *rpb1* gene. Monocots and Magnoliales are monophyletic, whereas Piperales, Laurales, and Winterales are represented by only one species. The combined data sets can not resolve the interrelationship among these subclades. The maximum likelihood tree based on 18S/*atpB*/*rbcL* DNA sequences and the first and second positions of *rpb1* DNA sequences place the Piperales (*Saruma*) and Winterales (*Drimys*) clade as a sister group of the Magnoliales (*Liriodendron* and *Magnolia*) and Laurales (*Persea*) clade. This is the same as shown in Figure 1 of Kuzoff and Gasser (2000). However, because the support values are low, the interrelationship of eumagnoliides is therefore not resolved by our data sets.

Eudicots, which include *Pisum*, *Arabidopsis*, *Nicotiana*, *Beta*, and *Papaver* in this study, are monophyletic in the phylogenetic trees based on 18S/*atpB*/*rbcL* DNA sequences and the first and second positions of *rpb1* DNA sequences (Figure 12). Ranunculales (*Papaver*) is sister to the other Eudicots (83%). The result of Ranunculales being the first-branching eudicots is the same as the analyses of 18S (Soltis *et al.*, 1997), *rbcL* (Chase *et al.*, 1993), *atpB* (Savolainen *et al.*, 2000) and 18S/*atpB*/*rbcL* (Soltis *et al.*, 2000). Following Ranunculales, Caryophyllales (*Beta*) is a sister group of the Asterids (*Nicotiana*) and monophyletic Rosids (*Pisum* and *Arabidopsis*) clade. The relationship among these three subclades is unclear in former studies (Soltis et al., 1999; Kuzoff and Gasser, 2000). Furthermore eudicots are sister to monocots to form a strong clade (84%) in the maximum parsimony tree (Figure 12, left). In the phylogenetic trees of 18S DNA sequences and the first and second positions of *rpb1*/*atpB*/*rbcL* DNA sequences, *Beta* is placed out of eudicots because of its *rpb1*gene.

The topologies of gymnosperms are different due to the position of gnetales. Gnetales (*Welwitschia* and *Ephedra*) are monophyletic with strong support. They are sister to either *Pinus* (Figure 11, left), all conifers (Figure 11, right; Figure 12, right), or the remaining gymnosperms and all angiosperms (Figure 12, left). So the position of gnetales is still unclear. The controversial Gnepine Hypothesis (Chaw *et al.*, 2000; Qiu *et al.*, 1999; Bowe *et al.*, 2000) where gnetales and *Pinus* form a clade sister to other conifers, is not strongly supported (59%, Figure 11, left). The sister relationship between *Zamia* and *Gingko* is contrary to previous studies where Cycads (*Zamia*) were shown to be sister to all the other gymnosperms and followed by *Gingko* (Kuzoff and Gasser, 2000).

Conclusion

Rpb1 genes from twenty-six land plant species (8 gymnosperm species, 17 angiosperm species and *Psilotum*, a fern ally, as the outgroup) were used to elucidate the phylogeny of seed plants. Among these twenty-six species, sixteen new *rpb1* genes were sequenced in this study. This study provides the further evidence that *rpb1* genes are good phylogenetic markers to study the phylogeny of seed plants (Nickerson and Drouin, in press). The length of *rpb1* genes is conserved (3027~ 3029 nucleotides) and they contain no base compositional bias. Whereas the first and second codon positions are useful for phylogenetic inference, the third codon positions are not because synonymous substitutions of *rpb1* sequences are saturated when compared to the outgroup. The nonsynonymous sites of *rpb1* genes evolve 2.5 times faster than the chloroplast *atpB* and *rbc*L genes (Ka = 0.067 ± 0.004, 0.022 ± 0.004, 0.026 ± 0.004, respectively) and provide useful phylogenetic information.

The topology of first and second codon positions trees of *rpb1* was in line with that of protein trees. These trees show both angiosperms and gymnosperms as being monophyletic. *Amborella* was found at the base of the angiosperm tree, followed by *Nymphaeales*, then *Illicium*. The Anthophyte Hypothesis was rejected. These results are consistent with recent molecular phylogenetic studies. The Gnepine Hypothesis was weakly supported by the maximum likelihood tree. However, the interrelationships among eumagnoliids and eudicots are not resolved. Although *rpb1* genes are useful, they still have a weaker phylogenetic signal than data sets made of multiple genes. A data set with multiple genes is more likely to contain more phylogenetic information across the various hierarchical levels of a phylogeny than one data set alone (Donoghue and Sanderson, 1992; Hoot, 1995). The multiple gene data sets were made up of *rpb1* genes, 18S ribosomal RNA gene and chloroplast *atpB* and *rbcL* genes. The combined data sets had a topology similar to that of the *rpb1* genes, except that they provided better resolution of the interrelationship among eumagnoliids and eudicots.

Improved inferences of land plant reconstruction will require additional data. Hajibabaei (2003) used three nuclear encoded RNA polymerase genes (*rpa1*, *rpb1*, and *rpc1*), four chloroplast genes (*psaA*, *psbB*, *rbcL* and *atpB*), two mitochondrial genes (*cox1* and *atpA*) and nuclear 18S rRNA (total of 16223 nucleotides) to evaluate the Gnepine Hypothesis. He combined 1+2 codon positions of chloroplast and nuclear genes and all codon positions of mitochondrial genes and 18S rRNA in a data set. A single topology was obtained with high bootstrap support in both maximum likelihood and maximum parsimony trees, and the Gnepine Hypothesis was strongly supported. Therefore, a larger data set should be built for twenty-six plant species used in this study to perform phylogenetic analyses. Such a large data set should produce a more robust phylogeny of seed plants. Chloranthales is an important clade of eumagnoliids. As the placeholder of Chloranthales, *Chloranthus* should have been included in my study. Unfortunately, I was not able to clone the DG fragment of this *rpb1* gene.

Analysis of combined data would likely be most informative for land plant phylogeny These combined data are not only from combined DNA sequence data set, but also from other DNA characters (for example, cpDNA structural changes, gene and gene intron loss; Downie and Palmer, 1992), nonmolecular data sets (for example, morphology and anatomy), and fossils (Doyle, 1998). This approach is encouraged by the Green Plant Phylogeny Research

Coordination Group, and it should ultimately provide the best estimate of plant phylogeny (Soltis and Soltis, 2000).

References

Bahloul, M. and Burkard, G. (1993). An improved method for the isolation of total RNA from spruce tissues. *Plant Mol. Biol. Rep.* 11: 212-215.

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 analyses of sequences from all three genomic compartments converge on the root of flowering plant phylogeny. *Proc. Natl. Acad. Sci. USA* 97: 13166-13171.

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. *Proc. Natl. Acad. Sci. USA* 97: 4092-4097.

Bull, J. J., Huelsenbeck, J. P., Cunningham, C. W., Swofford, D. L., and Waddell, P. J. (1993). Partitioning and combining data in phylogenetic analysis. *Syst. Biol.* 42:384-397.

Chase, M. W., Soltis, D. E., Olmstead, R. G., Morgan, D., Les, D. H., Mishler, B. D., 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., Furnier, G. R., Strauss, S. H., Xiang, Q-Y., Plunkett, G. M., Soltis, P. S., Swensen, S. M., Williams, S. E., Gadek, P. A., Quinn, C. J., Eguiarte, L. E., Golenberg, E., Learn, G. H. Jr., 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 plastid gene *rbc*L. *Ann. Missouri Bot. Gard.* 80: 528-580.

Chase, M. W. and Cox, A. V. (1998). Gene sequences, collaboration, and analysis of large data sets. *Aust. Sys. Bot.* 11: 215-229.

Chaw, S-M., Zharkikh, A., Sung, H-M., Lau, T-C. and Li, W-H. (1997). Molecular phylogeny of extant gymnosperms and seed plant evolution: Analysis of nuclear 18S rRNA sequences. *Mol. Biol. Evol.* 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. *Proc. Natl. Acad. Sci. USA* 97: 4086-4091.

Conaway, R. C. and Conaway, J. W. (1997). General transcription factors for RNA polymerase II. *Prog. Nucleic Acid Res. Mol. Biol.* 56: 327-346.

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

Crane, P. R., Friis, E. M. and Pedersen, K. R. (1995). The origin and early diversification of angiosperms. *Nature* 374: 27-33.

Crepet, W. L. (2000). Progress in understanding angiosperm history, success, and relationships: Darwin's abominably "perplexing phenomenon". *Proc. Natl. Acad. Sci. USA* 97: 12939-12941.

Crepet, W. L. (1996). Timing in the evolution of derived floral characters: Upper Cretaceous (Turonian) taxa with tricolpate and tricolpate-derived pollen. *Rev. Paleob. Palyn.* 90: 339-359.

Denton, A. L., McConaughy, B. L. and Hall, B. D. (1998). Usefulness of RNA polymerase II coding sequences for estimation of green plant phylogeny. *Mol.*

Biol. Evol. 15: 1082-1085.

Dilcher, D. L. (1989). The occurrence of fruits with affinities to Ceratophyllaceae in lower and mid-Cretaceous sediments. *Am. J. Bot. (Suppl.)* 76:162.

Downie, S. R. and Palmer, J. D. (1992). Use of chloroplast DNA rearrangements in reconstructing plant phylogeny. *In* "Molecular Systematics of Plants" (Soltis, P. S., Soltis, D. E., and Doyle, J. J., Eds.), pp. 14-35, Chapman & Hall, N.Y.

Doyle, J. A. (1996). Seed plant phylogeny and the relationships of Gnetales. *Int. J. Plant Sci.* 157: S3-S39.

Doyle, J. A. (1998). Molecules, morphology, fossils, and the relationship of angiosperms and Gnetales. *Mol. Phylogenet. Evol.* 9: 448-462.

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

Doyle, J. A. and Donoghue, M. J. (1992). Fossils and seed plant phylogeny revisited. *Brittonia* 44: 89-106.

Doyle, J. A. and Donoghue, M. J. (1993). Phylogenies and angiosperm diversification. *Paleobiology* 19: 141-167.

Doyle, J. A., Donoghue, M. J. and Zimmer, E. A. (1994). Integration of morphological and ribosomal RNA data on the origin of angiosperms. *Ann. Mo. Bot. Gard.* 81: 419-450.

Drouin, G., Part, F., Ell, M. and Clarke, G. D. (1999). Detecting and characterizing gene conversions between multigene family members. *Mol. Biol. Evol.* 16: 1369-1390.

Ebright, R. H. (2000) RNA polymerase: structural similarities between bacterial RNA polymerase and eukaryotic RNA polymerase II. *J. Mol. Biol.* 304: 687-698.

Floyd, S. K., Lerner, V. T., and Friedman, W. E. (1999). A developmental and evolutionary analysis of embryology in *Platanus* (Platanaceae), a basal eudicot. *Am. J. Bot.* 86: 1523-1537.

Friedman, W. E. and Floyd, S. K. (2001). Perspective: The origin of flowering plants and their reproductive biology - A tale of two phylogenies. *Evolution* 55: 217-231.

Friedman, W. E. (1990). Double fertilization in Ephedra, a nonflowering seed plant: Its bearing on the origin of angiosperms. *Science* 247: 951-954.

Friedman, W. E. (1994). The evolution of embryogeny in seed plants and the developmental origin and early history of endosperm. *Am. J. Bot.* 81: 1468-1486.

Friis, E. M., Pedersen, K. R. and Crane, P. R. (1994). Early angiosperm diversification: the diversity of pollen associated with reproductive structures in early Cretaceous floras from Portugal. *Ann. Mo. Bot. Gard.* 86: 259-296.

Friis, E. M., Crane, P. R. and Pedersen, K. R. (1986). Floral evidence for Cretaceous chloranthoid angiosperms. *Nature* 320: 163-164.

Graham, S. W. and Olmstead, R. G. (2000). Utility of 17 chloroplast genes for inferring the phylogeny of the basal angiosperms. *Am. J. Bot.* 87: 1712-1730.

Guillemaut, P. and Marechal-Dourard, L. (1992). Isolation of plant DNA: a fast inexpensive and reliable method. *Plant Mol. Biol. Rep.* 11: 32-37.

Hajibabaei, M. (2003). Molecular evolution of the RNA polymerase genes and the phylogeny of seed plants. Ph.D thesis. University of Ottawa.

Hall, D. B. Molecular systematics using gene sequences encoding nuclear RNA polymerase subunits. Background Information for Ben Hall's talk at "Deep Hypha I" Salt Lake City, Utah August 25, 2001. Available from World Wide Web:

<http://faculty.washington.edu/benhall/>

Hamby, R. K. and Zimmer, E. A. (1992) *in* "Molecular Systematics of Plants" (Soltis, P. S., Soltis, D. E. and Doyle, J. J. Eds.), pp. 50-91, Chapman & Hall, N.Y.

Hansen, A., Hansmann, S., Samigullin, T., Antonov, A. and Martin, W. (1999). *Gnetum* and the angiosperms: Molecular evidence that their shared morphological characters are convergent, rather than homologous. *Mol. Biol. Evol.* 16: 1006-1009.

Hickey, L. J. and Taylor, D. W. (1996) *In* "Flowering Plant Origin, Evolution and Phylogeny" (Taylor, D. W. and Hickey, L. J., Eds.), pp. 176-231, Chapman & Hall, N.Y.

Hillis, D. M. and Bull, J. J. (1993). An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. *Syst. Biol.* 42: 182-192.

Hirt, R. P., Logsdon, J. M. Jr., Healy, B., Dorey, M. W., Doolittle, W. F. and Embley, T. M. (1999). Microsporidia are related to fungi: evidence from the largest subunit of RNA polymerase II and other proteins. *Proc. Natl Acad. Sci. USA* 96: 580-585.

Hoot, S. B., Kadereit, J. W., Blattner, F. R., Jork, K. B., Schwarzbach, A. E. and Crane, P. R. (1997) Data congruence and phylogeny of the Papaveraceae s.l. based on four data sets: atpB and rbcL sequences, trnK restriction sites, and morphological characters. *Syst. Bot.* 22: 575-590.

Hoot, S. B., Culham, A., and Crane, P. R. (1995). The utility of *atpB* gene sequences in resolving phylogenetic relationships: Comparison within *rbcL* and 18S ribosomal DNA sequences in the Lardizabalaceae. *Ann. Mo. Bot. Gard.* 82: 194–207.

Hoot, S. B., Megallon-puebla, S. and Crane, P. R. (1999). Phylogeny of basal eudicots based on three molecular data sets: *atpB, rbcL,* and 18S nuclear ribosomal DNA sequences. *Ann. Mo. Bot. Gard.* 86: 1–32.

Huelsenbeck, J. P., Bull, J. J. (1996). A likelihood ratio test for detection of conflicting phylogenetic signal. *Syst. Biol.* 45: 92–98.

Iwabe, N., Kuma, K-I., Kishino, H., Hasegawa, M. and Miyata, T. (1991). Evolution of RNA polymerases and branching patterns of the three major groups of archaebacteria. *J. Mol. Evol.* 32: 70-78.

Jokerst, R. S., Weeks, J. R., Zehring, W. A. and Greenleaf, A. L. (1989). Analysis of the gene encoding the largest subunit of RNA polymerase II in Drosophila. *Mol. Gen. Genet.* 215: 266-275.

Källersjö, M., Farris, J. S., Chase, M. W., Bremer, B., Fay, M. F., Humphries, C. J., Petersen, G., Seberg, O. and Bremer, K. (1998). Simultaneous parsimony jackknife analysis of 2538 *rbcL* DNA sequences reveals support for major clades of green plants, land plants, seed plants and flowering plants. *Plant Syst. Evol.* 213: 259-287.

Källersjö, M., Albert, V. A., and Farris, J. S. (1999). Homoplasy increases phylogenetic structure. *Cladistics* 15: 91-93.

Karol, K. G., McCourt, R. M., Cimino, M. T. and Delwiche, C. F. (2001). The closest living relatives of land plants. *Science* 294: 2351-2353.

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

Kumar, S., Tamura,K., Jakobsen, I. B. and Nei, M. (2001). MEGA2: molecular evolutionary genetics analysis software. *Bioinformatics* 17:1244-1245.

Kuzoff, R. K. and Gasser, C. S. (2000). Recent progress in reconstructing angiosperm phylogeny. *Trends Plant Sci.* 5: 330-336.

Larson, A. (1994). The comparison of morphological and molecular data in phylogenetic systematics. *In* "Molecular Ecology and Evolution: Approaches and Applications" (Schierwater, B., Streit, B., Wagner, G. P. and DeSalle, R., Eds.), pp. 371–390, Birkhaüser Verlag, Basel, Switzerland.

Liu, Y., Whelen, S. and Hall, B. D. (1999). Phylogenetic relationships among ascomycetes: Evidence from an RNA polymerase II subunit. *Mol. Biol. Evol.* 16: 1799-1808.

Lloyd, A. T. and Sharp, P. M. (1992). CODONS: a microcomputer program for codon usage analysis. *J. Hered.* 83: 239-240.

Loconte, H. (1996). Comparison of alternative hypotheses for the origin of angiosperms. *In* "Flowering Plant Origin, Evolution and Phylogeny" (Taylor, D. W. and Hickey, L. J., Eds.), pp. 267-285, Chapman & Hall, N.Y.

Loomis, W. D. (1974). Overcoming problems of phenolics and quinones in the isolation of plant enzymes and organelles. *Meth. Enzymol.* 31: 528–545.

Lutzoni, F. and Vilgalys, R. (1995). Integration of morphological and molecular data sets in estimating fungal phylogenies. *Can. J. Bot. (Suppl.)* 73: S649–S659.

Matheny, P. B., Liu, Y. J., Ammirati, J. F. and Hall, B.D. (2002). Using RPB1 sequences to improve phylogenetic inference among mushrooms (Inocybe, Agaricales). *Am. J. Bot.* 89: 688-698.

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

Magallón, S. and Sanderson, M. J. (2002). Relationships among seed plants inferred from highly conserved genes: sorting conflicting phylogenetic signals among ancient lineages *Am. J. Bot.* 89:1991-2006.

Mitra, D. and Kootstra, A. (1993). Isolation of RNA from apple skin. *Plant Mol. Biol. Rep.* 11: 326-332.

Miyamoto, M. M. and Fitch, W. M. (1995). Testing species phylogenies and phylogenetic methods with congruence. *Syst. Biol.* 44: 64–76.

Nandi, O. I., Chase, M. W. and Endress, P. K. (1998). A combined cladistic analysis of angiosperms using *rbcL* and non-molecular data sets. *Ann. Mo. Bot. Gard.* 85: 137–212.

Nawrath, C., Schell, J. and Koncz, C. (1990). Homologous domains of the largest subunit of eukaryotic RNA polymerase II are conserved in plants. *Mol. Gen. Genet.* 223: 65-75.

Nickerson, J. and Drouin, G. (2003). The sequence of the largest subunit of RNA polymerase II is a useful marker for inferring seed plant phylogeny. In press

Nixon, K. C., Crepet, W. L., Stevenson, D. W and Friis, E. M. (1994). A reevaluation of seed plant phylogeny. *Ann. Mo. Bot. Gard.* 81: 484–533.

Otero, G., Fellows, J., Li, Y., de Bizemont, T., Dirac, A. M. G., Gustafsson, C. M., Erdjument-Bromage, H., Tempst, P. and Svejstrup, J. Q. (1999). Elongator, a multisubunit component of a novel RNA polymerase II holoenzyme for transcriptional elongation. *Mol. Cell* 3: 109-118.

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. *Mol. Biol. Evol.* 17: 1131-1145.

Page, R. D. M. (1996). TREEVIEW: An application to display phylogenetic trees on personal computers. *Comput. Appl. Biosci.*12: 357-358.

Pamilo, P. and Bianchi, N. (1993). Evolution of the Zfx and Zfy genes: rates and interdependence between the genes. *Mol. Biol. Evol.* 10: 271-281.

Parkinson, C. L., Adams, K. L. and Palmer, J. D. (1999). Multigene analyses identify the three earliest lineages of extant flowering plants. *Curr. Biol.* 9: 1485-1488.

Pühler, G., Leffers, H., Gropp, F., Palm, P., Klenk, H-P., Lottspeich, F., Garrett, R. A. and Zillig, W. (1989). Archaebacterial DNA-dependent RNA polymerases testify to the evolution of the eukaryotic nuclear genome. *Proc. Natl. Acad. Sci. USA* 86: 4569-4573.

Qiu, Y-L, Chase, M. W., Les, D. H., Parks, C. R. (1993). Molecular phylogenetics of the Magnoliidae: cladistic analyses of nucleotide sequences of the plastid gene *rbcL*. *Ann. Mo. Bot. Gard.* 80: 587–606.

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: from mitochondrial, plastid, and nuclear genomes. *Nature* 402: 404-407.

Robinson-Rechavi, M., Gout, M., Gauthier, C. and Mouchiroud, D. (1998). Sensitivity of the relative-rate test to taxonomic sampling. *Mol. Biol. Evol.* 15: 1091-1098.

Rydin, C., Källersjö, M. and Friist, 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. *Int. J. Plant Sci.* 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. *Mol. Biol. Evol.* 17: 782-797.

Savolainen, V., Chase, M. W., Hoot, S. B., Morton, C. M., Soltis, D. E., and Bayer, C. (2000). Phylogenetics of Flowering Plants Based on Combined Analysis of Plastid *atpB* and *rbcL* Gene Sequences. *Syst. Biol.* 49: 306-362.

Savolainen, V., Spichiger, R. and Manen, J-F. (1997) Polyphyletism of Celastrrales deduced from a chloroplast noncoding DNA region. *Mol. Phylogenet. Evol.* 7: 145-157.

Seelanan, T., Schnabel, A. and Wendel, J. (1997). Congruence and consensus in the cotton tribe (Malvaceae). *Syst. Bot.* 22: 259–290.

Shultz, J. W and Regier, J. C. (2000). Phylogenetic analysis of anthropods using two nuclear protein-encoding genes supports a crustacean + hexapod clade. *Proc. R. Soc. Lond., B, Biol. Sci.* 267: 1011-1019.

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

Soltis, D. E. and Soltis, P. S. (2000). Contributions of plant molecular systematics to studies of molecular evolution. *Plant Mol. Biol.* 42: 45-75.

Soltis, D. E., Soltis, P. S., Mort, M. E., Chase, M. W., Savolainen, V., Hoot, S. B. and Morton, C. M. (1998). Inferring complex phylogenies using parsimony: an empirical approach using three large DNA datasets for angiosperms. *Syst. Biol.*

47: 32–42.

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, *rbcL*, and *atpB* sequences. *Bot. J. Linn. Soc.* 133: 381-461.

Soltis, D. E., Soltis, P. S., Nickrent, D. L., Johnson, L. A., Hahn, W. J., Hoot, S. B., Sweere, J. A., Kuzo, R. K., Kron, K. A., Chase, M. W., Swensen, S. M., Zimmer, E. A., Chaw, S. M., Gillespie, L. J., Kress, W. J. and Sytsma, K. J. (1997) Angiosperm phylogeny inferred from 18S ribosomal DNA sequences. *Ann. Mo. Bot. Gard.* 84: 1–49.

Soltis, D. E., Soltis, P. S. and Zanis, M. J. (2002). Phylogeny of seed plants based on evidence from eight genes. *Am. J. Bot.* 89: 1670-1681.

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

Stiller, J. W. and Hall, B. D. (1998). Sequences of the largest subunit of RNA polymerase II from two red algae and their implications for rhodophyte evolution. *J. Phycol.* 34: 857-864.

Stiller, J. W. and Hall, B. D. (1997). The origin of red algae: Implications for plastid evolution. *Proc. Natl. Acad. Sci. USA* 94: 4520-4525.

Stokstad, E. (2002). Fossil Plant Hints How First Flowers Bloomed. *Science* 296:821.

Strimmer, K. and Von Haeseler, A. (1996). Quartet puzzling: a quartet maximum likelihood method for reconstructing tree topologies. *Mol. Biol. Evol.* 13: 964-969.

Sun, G., Ji, Q., Dilcher, D. L., Zheng, S., Nixon, K. C. and Wang, X. (2002). Archaefructaceae, a new basal angiosperm family. *Science* 296: 899-904.

Swofford, D. L. (1998). PAUP: phylogenetic analysis using parsimony. Version 4.0b2. Sinauer, Sunderland, Mass.

Thompson, J.D., Higgins, D.G., and Gibson, T.J. (1994). CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22: 4673–4680.

Thuriaux, P. and Sentenac, A. (1992) *in* "The Molecular and Cellular Biology of the Yeast, Saccharomyces: Gene Expression" (Jones, E. W., Pringle, J. R. and Broach, J. R., Eds.), Vol. II, pp. 1–48, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

Treich, I., Riva, M. and Sentenac, A. (1991). Zinc-binding subunits of yeast RNA polymerase. *J. Biol. Chem.* 266: 21971-21976.

Upchurch, G. R., Crane, P. R. and Drinnan, A. N. (1994). The megaflora from the Quantico locality (Upper Albian), lower Cretaceous Potomac group of Virginia. *Va. Mus. Nat. Hist. Mem.* 4: 1-57.

Walker, J. W., Brenner, G. J. and Walker, A. G. (1983). Winteraceous pollen in the lower Cretaceous of Israel: early evidence of a magnolialean angiosperm family.

Science 220: 1273-1275.

Wang,C-S. and Vodkin L.O. (1994). Extraction of RNA from tissues containing high levels of procyanidins that bind RNA. *Plant Mol. Biol. Rep.*12: 132-145.

Winter, K-U., Becker, A., Munster, T., Kim, J. T., Saedler, H. and Thiessen, G. (1999). MADS-box genes reveal that gnetophytes are more closely related to conifers than to flowering plants. *Proc. Natl. Acad. Sci. USA* 96: 7342-7347.

Woychik, N. A. (1998). Fractions to functions: RNA polymerase II thirty years later. *Cold Spring Harb. Symp. Quant. Biol.* 63: 311-317.

Young, R. A. (1991). RNA polymerase II. *Ann. Rev. Biochem.* 60: 689-715.

Zillig, W., Klenk, H-P., Palm, P., Puhler, G., Gropp, F., Garrett, R. A. and Leffers, H. (1989). The phylogenetic relations of DNA-dependent RNA polymerases of archaebacteria, eukaryotes, and eubacteria. *Can. J. Microbiol.* 35: 73-80.

Zurawski, G., Bottomley, W., and Whitfeld, P. R. (1982). Structure of the genes for the beta and epsilon subunits of the spinach chloroplast ATPase indicates a dicistronic mRNA and an overlapping translation stop/start signal. *Proc. Natl. Acad. Sci. USA* 79: 6260–6264.

Appendix

Appendix A. Protein sequence alignment of sixteen RPB1 sequences sequenced in this study. All amino acids identical to the *Taxus* are represented by a dot and a dash represents a gap in the alignment. The conserved domains A to G are labeled based on the study of Nawrath *et al.* (1990).

```
                         Domain A
Taxus          HLELAKPMYH IGFLKTILSI LRCVCFNCSR ILADEDD-HR FKQALKVKNP
Thuja          .......... .......... .......... .....E.-.. ..........
Podocarpus     .......... .......... .......... .......-.. ..........
Gingko         .......... ......V... .......... .....E.-.. ......IR..
Drimys         ........F. ......V... M........K V....E.-.K ....M.....
Saruma         ........F. ......V... M........K .......-.. ...S...R..
Liriodendron   ........F. ......V... M........K ..V..EE-.K ....Q.IR..
Ceratophyllum  ........F. ......VCNL M..I.....K .....E.-.K ....V.I...
Beta           ........F. ......V... M........K .....E.-.K ......IR..
Persea         ........F. ...M..V.T. M........K ..V.-..-LK ...SQ.IR..
Pisum          Y.......F. ......V... M........K .....Q.-.K .....RI...
Nicotiana      ........F. ...M.PV... .........K .....E.-PK ....MRIR..
Papaver        ........F. ...M.IV... M........K .....E.-.K ...SQRI...
Asparagus      ........F. ...M..V.A. M........K .......-V. ......IR..
Illicium       ........F. ......V... M.S......H ....QE.-.. ......IR..
Ephedra        .......... ......V... .......... .....E.-.. ....M.....

Taxus          KHRLKKVLDC CKNKTKCEGG DVIDEDQGQD -GD--DEVKK KRQGGCGAQQ
Thuja          ....R..... .......... .......... -.--...... ..........
Podocarpus     .......... .......... .A.....V.. -..--E.... ..........
Gingko         ....R..... .......... .E..D..... -.E--EAA.. ..H.......
Drimys         ....R....A .......D.. .E..-V.... -SE--EP... P-R.......
Saruma         ....R....A ....Q..... .E..-V...E -S.--EP... .-R.......
Liriodendron   ....R....A .....R.... .D.E-VRD.. -LE--EP... R-R.......
Ceratophyllum  ....R..M.A ..S.Q..... .EM.-V.... -E.--EP-.. .-R....S..
Beta           .N...RI.EA ....K..... .SL.-V.D.. -Q.DGESA.. T-R.......
Persea         .....R.F.A ....Q..... .EM.-V...G -SE--EP... S-R.......
Pisum          .N..R.I..A .......... .DL.-LP... -TE--EP... S-R.......
Nicotiana      .N....M..A .......... .E..-VR.EE -PE--AP... S-R.......
Papaver        ....R....A .......D.. .E..-V.DE. -LE--KP... S-R.......
Asparagus      .N...RIY.A ..S.K..A.. .D..-V...Q DS.--EP... S-R.......
Illicium       ....R..... ..S.Q..... .D.....A.. -SE--.G... .GH.......
Ephedra        .....R.... .......... .S.E...A.. -A.--E.... .PH....S..

Taxus          PKIYIDGMKI IAEYKAPRKK ADDPEQLMPE PVERKQQLSA ERVLSVLKRI
Thuja          .......... M......... .......... .......... ..........
Podocarpus     .......... V......... ...Q...... .A......P. ....N.....
Gingko         ...S.....M V....L.... ..EQ...... .I......T. .K........
Drimys         ..LTVE...M ......TK.. N..Q-EQL.. .......... .....I....
Saruma         ..LTVE...M ......-K.. N..Q-EQL.. ........T. .K........
Liriodendron   ..LTVE...M ......T... N..Q-EQL.. .......... ..........
Ceratophyllum  ...AVE.... ...F...K.R S.EQ-DQL.. .......... .....I....
Beta           .TFT.E.... .....MTK.. S.EQ-EQL.. .A........ .K.......V
```

Persea	..LT.E...M	.....GS...	N.E--EPL..	....	....
Pisum	...T.E....	N.....Q...	S..Q-DQL..	......T...	....
Nicotiana	...S.....M	V....LQK..	S...-EQ...	....	.....I...V
Papaver	..LT.E...M	.....IQ...	T...-EQL..	........T.	....
Asparagus	.N.T.....M	V.....TK..	S...-EQL..	....	....
Illicium	...TVE...M	...F.QT...	S.EQ..ML..	.A........	.....I....
Ephedra	...T......	....R.....	..EQD.....	........T.	....NI....

Domain B

Taxus	TDEECQMLGL	NPKYSRPDWM	ILQVLPVPPP	PVRPSVMMDT	TSRSEDDLTH
Thuja	S.....L...	....A.....	....	....	....
Podocarpus	S........F	....A.....	....	....A.....	....
Gingko	S.....L...	....A.....	......I...	....	....
Drimys	S..D.RL...	D...A.....	......I...	....	S.........
Saruma	S..D.IL...	S...A.....	......I...	....	S.........
Liriodendron	S..D.LL...	....A.....	......I...	....	S.........
Ceratophyllum	S..D.VL..F	....A.....	......I...	....	S.........
Beta	S..D..LM.F	....A.....	......I...	....	S.........
Persea	S..D.ML...	....A.....	......I...	....	S.........
Pisum	S..D..L...	....A....L	......I...	....	S.........
Nicotiana	S..D.LL...	..E.A.....	...A..I...	....	S.........
Papaver	S..D.IL...	....A.....	......I...	....	SA........
Asparagus	S..D.LL...	....A.....	......I...	....	SA........
Illicium	S..D.LL...	....A.....	......I...	....	S.........
Ephedra	S........F	D.QFA.....	......I...	........NS	....

Domain B

Taxus	QLAMIIRHNE	TLKKQEKTGA	PAHIISEFAQ	LLQFHIATYF	DNELPGQPRA
Thuja	....	N.........	....	....	....
Podocarpus	....	....	....	....	....
Gingko	....	N..R..RN..	.....T....	....	..D.......
Drimys	....	N.RR..RN..	....	....	....
Saruma	....	N.RR...N..	....	....	....
Liriodendron	....	N.RR..RN..	....	....	....
Ceratophyllum	....	N.RR..RN..	....	....	.....-....
Beta	.........D	N.RR..RN..	....	....	....
Persea	....	N.R...RN..	....	....	....
Pisum	....	N..R..RN.S	....	....	......L...
Nicotiana	....	N..R..RN..	....	....	..D.......
Papaver	....	N.R...RN..	.....L....	....	....
Asparagus	....	N.RR..RN..	.....T....	....	....
Illicium	....	N.RR..RN..	....	....	....
Ephedra	....	.......N..	....VT....	.....V....	....

Domain C

Taxus	TQRSGRPIKS	ICSRLKAKEG	RIRGNLMGKR	VDFSARTVIT	PDPNINIDEL
Thuja	....	....	....	....	....
Podocarpus	....	....	....	....	....
Gingko	....	..N.......	....	....	........Q.
Drimys	....	....	....	....	...T......
Saruma	....	....	....	....	...T......
Liriodendron	....	....	....	....	...T......
Ceratophyllum	....	....	....	....	...T......
Beta	....	..T.......	....	....	....
Persea	....	....	....	....	...T......
Pisum	....	....	....	....	...T....Q.

```
Nicotiana       .......... ......S... .......... .......... ...T....Q.
Papaver         .......... .......... .......... .......... ...T......
Asparagus       .......... .......... .......... .......... ...T......
Illicium        .......... .......... .......... .......... ...T......
Ephedra         .......... .......... .......... .......... ........Q.

                      Domain C
Taxus           GVPWSIALNL TYPETVTPYN IERLKELVEY GPHPPPGKTG AKYIIREDGQ
Thuja           .......... .......... .......... .......... ..........
Podocarpus      .......... .......... .......... .......... ..........
Gingko          .......... .......... .......... .......... ..........
Drimys          .......... .......... .......... .......... ..........
Saruma          .......... .......... .D........ .......... ..........
Liriodendron    .......... .......... .......... .......... ..........
Ceratophyllum   .......... .......... .......... .......... ..........
Beta            .......... .......... ........D. .......... ......D...
Persea          .......... .......... .......... .......... ..........
Pisum           ....T..... .......... .......... .......... ......D...
Nicotiana       .......... .......... .......... .......... ......D...
Papaver         .......... .......... .......... .....V.... ..........
Asparagus       .......... .......... .......... .......... ..........
Illicium        .......... .......... .......... .......... ..........
Ephedra         .......... ....V..... .......... .......... ......D...

                                                       Domain D
Taxus           RLDLRYLKKS SDHHLELGYK VERHLNDGDF VLFNRQPSLH KMSIMGHRIR
Thuja           .......... .......... .......... .......... ..........
Podocarpus      .......... .......... .......... .......... ..........
Gingko          .......... .......... .......... .......... .........K
Drimys          .......... .......... .......... .......... .........K
Saruma          .......... .......... .........Y .......... .........K
Liriodendron    .......... .......... .......... .......... .........K
Ceratophyllum   .......... .......... .......... .......... .........K
Beta            .......... .......... .......... .......... .........K
Persea          ......S... .......... .......... .......... .........K
Pisum           .......... .......... .......... .......... .........K
Nicotiana       .......... ..Q....... .......... .......... .........K
Papaver         .......... .......... .......... .......... .........K
Asparagus       .......... .......... .......... .......... .........K
Illicium        .......... .......... .........Y .......... .........K
Ephedra         ........R. .......... .......... .......... .........K

                                  Domain D
Taxus           IMPYSTFRLN LSVTSPYNAD FDGDEMNMHV PQSFETRAEV LELMMVPKCI
Thuja           .......... .......... .......... .......... ..........
Podocarpus      .......... .......... .......... .......... ..........
Gingko          .......... .......... .......... .......... ..........
Drimys          .......... .......... .......... .......... ..........
Saruma          .......... .......... .......... .......... ..........
Liriodendron    .......... .......... .......... .......... ..........
Ceratophyllum   .......... .......... .......... .......... ..........
Beta            .......... .......... .......... .......... -.........
Persea          .......... .......... .......... .......... ..........
Pisum           .......... .......... .......... .......... ..........
Nicotiana       .......... .......... .......... .......... ..........
Papaver         .......... .......... .......... .......... ..........
```

```
Asparagus      .......... .......... .......... .......... ..........
Illicium       .......... .......... .......... .......... ..........
Ephedra        .......... .......... .......... .........- ..........
```

```
Taxus          VSPQSNRPVM GIVQDTLLGC RKITKRDTFI EKDVFMNILM WWEDFDGKIP
Thuja          .......... .......... .......... .......... ..........
Podocarpus     .......... .......... .......... .......... ..........
Gingko         .......... .......... ..V....... .......... ..........
Drimys         .......... .......... .......... .......... ........V.
Saruma         .......... .......... .......... .......... ..........
Liriodendron   .......... .......... .......... .......... ..........
Ceratophyllum  .......... .......... .......... .......... ..........
Beta           .......... .......... .......... .......... ..........
Persea         .......... .......... .......... .......... ..........
Pisum          ....A..... .....S.... .......... A......... ........V.
Nicotiana      ....A..... .......... .......... .......... ........V.
Papaver        .......... .......... .......... .......... ........V.
Asparagus      .......... .......... .......... .......... ........V.
Illicium       .......... .......... .......... .......... ..........
Ephedra        .......... .......... .......... .......... ........V.
```

Domain E

```
Taxus          APTILKPRPL WTGKQVFNLI IPKQINLIRQ SAWHAESETG FITPGDTIVR
Thuja          .......... .......... .........Y .......... .......C..
Podocarpus     S.A....... .......... .........Y .......... .L.A...V.H
Gingko         S......... .......... ..R......Y ....S..... .......C..
Drimys         ..A....... .......... .......... ....S.T... .......Q..
Saruma         ..AV...... .......... .......L.T .......... .......Q..
Liriodendron   ..AV...... .......... .........T ....S.A... .......Q..
Ceratophyllum  .......... .......... .........T ....S.A... .......Q..
Beta           .......... .......... .........T ....S.A... .......Q..
Persea         ..A....... .......... ........-T ....S..... .......Q..
Pisum          ......E... .......... .........Y .S..N.A.R. P......M..
Nicotiana      ..A....... .......... .......L.Y ....ND..K. P......Q..
Papaver        .......... .......... .......L.T ....S.N... .......Q..
Asparagus      T.A....... .......... .......S.A ....S..... .......Q..
Illicium       ..A....... .......... .........T ....S..... .......H..
Ephedra        ..A.....I. .......... .......V.Y ....N..DR. HL.L...V..
```

```
Taxus          IEKGEVLSGT LCKKTLGTSS GSLIHVIWEE VGPDAARKFL GHTQWLVNYW
Thuja          ......I... .......... .......... .......... ..........
Podocarpus     .......... .........T .......... .......... S.........
Gingko         .......... .......... .......... .......... ..........
Drimys         ..R..L.C.. .........T .......... .......... ..........
Saruma         ..R..L.A.. ....A....T .......... .......... ..........
Liriodendron   ..R..L.A.. .........T .......... .......... ..........
Ceratophyllum  ..R..L.A.. .........T .......... .......... ..........
Beta           ..R..L.A.. .........T .......... .......... ..........
Persea         .....L.A.. .........T .......... .......... ..........
Pisum          .....L.T.. ........GT ......V... .......... ..........
Nicotiana      .....L.T.. .........T .......... .......... ..........
Papaver        ..R..L.T.. ....A....T .......... .......... ..........
Asparagus      .....L.T.. .........T ........-- .......... ..........
Illicium       .....L.... .........T .......... .......... ..........
```

```
Ephedra        ......IT.. .......A.. .......... .......... ..........

Taxus          LLQQGFSIGI GDTIADAATM EVINETISKA KHEVKQLIKA AQDKQLEAEP
Thuja          .......... .......... .......... .N........ ..........
Podocarpus     .......M.. .......... ........R. .N........ ..E.......
Gingko         .......... .......... .T........ .N........ ..E.A.....
Drimys         ...N...... .......... DT.S...... .ND..E..M. ....A.....
Saruma         ...NA..... .......... .K..D..T.. .N...E.... .HE.......
Liriodendron   ...N...... ....:..... .K........ .N...E.... ..........
Ceratophyllum  ...N...... .......S.. .K........ .N...E.... ..E.......
Beta           ...N...... .......S.. .K........ .N...E.... ..E.......
Persea         ...N...... .......... .K.....L.. .S...D.... ..E.S.....
Pisum          ...NA..... .......S.. .T..Q...Q. .DK.....RE ..E.K.....
Nicotiana      ...NA..... .......S.. .K......N. .SK..E.... ..E.......
Papaver        ...N...... .......S.. .K..D..... .N...D..RM ..E.S..P..
Asparagus      ...NA..... .......... .K........ .ND..D...L .......P..
Illicium       ...N...... .......... .K........ .N...E.... ..E.......
Ephedra        .......M.. .......T.. DT.....QD. .IK.QE..EK YMAHK..Q..

                                                        Domain F
Taxus          GRTMMESFEN RVNQVLNKAR DDAGSSAQKS LSESNNLKAM VTAGSKGSFI
Thuja          .......... .......... .......... .......... ..........
Podocarpus     .......... .......... .......... .......... -.........
Gingko         .......... .......... ........R. .......... ..........
Drimys         ....T..... .......... ........N. .......... ..........
Saruma         .......... .......... .......... .......... ..........
Liriodendron   .......... .......... .......... .......... ..........
Ceratophyllum  .......... .......... .......... .......... ..........
Beta           .......... .......... .......... .......... ..........
Persea         .......... .......... .......... .......... ..........
Pisum          .....D.... ....T..... ....N..... .......... ..........
Nicotiana      .......... .......... .......E.. .......... ..........
Papaver        .......... .......... .......... .......... ..........
Asparagus      .......... .......... .......... .......... ..........
Illicium       .......... .......... .......... .......... ..........
Ephedra        ...LL..... Q......... ....N...R. .......... ..........

                                   Domain F
Taxus          NISQMTACVG QQNVEGKRIP YGFDGRTLPH FTKDDKGPES RGFVENSYLR
Thuja          .......... .......... .......... .......... ..........
Podocarpus     .......... .......... .......... .......... ..........
Gingko         .......... .......... .......... .....N.... ..........
Drimys         .......... .......... ...I...... .....N.... ..........
Saruma         .......... .......... ...IN..... ..Q..N.... ..........
Liriodendron   .......... .......... ...ID..... .....Y.... ..........
Ceratophyllum  .......... .......... ...ID..... .....Y.... ..........
Beta           .......... .......... ...ID..... .....Y.... ..........
Persea         .......... .......... ...N...... .....Y...- ..........
Pisum          .......... .......... F..ID..... .....Y.... ..........
Nicotiana      .......... .......... F..ID..... .....Y.... ..........
Papaver        .......... .......... ...ID..... .....Y.... ..........
Asparagus      .......... .......... ...I...... .....Y.... ..........
Illicium       .......... .......... ...ID..... .....Y.... ..........
Ephedra        .......... .......... ...Q.....- .V........ ..........
```

Domain F

Taxus	GLTPQEFFFH	AMGGREGLID	TAVKTSETGY	IQRRLVKAME	DIMVKYDGTV
Thuja	....	....	....	....	....
Podocarpus	....	....	....	....	....
Gingko	....	....	....	....	....
Drimys	....	....	....	....	....
Saruma	....	....	....	....	....
Liriodendron	....	....	....	....	....
Ceratophyllum	....	....	....	....	....
Beta	....	....	....	....	....
Persea	....	....	....	....	....
Pisum	....	....	....	....	....
Nicotiana	....	....	....	....	....
Papaver	....	....	....	....	....
Asparagus	.........Y	....	....	....	....
Illicium	....	....	....	....	....
Ephedra	....	....	....	....	.......A..

Taxus	RNSLGDVIQF	LYGEDGMDAV	WIESQKLDSL	KMKKKEFENV	YRYEIDQENW
Thuja	....	....	....	...R......	....
Podocarpus	....	....	....	...R......	.K..V.....
Gingko	....	....	....	...R......	.K....N...
Drimys	....	........S.	....	....A..D..	F.....ND..
Saruma	....	....	....	....A..DK.	F...F.DA..
Liriodendron	....	....	....	....R..D..	F...F.NA..
Ceratophyllum	....	....	....	....R..D..	F...F.NA..
Beta	....	....	....	....R..D..	F...F.NA..
Persea	....	........S.	....	....R..L..	......N...
Pisum	....	....	...T......	....TD.DRA	F...F.D...
Nicotiana	....	....	...T......	.A..ST.DDL	.A....DP..
Papaver	....	....	....	....P..DK.	..F.F.DD..
Asparagus	....	....	....	.......N..	......DNQ.
Illicium	....	....	....	....	F.....ND..
Ephedra	....	....	R.......H.	.......DET	FKH.LG....

Taxus	NPSYMLPEHV	DDLKTIREFQ	NVFDAELQKL	ESDRRQLGTE	IATTGDNSWP
Thuja	....	....	....	....	..PS......
Podocarpus	..I.......	E.........	......V...	....	..PS....C.
Gingko	..G...AD..	E........R	......V...	.A........	..PS......
Drimys	.....V....	E........R	......V...	.A..Y...S.	....
Saruma	....	E........R	S.....V...	.A..Y.....	.T........
Liriodendron	....	E........R	......V...	.A..Y.....	....
Ceratophyllum	....	E........R	......V...	.A..Y.....	....
Beta	....	E........R	......V...	.A..Y.....	....
Persea	..G.......	E........R	......V...	.A..Y.....	....
Pisum	K.T...E.P.	E........R	...E..V...	DA..F..A..	......S.L.
Nicotiana	........AI	E.......IR	S.....V...	QA..K.....	..V.......
Papaver	T.N..QSDY.	E...N....R	H.....VK..	AD..L.....	.T........
Asparagus	D.........	E.....P..R	......V...	.A..F....D	.......T..
Illicium	....	K........R	......V...	.A..F.....	....
Ephedra	H.T.....MA	..I.N.S...	D.LV..YK..	QL......M.	..P..E....

```
Taxus          MPVNLKRLIW NAQKTFKIDP RKRSDMHPME IVEAIDKLQE RLKVVPGDDL
Thuja          .......... .......... .......... .......... .........-
Podocarpus     .......... .......... ..P....... ......R... ..........
Gingko         .......... .........L ..P....... ....V..... ..........
Drimys         .......... .......V.L .RP.G..... ....V..... .........P
Saruma         .......... .......V.L .RP....... ....V..... .......E..
Liriodendron   .......... .......V.L .RP....... ....V..... .........F
Ceratophyllum  .......... .......V.L .RP....... ....V..... .........F
Beta           .......... .......V.L .RP....... ....V..... .........F
Persea         .......... .......V.L .RP....... ....V..... .........F
Pisum          L......... .......V.F .RP....... .......... ..R.......
Nicotiana      L...IQ..VL .........F .R........ ....V..... .........Y
Papaver        .....R.... .......V.. .RP....... .......... .........S
Asparagus      .......... .......V.. .RP....... ....V..... K.R......F
Illicium       .......... .......V.L .RP....... ....V..... .........F
Ephedra        L......... ....I..V.. ..T...Q.L. ..DT...... .........P
```

Domain G

```
Taxus          MSIEAQKNAT LFFNILLRST FASKRVLKEY HLTKEAFEWV VGEIESRFLQ
Thuja          I......... .......... .......... .......... ..........
Podocarpus     I......... .......... .......... .......... ..........
Gingko         .......... .......... .......... R..R...... I.........
Drimys         I.M....... .......... .......... R..R...... I.........
Saruma         ..M....... .......... .S........ R..R...... I.........
Liriodendron   L.M....... .......... .......... R..R...... I.........
Ceratophyllum  L.M....... ........T. .......... R..R...... I.........
Beta           L.M....... ........T. .......... R..R...... I.........
Persea         L.M....... .......... .......... R..R...... I.........
Pisum          L.Q....... .L........ .......E.. R.SR...... ..........
Nicotiana      L.M....... .........A L......... R.SR...D.. I.........
Papaver        ..M....... .......... .........H ...R...D.. I..V......
Asparagus      L.M....... .L........ .......... R..R...... I.........
Illicium       L.M....... .......... .......... R..R...... ..........
Ephedra        L.F....... .........V .......... ...R...D.. I..V.N....
```

Domain G

```
Taxus          SLVAPGEMIG CVAAQSI
Thuja          .......... .......
Podocarpus     .......... .......
Gingko         .......... .......
Drimys         .......... .......
Saruma         .......... .......
Liriodendron   .......... .......
Ceratophyllum  .......... .......
Beta           .......... .......
Persea         .......... .......
Pisum          ....S..... .......
Nicotiana      .......... .......
Papaver        .......... .......
Asparagus      .......... .......
Illicium       .......... .......
Ephedra        .......... .......
```

Appendix B. Nucleotide sequence alignment of sixteen *rpb1* genes sequenced in this study. All nucleotides identical to the *Taxus* are represented by a dot. A dash represents a gap and a question mark represents an uncertain nucleotide in the alignment.

```
Taxus         CACCTGGAGC TGGCCAAGCC AATGTATCAT ATTGGGTTTT TGAAGACAAT TCTTTCGATT
Thuja         .....T.... .C........ C......... .......... .......C.. ..........
Podocarpus    ...T.....T ....T..... T......... .......... .......T.. ...G..A...
Gingko        .....T..AT ....A..A.. C......... .......... .......TG. ......A..A
Drimys        ..T..T.... .T..T..A.. T....TC..C ..A..C...C .T.....CG. A..CAGT...
Saruma        .....C.... .C..G..... .....T.... .....A...C .T.....CG. G..C..T...
Liriodendron  .......... .C.....A.. C....TC... ........C. ....A..CG. G..C..T..C
Ceratophyllum ..TT.A..A. .T..A..A.. C....T...C .....A.... ........G. .TG.AATC.C
Beta          .....T.... .C..T..A.. T....T.... .....A.... .......TG. G...AGT...
Persea        .....T.... .T.....A.. T....T.... .........A .......TG. G...A.C..C
Pisum         T..T...... .T..A..A.. T....T.... .......... .......TG. ...CAGT..A
Nicotiana     .....T.... .A..A..A.. T....T.... ..C..A...A ....AC.TG. A...AGT..C
Papaver       ...T.A..A. .T..T..A.. T....T.... .........A ......TTG. ....AGT..A
Asparagus     .....C.... .C..A..... C....T.... .....A..CA .......TG. G..CG.T..C
Illicium      ...T...... .T..A..A.. C....TC... .....T..C. .A.....TG. ......C..C
Ephedra       .....T...T ....T..A.. T......... ..A..T..C. ....A..TG. ......A..C

Taxus         TTGCGTTGCG TATGCTTCAA CTGCTCAAGG ATTCTCGCAG ATGAGGATGA T---CATCGT
Thuja         ........T. .......... .......... .....T.... .......G.. .---......
Podocarpus    ........T. .G........ T..T..T... ...T.A.... .......... C---......
Gingko        .....C..T. .C........ T..T..GC.T .....T..T. .......G.. C---......
Drimys        A....C..T. .C........ T.....G.AA G.C..A..T. .......G.. C---..CAAA
Saruma        A....G..T. .C........ T.....C.A. .....T..G. .......... C---.....G
Liriodendron  A....C.... .G.....T.. T.....T.AA ..A..T.TT. .......A.. A---..CAAG
Ceratophyllum A....C..TA ....T..T.. T..T..G.A. .....T..G. .......A.. C---..CAAA
Beta          A.......T. .T........ T.....T.A. ........G. .......A.. C---..CAAA
Persea        A....C.... .G........ T.....C.A. .....A.TT. ..?....... C---.T.AAA
Pisum         A.......T. .T.....T.. ......T.A. .....T..T. .....C.A.. .---...AAG
Nicotiana     C.T....... .C........ ......G.AA .....G.... .......G.. .---.C.AAA
Papaver       A....A..T. .C........ ...T..T.AA ..C..T.... .......G.. .---..CAAG
Asparagus     A....C.... .T........ T.....C.A. ........C. .C.....C.. .---GTC..G
Illicium      A......CT. .T..T..T.. T.....CCAT .....A..T. .CC....G.. C---...A.A
Ephedra       C.C..A..T. .T........ T.....CC.C .....A..T. .......G.. .---......

Taxus         TTCAAACAAG CATTGAAAGT AAAGAACCCA AAACACCGGC TGAAAAAGGT ATTGGATTGT
Thuja         .......... .......... .......... ......A... .C.G...... T.........
Podocarpus    .......... ....A..... .......... ..G..TA... .C.....A.. T..A......
Gingko        .......... .G......A. C.G...T..C ......A... .C.G...... C..A..C...
Drimys        .....G.... .TA....... T..A..T... ..G..T..A. .A.G...... T......GCC
Saruma        .....G...T .T........ T.G......T .....T..C. ...G...... T......GCC
Liriodendron  .....G.... .TCA....A. T.GA..T... .....T..A. .C.GG..... T......GCA
Ceratophyllum ..T..G..G. .CG.A...A. T..A..T... ......A.A. ...G...A.. CA....CGC.
Beta          ..T..G.... ........A. T.G...T..G ...A.T...T ....G.GAA. ......AGCC
Persea        .....G...T .TCA....A. C.GA.....C ..G..TA.A. .A..G.GA.. T..T...GCA
Pisum         ..T..G.... .T....GGA. ......T... ...A.TA... ...G....A. T......GC.
Nicotiana     .....G.... .CA...G.A. ..G...T..T ...A..A... ........A. GC.....GC.
Papaver       .....G...T .TCA..GGA. ......T... .....T..A. .A.G...... TC.T...GCC
```

```
Asparagus       ..T..G..G. ..C.....A. T.GA.....G ..GA..A... ....G.G.A. T.AT...GCC
Illicium        .....G.... ........A. ..GA..T..T ......A.A. .A.G...A.. ...A.....C
Ephedra         .......... .TA....... T..A...... ..G...A... .T..G.GA.. CC.A..C...

Taxus           TGTAAGAATA AAACGAAATG TGAGGGTGGA GATGTCATTG ACGAGGATCA AGGACAAGAT
Thuja           ..C....... ....T..... ...A...... .......... .......C.. ...T......
Podocarpus      ..C....... ....A..... .......... ....CA.... .T..A..... ..TT......
Gingko          ........C. ....A..... .........G ....AG.... .T..T..... ...G......
Drimys          .....A..C. ....A..... C..T.....T ....AA..C. .T---.TA.. ...T......
Saruma          .....A..C. ..CAA..... ...A.....T ....AA..C. .T---.TC.. ...T.....G
Liriodendron    .....A..C. ....A.G... ...A..C..T ....AT.... .G---.TC.G ..AT......
Ceratophyllum   ..C..A.GC. ..CAA..... ...A.....C ....AA..G. .T---.TC.. ...T......
Beta            ..C..A..C. ...A...... ...A..A..T ...TC.T.A. .T---.TA.. G.AT..G...
Persea          .....A..C. .GCAA..... ...A..G..T ....AG..G. .T---.TC.. G..C....G.
Pisum           .....A..C. ....C..G.. ...A..A..T ....A.C... ..---CT..C T..T......
Nicotiana       ..C..A.... ....A..G.. C..A..A..G ..C.AA.... .T---.T..G ...CG....G
Papaver         .......... ....A..G.. ...T..G..T ....AA.... .T---.TC.. ..ACG.....
Asparagus       .....A.GC. .G.AA..... ..CA..C..G ....AT..C. .T---.T... ...T...C.G
Illicium        .....A.GC. ..CA...... C..A.....T ....A..... .T.....C.. ..CT......
Ephedra         .....A.... ....A..G.. ...A...... ...TCA.... .A........ ..CC......

Taxus           ---GGAGAT- -----GACGA GGTTAAAAAG AAACGGCAAG GAGGATGTGG TGCTCAACAG
Thuja           ---......- -----..... ......G... ..G..A.... ....T..... ......G...
Podocarpus      ---..G...- -----..G.. .......... ..G..T.... ....T..... ...C......
Gingko          ---..T..A- -----..A.C A.C...G... ..G..T..T. ....T..C.. ...G.....A
Drimys          ---TC...A- -----..ACC ...A..G... CCG---.GG. .G..T..... ...C......
Saruma          ---TCT...- -----..GCC T..G..G... ..G---.GG. .T..C..... G..C..G...
Liriodendron    ---TT...A- -----..ACC A..A..G... CGC---.G.. .G..T..... ..........
Ceratophyllum   ---.A....- -----..ACC .?.C..G..A ...---.G.. .G..T..... .T.C.....A
Beta            ---CAG...G ATGGA..GTC T.CC...... .CT---.GT. .T..C..... ...A......
Persea          ---TC...A- -----..ACC T..A..G... .GC---AG.. .G..T..... .........A
Pisum           ---ACC..G- -----..GCC .......... .GT---.GC. .T..C..... ......G..A
Nicotiana       ---CCT..A- -----.CACC T..C..G... .GT---AG.. .T..C..... ...A..G...
Papaver         ---TTG..A- -----A.ACC A..G..G... .GT------A ....GG..T. C.G.GCT..A
Asparagus       GATTC....- -----..ACC A..A..G... .GC---.G.. .C..T..... ......G...
Illicium        ---TC...A- -----..T.G T..A..G... ..GG.C..T. ....T..... ...C.....A
Ephedra         ---.CT...- -----..G.. ...G...... ..G.CT..T. ....C..... .T.G..G..A

Taxus           CCCAAGATCT ATATTGATGG TATGAAAATA ATTGCTGAAT ATAAAGCTCC TCGCAAGAAA
Thuja           .......... .C........ A......... ..G..C.... .......... A.........
Podocarpus      ........T. ....C..... A........T G......... .......... A..G......
Gingko          ..T.....T. CC........ .........G G....C..G. .....CT... A.....A..G
Drimys          ......T.AA CGG....A.. .........G .......... .C..G...A. GAAG......
Saruma          ..A...T.AA C.G....A.. ......G..G .......... ....G..AAA GAAG...---
Liriodendron    ..A...C..A C.G....A.. .........G .....A.... .C..G...A. AA.G..A..G
Ceratophyllum   ..A.....AG CCG....A.. ......G... .....A..G. TC..G..C.. .AAG....G.
Beta            ..G.CCT..A C......A.. .........T ..A....... .C..GATGA. AAAG......
Persea          ..A...T.AA CC.....A.. .........G .....A.... .C..G.G.T. AA.G......
Pisum           ..A......A C......G.. G........T .A...A..G. .........A AA.G......
Nicotiana       ..A..A..T. CA..A..... C........G G......... ....GCT..A GAAG..A..G
Papaver         .AACC..AGC TC.C.ATC.A AGGT.TG.A. ..GAT..C.G .AT.CAAAAT ?.AA.G...G
Asparagus       ..A..T..TA C......... .........G G.C..G..G. .C......A. GAAG......
Illicium        ..A.....AA C.G....G.. .........G .....A..G. TC...CAGA. ..........
Ephedra         .....A..TA CA..C..... .......... ..A..A.... ...G...... A.....A..G

Taxus           GCAGATGACC CAGAGCAGTT GATGCCAGAG CCTGTTGAGA GAAAACAGCA GCTATCAGCT
Thuja           .......... .......... ......G... .......... .......A.. .T....T...
```

```
Podocarpus    .......... AG......C. T.....T..A ....C...A. .......A.. ...TC.T...
Gingko        ........A. A...A...C. A......... ..CA.A.... .......... ...TA.T..A
Drimys        AAT....... A.---G.ACA .C.T..T..A ..C..A..A. .G..G..A.. A..T..T..A
Saruma        AAT....... A.---G..CA .C.T..T..A .....A..A. .......... A..TA.T..A
Liriodendron  AAT.....?. A.---G.ACA .C.T..T..A .....G..A. .......A.. ...T..T..A
Ceratophyllum T.C.....A. A.---G.TCA .C.T..T... ..A.....A. ....G..... A..T..T...
Beta          AGT.....G. AG---G.ACA AC.C..T... ....C..... ....G..... ...T..T...
Persea        AAT.....A- -----G..CC .C.T..T..A ........A. .......A.. ...T..T..A
Pisum         AGT.....T. AG---G.TCA .C.T..T..A .....A..A. .G......AC T..T..C..A
Nicotiana     AGT....... ..---G..CA .........A ..A.....A. .G.....A.. A..C..T...
Papaver       AA.AC....G ATCCCG.ACA .C.T..T..A ........A. .......A.. ...TA.T..A
Asparagus     T.T....... ..---G.ACA .C.T..T..A .......... .......A.. ...T..T..A
Illicium      T.C.....G. A...A...A. .C.A..G..A ..A.C...A. .......A.. A..T..G..G
Ephedra       ..T.....A. A...T..AC. C......... .....A.... .G........ ...TA....A

Taxus         GAGAGGGTTT TGAGTGTTTT AAAGCGAATA ACTGATGAGG AGTGTCAAAT GTTAGGATTG
Thuja         .......... .......... G.....C..T T......... ........T. ...G......
Podocarpus    .......... ...A...... ......T... T.......A. .A........ .C.G.....C
Gingko        ....A..... .......... G......... .GC..C.... ........T. ...G..CC..
Drimys        ..A......C .T...A..C. G...A.G... .G........ .C..C.G.T. ......G...
Saruma        ..A.A...CC .T......C. G..AA.G... .G...C.... .C..CATCC. C.....T...
Liriodendron  ..A......C .A......C. C.....G... .G........ .C..C.T.C. A..G..T...
Ceratophyllum ........G. .A..CA.A.. G...A..... .G......A. .T...GT.T. .C.......T
Beta          ..A.A...GC .C......C. G...A..G.G .G......A. .C......T. .A.G..T..C
Persea        ..A......C .T......C. C.....G... .G........ .C..CATGC. T.....T...
Pisum         ..A......C ........C. G..AA.G... .G........ .C..C..GT. ......C...
Nicotiana     ..A.....GC ....CA..C. G...A.GG.. .G......A. .T....TGT. .C.T..CC.A
Papaver       .......... .......CC. C.....G... .G...C.... .C...AT.T. ...G..T...
Asparagus     .........C .C......C. T...A.G... .GC.....A. .T..C.TCC. AC....T...
Illicium      .....A...C .A..CA.... G.....G... .G........ .C..CTTGC. .C.T..C...
Ephedra       ..A..A.... .A.ACA.C.. ....A..... .G......A. .A........ .C.T..G..T

Taxus         AATCCCAAAT ATTCTCGCCC AGATTGGATG ATTTTACAAG TGCTTCCAGT TCCTCCACCA
Thuja         ..C..A.... ..G....T.. G..C...... .......... .......... .........T
Podocarpus    ..C....... ..G....T.. .......... .......... .C.....C.. .........T
Gingko        .......... ..G....T.. C......... .....G..G. ........A. ......T..T
Drimys        G.C.....G. ..G.C..T.. .......... .......... .CT.G..TA. ......C..C
Saruma        .GC..A..G. ..G.C..A.. .......... ..C....... .T..G..TA. C.....T..T
Liriodendron  .....T..G. ..G.C..T.. .......... .......... .C..C..GA. ...C..T..T
Ceratophyllum .....A..G. ..G....... T......... .....G.... .C......A. ......T..T
Beta          .......... ..G....... T..C...... ...C.G.... .T..C..TA. .........C
Persea        ..C..T.... ..G.C..T.. .......... .......... .C.....TA. ......T..T
Pisum         .....T..G. ..G.C..... T..C...T.. ..C....... .T......A. ......T...
Nicotiana     ..C...G.G. ..G....... .......... ...C.T.... C......TA. A..A......
Papaver       ..C..T.... ..G.C..... .......... .......... .C..C..TA. A.....T..T
Asparagus     ..C..G.... ..G....T.. .......... ..A....... .T.....TA. ...A......
Illicium      ..C..T.... ..G....T.. .......... .....G.... .CT.G..GA. ...A..T..T
Ephedra       G.C..AC.G. T.G.CA.A.. ...C...... ..A..G.... .T.....TA. .........T

Taxus         CCTGTTAGAC CTTCAGTCAT GATGGACACC ACATCAAGGA GTGAGGATGA TCTCACCCAT
Thuja         ..A..G.... .......T.. .......... ..TAGT.... .......C.. ...T..T...
Podocarpus    .....CC... ..G....T.. .........T .......... .......... ..........
Gingko        ..G..C..G. ....T..... .........T .....C.... .......... ......T...
Drimys        .......... ....T..G.. ......T..G T.T..T.... .......... CT.A..T...
Saruma        ..A..G..G. ....T..G.. ......T..A T.T..TC.T. .......... CT.G..T...
Liriodendron  .....G.... ....T..G.. ......T..T T.T..T.... .......C.. CT.G..T...
Ceratophyllum .....G.... ....T..G.. ......T..T T.T..T.... .......... .T.G......
```

```
Beta           ..A..G..G. ....T..G.. ......T... T.T..T.... .......... .T.G......
Persea         ..A..G.... .A..T..G.. ......T... T.C..TC.T. .......... .T.G......
Pisum          .....G.... .......A.. .........A T.C..C..A. .......... .T.A..T...
Nicotiana      ......C... ....T..A.. ......T..T T.G....... .......... .T.A..A...
Papaver        .....G.... ....T..G.. ......T..A T..G.T.... .......... .T.A..T...
Asparagus      .....A.... .A..T..T.. .......... T.TG.T..A. .......... .T.G..T...
Illicium       ..A....... .A..C..T.. ......T..T T.G..T.... .......... .T.A..T...
Ephedra        .....C..G. .A..T..... ....A.T.G. ..T..G.... .......C.. ...T..T...

Taxus          CAGCTTGCTA TGATTATTAG ACACAATGAA ACTCTAAAAA AGCAGGAGAA AACTGGGGCT
Thuja          .......... .......... .......... .A...T.... .......... ......A...
Podocarpus     ........C. .......... G.....C... .....T.... .......... ..........
Gingko         .......... ....C..... ......C... .A.T....G. GA.......G ..A...A...
Drimys         ..A..G..A. .......... .........G .A...T.G.. G........G ..A...A...
Saruma         ...T.A..C. .......... .........G .A...T.G.. G......... ..A...C..C
Liriodendron   ...T.G..A. .......... .........G .A...G.G.. G........G ..A...A...
Ceratophyllum  ..A....... .......CC. G..T...... .A....CGG. GA.......G G.A...T..G
Beta           ..AT.A.... ........C. T..T.....C .A...G.G.. GA..A....G G.A......G
Persea         ...T.G..C. .......... .......... .A...C.G.. .........G ..A...A...
Pisum          .....G..C. ....C..C.. G......... .A...T..G. GA.......G G.AC..TT..
Nicotiana      ..A..G..C. ........C. T.....C..G .A.T.G..G. G......A.G G.A......A
Papaver        ..AT.G..G. .......... .........G .A...G.GG. .........G ..A...T...
Asparagus      .....A..A. ....A..C.. .......... .ACT.G.G.C GT..A....G G.AC..A...
Illicium       .....G..A. .......... .........G .A...T.GG. GA.......G ..A...A...
Ephedra        ..AT.A..A. ....C..... .........G .....G..G. ....A..A.. ..A...A...

Taxus          CCAGCTCATA TTATTTCTGA GTTTGCCCAA TTGCTTCAAT TCCACATTGC TACATATTTT
Thuja          .......... .......... .......... ...T.G.... .......... ..........
Podocarpus     .......... .......A.. .......... .....G.... ....T..... ......C...
Gingko         ........C. ....CA.... A.....A... .....G..G. .......... ..........
Drimys         ..T..C..C. .C..A..A.. A.....T... ..AT.G.... .T........ ..........
Saruma         ..T..C..C. .A.....A.. A.....T... ...T.G.... .T........ C.........
Liriodendron   ..T..C..C. .C..C..A.. A.....G... ..AT.G..G. .T........ ..........
Ceratophyllum  ..T.....C. .C..A..A.. A.....A... .....G..G. .T..T..A.. A..C......
Beta           ..T..A.... .C..C..C.. A........G .....G..G. .T..T..... ......C...
Persea         ..T..G..C. .C..C..A.. A.....T..G ...T.G..G. .T........ ...G.....C
Pisum          ..T..A.... ....A..C.. ......T..G ...T.G..G. .T........ C..G......
Nicotiana      ..G..A..C. .C.....A.. ...C..A..G ..AT.G.... .T..T..A.. ..........
Papaver        ..T..G..C. ......TG.. ...C..A... C.AT.A..G. ....T..... A..T......
Asparagus      ..T.....C. .C...A.A.. .........G ..AT.G.... .T..T..... A..T......
Illicium       ..T..C..C. .C..C..G.. A.....T..G ...T.G..G. .T.....C.. A..T......
Ephedra        .....A..C. ..G..A.G.. ......T..G ...T.G.... .T..TG.... ...G......

Taxus          GACAATGAGC TGCCTGGCCA GCCACGGGCA ACTCAGCGGT CAGGACGACC AATCAAGTCA
Thuja          .......... .......T.. ....A..... ..C...A... .C........ ...A..A...
Podocarpus     .....C.... ....C..... ....A..... ........T. .T.....T.. ...T.....T
Gingko         ........C. .......T.. ....A..... .......... .......G.. T........T
Drimys         ..T....... .T.....T.. ...GA....T ..A..A..T. .T..G..G.. T.....A..T
Saruma         ..T....... .......T.. ....A....T ........A. ....C..... T.....A..T
Liriodendron   .......... .T..G..A.. ...GA....T ..A...A... .G.....G.. T.....A..T
Ceratophyllum  .....C..AT ......?T.. ....A..... ........C. .T........ C.....A..C
Beta           .....C.... .......A.. ....A....T ..G..A..A. ....TA.... ...T..A..C
Persea         ..T....... .......T.. A...A....C ........A. .T........ T.....A..T
Pisum          ..T......T .......ATT ....A....T .....AA.A. .....A.G.. T.....A...
Nicotiana      ........T. ....G..A.. A...A.A..T ..A..AA.A. ....TA.... T..T..A...
Papaver        .......... .......T.. A...A....T ..A..A..T. .T..A..... ......A..T
Asparagus      ..T....... .T.....A.. ....A....T ..A....... .T........ T........T
```

```
Illicium      .......... .T.....A.. ....A..... ..A.....T. ....GA.G.. C.....A..T
Ephedra       ..T....... .T.....A.. .......... .....AA.A. ....T..... C..A.....T

Taxus         ATATGTAGTA GATTGAAAGC CAAAGAGGGT CGCATCAGGG GAAATCTAAT GGGAAAGCGT
Thuja         .....C...C .......... T......... .......... .......G.. ...G......
Podocarpus    .....C.... .......G.. .......... .......... .T...T.... ...G..?...
Gingko        .....C.A.. ..C....... T..G...... ..T.....A. .G..CT.G.. ...T......
Drimys        .......... .GC.T..G.. T.....A... ..G..T..A. ....CT.G.. ...G..A...
Saruma        .......... .GC.C..G.. T.....A..C ..A..T..A. .T...T.G.. ...G..A...
Liriodendron  .....C...C .GC.T..G.. A.....A... ..G..T..A. .C..CT.G.. ..........
Ceratophyllum .......... ..C.C..G.. .........A ..T.....A. .T...T.G.. ..........
Beta          .......CA. .GC....... ......A... ..G..A..A. .C...T.G.. ......A..A
Persea        .....C.... .GC.T..G.. T.....A... ..G..T..A. .G..CT.G.. ...G..A...
Pisum         .....C..C. .GC.T..... G.....A..A ..T..T..A. .C..CT.... ......AA.A
Nicotiana     .......... ....A...T. A.....A... ..G.....A. .C.....G.. ...C..AA.A
Papaver       .....C.... .GC.T..G.. G..G..A... ..A..T.... .T..CT.... ...G......
Asparagus     .....C.... ..C.T..... A.....A... ..G..T..A. .G..CT.G.. ..........
Illicium      .......... .GC.A..G.. T..G..A..C ..T..T.... .T...T.G.. ...G..A...
Ephedra       .....C..C. .......G.. G..G..A..A A.A..T..A. .G...T.G.. ..........

Taxus         GTAGATTTTT CTGCTCGTAC AGTTATTACT CCTGATCCTA ATATCAACAT TGATGAACTA
Thuja         .......... ....A..... .........A .......... ....T..... ..........
Podocarpus    .......... ....A..... T........A ........A. ....T..... ........T.
Gingko        .......... .C..A..... T......... ........A. .......... ....C.G..G
Drimys        ..G....... ....A..G.. T..C..C..A ..G.....C. CA..T..... .........G
Saruma        .......... ....A..A.. ...G..C..C ..A..C.... C...T..... C........G
Liriodendron  ..G....... .A..A..A.. G..C..C..A ..A.....G. CA..T..... .........T
Ceratophyllum ..T....... .A........ ...C..C... ..C..C..A. CC..T..T.. .........T
Beta          ..G..C..C. .G..A..... T..C...... ........A. ....T..T.. C........C
Persea        ..G....... ....A..G.. T..C..C..A ..A.....A. CC..T..... .........G
Pisum         ..T....... .......A.. ...A..C..A ..A.....A. CC.....T.. C...C..T.G
Nicotiana     .....C.... .......G.. .......... ........A. C...T..... ....C..T.G
Papaver       ..T....... .A........ T..C..C..G ..A.....A. CA........ .........G
Asparagus     ..T....... .A..A..A.. ......C..C ........A. CA..T..T.. .........G
Illicium      ..G....... .G..A..G.. T..C..C..A ..A..C.... CA..T..... .......G.G
Ephedra       .....C.... ....A..... ......A..A ..G.....A. ....T..T.. ...CC.G..T

Taxus         GGTGTCCCCT GGAGTATTGC CCTGAACTTG ACATACCCAG AAACTGTCAC TCCATACAAC
Thuja         ..A..T.... .......... G......C.. .......... .G........ ..........
Podocarpus    ..A..A..A. .......... A......... .....T.... .......G.. .........T
Gingko        ..A..G.... .......... .......... ........G. .......G.. A.....T...
Drimys        ..A..A..G. .......... TT.A..T..A .......... .G.....G.. A.....T...
Saruma        .....G..G. .......... T.....T... ........T. .......G.. A.....T...
Liriodendron  ..A..G..A. .......... T..T..T..A .....T.... .......G.. A.....T...
Ceratophyllum ..A..A.... .......... A..C..T... ..T.....T. ....A..G.. ......T...
Beta          .....G..T. .......A.. AT....TC.T .....T.... .G.....G.. ...T..T..T
Persea        ..A..A.... .......... G..A..T..A .....T.... .G.....G.. A........T
Pisum         ..A..G..A. ...C...... T..A..TC.T ........C. .G.....T.. ...T......
Nicotiana     ..A.....G. .......... T.....TC.C ..T..T.... ....A..G.. A.....T...
Papaver       ..A.....T. .......... TT.A..TC.. .....T.... ....A..G.. A..C..T...
Asparagus     .....G..T. ....C..... .T.A..TC.. ..G....... .......G.. ...G..T...
Illicium      ..A..G..A. .......... T..C..T..A .......... .......A.. A..T..T...
Ephedra       ..A..A..T. .......... T.....T..A ..T..T..G. ..GT...G.. ......T...

Taxus         ATAGAAAGAT TAAAGGAGCT TGTTGAATAT GGACCTCATC CTCCCCCAGG GAAAACTGGG
Thuja         .......... .......A.. .......... .......... ....A..... .........T
Podocarpus    .......... .G........ .......... .......... .C..A..G.. T........T
```

```
Gingko        ..T..G.... .......... .......... .......... .C..T..... .........T
Drimys        ..T..G..G. .G......T. G..A...... ..G..C.... ....T..T.. T.....C..T
Saruma        ..C..C.... .G........ ...A...... ..C..A.... .A.....T.. C........T
Liriodendron  ..T.....G. .G.....A.. ...A...... ..G..C.... ....T..T.. T........T
Ceratophyllum ..T.....G. .......A.. A..A...... .......... .C..T..T.. A........T
Beta          ..T.....G. ....A..... G.....C... ..T....... ....T..C.. T.........
Persea        ..T..G..G. .G........ ...G..G... ..G....... .......... T........T
Pisum         ..T....... .G.....A.. ......G... .....C.... ....A..T.. A........T
Nicotiana     ..T.....G. .......... .......... .......... .A..T..T.. T.........
Papaver       ..T.....G. .G........ .......... ..G....... .....GTT.. T.........
Asparagus     ..T...C... .G.....A.. ...A...... ..C....... .......... T..G.....T
Illicium      ..T....... .......... ...C..G... ..G..A.... .C..T..T.. T.....G...
Ephedra       ..C,.G.... .......A.. ...A...... ..T..A.... ....T..... A........T

Taxus         GCCAAATACA TTATCAGGGA GGATGGCCAG AGACTTGATC TCCGTTACTT GAAAAAGAGC
Thuja         .......... .......... ......T... ........C. .......... ..........
Podocarpus    ..T....... .......... .......... .......... .T........ ..........
Gingko        ..A..G.... .......... A......... .......... .T..C..... ..........
Drimys        .....G.... ....A..... A.....G... ..G....... ....G..... ...G..A..T
Saruma        .......... ....A..... A.....A... .......... .T..G...C. ...G..A..T
Liriodendron  .....G.... ....A..... A.....A... ..G....... .T..G..... ......A..T
Ceratophyllum ........T. .A..A..A.. ......T... .......... .T..G..T.. ...G......
Beta          ........T. .C.....A.. T.....T... ..G....... .......TC. ...G.....T
Persea        .....G..T. .C..A..A.. A.....A... ..G....... .TA.G....C .........T
Pisum         ..T..G..T. .C..TC.T.. T.....A..A ..G....... ..A.A..T.. ...G..A..T
Nicotiana     ..A..G..T. .A..T..... T..C.....A ..G....... ....A..... ...G......
Papaver       ..A..G.... ....A..... A.....A... .......... .T.....T.. ......A..T
Asparagus     ..A..G.... .A........ A..C..T... ..G....... .T......C. C.....A...
Illicium      .....G.... .C..A..... A.....T... ..G....... .T..G..... ...G..A..T
Ephedra       ..G.....T. .C.....A.. C.....T... ...T.G.... .G.....T.. .....G....

Taxus         AGTGACCATC ATCTGGAGTT GGGCTATAAG GTGGAGAGAC ACTTGAATGA TGGAGACTTT
Thuja         ..C..T..C. ....A..A.. .......... .......... .......... ..........
Podocarpus    ........C. ..T.A..... A..T...... ..T.....G. .......... ...G.....C
Gingko        ..C.....C. ....A..... .......... ........G. .T........ ..........
Drimys        .....T..C. ........C. A..A..C... ........G. .......... ...T......
Saruma        .....T.... .C......C. C.....C... .....A..G. .......... ...G....A.
Liriodendron  .....T.... ........C. T..T...... ..A.....G. .......... ...G......
Ceratophyllum .....T..C. .C..C..AC. T..A...... ..T...C.T. .......... ...G..T...
Beta          .......... ..T....A.. ...T..C... ..T.....G. .T........ ..........
Persea        .....T.... ........C. T..A...... ..A.....G. .......... ...G..T...
Pisum         ..C.....C. ..T.....C. T..A..C... ......C.T. .T........ ..........
Nicotiana     ..?..T..G. ....T...C. T..A...... ..T.....G. ....A..... ...G......
Papaver       .....T.... ..T.A...C. T..A...... ..A....... .......... ...C......
Asparagus     .....T.... ....C...C. T..G..C... ..A.....G. .T..A..C.. C..G......
Illicium      .....T..C. ..T.....C. T..A..C... ..T...C... .......... ........AC
Ephedra       .....T..C. .......AC. ...T..C... .....A.... .T........ ...T.....C

Taxus         GTTCTTTTTA ATCGACAGCC GAGTCTTCAT AAGATGTCTA TCATGGGGCA TAGAATCAGA
Thuja         .......... .......A.. A......... .......... .......... ..........
Podocarpus    .......... ....G..A.. T.....C... .......... .T.....T.. ..........
Gingko        .......... ....G..A.. C.....C... .......... .......... ........A.
Drimys        ..C....... ....G..A.. T......... ..A....... .......... ........A.
Saruma        .....C..C. ..A....A.. T......... ..A.....C. .......A.. ...G....A.
Liriodendron  ........C. ....G..A.. A........C ..A....... .A........ .C......A.
Ceratophyllum ........C. ..A.G..... T......... ..A....... .T.....T.. ......T.A.
Beta          ..G....... .......A.. C......... ..A.....C. .A.....T.. C.....A.AG
```

```
Persea        .....G.... ....G..A.. T......... .......... .......... ........A.
Pisum         .....C..C. ....T..... T......... .......... .......... CC......A.
Nicotiana     .....C..C. ....G..A.. C.....A..C ..A....... .T........ ...G....AG
Papaver       ..C.....C. .C..T..A.. A........C ..A....... .......... ........AG
Asparagus     ........C. .CA....A.. A......... ..A.....C. .T.....A.. C..G..A.A.
Illicium      ........C. .C..G..A.. A.....C... ..A....... .......C.. ......T.AG
Ephedra       ..C.....C. ....G..A.. T..C...... ..A....... .T........ ........A.

Taxus         ATCATGCCAT ATTCAACATT CCGATTGAAT TTGTCAGTCA CATCTCCGTA TAATGCGGAT
Thuja         ........T. .C.....T.. ......A..C .......... ....A..C.. C.....T...
Podocarpus    .......... .C........ T.....A... .......... .T..A..C.. C.....T...
Gingko        .......... .C........ TA........ ..A..G.... .T..A..A.. C.....T...
Drimys        ..T....... .C.....C.. T..C...... .....C..G. .T.....T.. C.....T...
Saruma        .......... .C.....G.. T..C.....C .....T..G. .T..C..A.. ......T...
Liriodendron  .......... .C.....T.. T..C.....C .....T..A. .T........ C.....T...
Ceratophyllum ..T....... .C.....C.. T.....A... ..A..?..G. .T..A..T.. C.....T...
Beta          ..T.....C. ....C..... .......... C....T..T. ....A..T.. C.....T...
Persea        ..T....... .C.....T.. T..T...... .....T..A. .......A.. C..C..?...
Pisum         ........T. ....T..T.. T..GC.T..C .....T..T. .C..A..A.. C.....T...
Nicotiana     .......... .C.....C.. ...G...... ..A..T..T. .T.....T.. C.....T...
Papaver       ..T.....T. .C.....C.. T..TC....C .....T..A. .T..A..A.. C.....T...
Asparagus     .......... .......T.. T..C..A... ..A.....T. .T..C..A.. C.....T...
Illicium      .......... .C.....C.. T..T...... ........G. .T........ C.....T...
Ephedra       ........T. .C........ .A..C..... ..A.....T. ....A..T.. ......?...

Taxus         TTTGATGGGG A?GAAATGAA ?ATGCATGTT CCACAGTCAT TTGAGACAAG AGCAGAAGTA
Thuja         ..C..C..A. .T........ C......... ..C....... .......... .........T
Podocarpus    ........A. .T........ C......... ..T..A.... .......C.. .........T
Gingko        ........A. .T........ C........G ..C....... .......... G........G
Drimys        ........A. .T........ C........A ..T.....G. ....A..T.. G........C
Saruma        ........A. .T........ C......... ..T....... ....A..T.. G.....G..C
Liriodendron  .......... .T........ C......... ..T..A.... ....A..T.. .........C
Ceratophyllum ........?. .T........ C......... ..C.....?. ....A..... .........?
Beta          ..?.....?. .T........ C......... ..T..?..?. ....?..T?. ?........?
Persea        .....C.... .T........ C......... ..T..A.... ....A..T.. .........C
Pisum         .....C..T. .T........ T......... ..T..A.... ....A..C.. ...T..G..G
Nicotiana     ........T. .C........ C......... ..C....... ....A..C.. .........G
Papaver       .......... .T..G..... T......... ..T....... .......G.. G........T
Asparagus     ........T. .T..G..... C.....C... ..T.....C. ....A..T.. ...T.....T
Illicium      .....C.... .T........ T.....C... ..T.....T. .C.....T.. .........T
Ephedra       ..?.....?. .T........ C......... ..?..?..?. ....?..C.. ?..?...?.?

Taxus         TTGGAACTAA TGATGGTGCC AAAATGCATT GTCTCTCCTC AGTCCAATAG GCCAGTTATG
Thuja         ..A.....C. .......C.. C........C .......... .......... ..........
Podocarpus    ........T. .......C.. C........C .......... ....A..... ..........
Gingko        .....GT... .......... ...G...... ..T.....A. .......C.. ...T......
Drimys        .....GT.G. .......A.. .......... .....A.... ....A...C. T..T......
Saruma        .....G..G. .......C.. C........C ..T..A.... .A..G...C. T..T......
Liriodendron  .....G..G. .......C.. C........C .....A.... .A..A...C. C..T......
Ceratophyllum ..?..?..?. .......?.. ?.....?..? .....G.... .A..A...C. C..T......
Beta          ?.?..G..G. .......A.. G......... .....G.... .A..A...C. C..T......
Persea        ........G. .......... .......... ..G....... .A..A...C. T..T......
Pisum         .....G..G. .......... T......... ..G..A.... ..G.A..C.. ......A...
Nicotiana     C....G..T. .......C.. C..G..T..? ..T..A..C. ..G.A..C.. ...T..A...
Papaver       .....G..C. .......T.. T..G.....C ..G..A.... ....A..... ...T......
Asparagus     .....G..G. .......C.. C..G.....C ..A..A.... ....A...C. ...T......
Illicium      .....GT.G. .......... C..G...... .....A.... .A..T...C. A.....C...
```

```
Ephedra        ?....??.G. .......?.. ?.....?..? ..G..C.... .......... ...T......

Taxus          GGTATTGTGC AAGACACACT TTTGGGTTGT CGAAAGATTA CCAAAAGAGA CACATTTATA
Thuja          .......... ....T..... .......... .......... .A........ ..........
Podocarpus     .......... ....T..T.. CC.A...... .......... .A........ T.....C...
Gingko         ........T. .G.....T.. .........C A.....G.G. .G.....G.. T.........
Drimys         ..C..C..C. .G..T..... C..A..A..C .......... .T........ T..C..C...
Saruma         ........C. .G..T..... C.....C..C ..G.....C. .T........ T..C..C...
Liriodendron   .....C..C. .G..T..... C..A..A..C ..G.....A. .T........ ...C..C..T
Ceratophyllum  ........C. .G..T..... C..A..A..C ..G.....A. .T........ T..C..C..T
Beta           ........C. .G..T..... C..A..A..C ..G.....A. .T........ T..C..C..T
Persea         ........C. .G..T..... C..A..G..C ..G.....C. .T........ T.....C...
Pisum          ..A.....C. ....TT.G.. ...A..A..C A.G..A..C. .......... T........T
Nicotiana      .....C.... .G..T..C.. ...A.....C .....A..C. .T........ T..C......
Papaver        ........C. .G........ C..A..G..C ..G..A..C. .......... T..G......
Asparagus      ..C.....C. .G..T..... C..A..A... ..G.....C. ....G..... ...C.....T
Illicium       ..C.....C. .G.....C.. ...A..A..C ........C. .A...C.T.. ...T..C...
Ephedra        ........C. .G..T..T.. .........C A.......C. .......... .........T

Taxus          GAGAAGGATG TCTTTATGAA TATACTAATG TGGTGGGAGG ATTTTGATGG AAAAATACCT
Thuja          .......... .G........ ...C...... .......... .......... G..G......
Podocarpus     .......... .......... ...CT..... .......... .......... T.........
Gingko         .......... ....C..... ...CT..... .......... .......... C........A
Drimys         .......... .......... C..CT.G... ........A. .......... G...G.....
Saruma         .......... .A........ C...T.G... .......... .C........ G.....T...
Liriodendron   .......... .......... ...CT.G... .......... .......... G.........
Ceratophyllum  .......... .......... ...CT.G... .......... .......... G.........
Beta           .......... .......... ...CT.G... .......... .......... G.........
Persea         .......... ....C..... ...CT.G... .......... .......... ..........
Pisum          .CA....... .T........ ....T.G... ........A. .......... G...G.T...
Nicotiana      ..A....... .......... C..TT..... .......... .......C.. G...G....C
Papaver        .......... ....C..... C..CT.G... .......... .......... ....G.....
Asparagus      .......... .A........ ...TT.G... ........A. .C........ G..GG.G...
Illicium       ..A....... ....C..... C...T.G... .......... .......... G.....T...
Ephedra        .......... .T..C..... ...CT.G... ........A. .C........ G...G.....

Taxus          GCTCCAACTA TTCTAAAGCC AAGGCCGCTC TGGACTGGTA AACAAGTATT CAATCTGATT
Thuja          .......... .......... ......A..T ........C. .......C.. ..........
Podocarpus     T....CG... .......... ...A..A..T ........A. .G........ T........A
Gingko         T.......A. .......... C.....A..T ........C. .......... ......T..C
Drimys         ......G.C. ..T.G..... G..A..T..T .....G..A. .......... T.....T..C
Saruma         ..C...G..G ..T.G..... C.....T..T .....A..A. .......... T.....T..A
Liriodendron   ......G..G ....T..... T..A..A..T ........A. .G.....G.. .........C
Ceratophyllum  .......... ....G..... T..A..T..T .....A..A. .G.....G.. .........C
Beta           .......... ....G..... T..A..T..T .....A..A. .G.....G.. .........C
Persea         ......G.C. ....G..... ......T..T .....G..A. .......... ......C...
Pisum          .....T..G. .AT.G..A.. .GAA..A..G ........G. .......T.. ..........
Nicotiana      ..G..GG.G. ....G..A.. ......A..T .....C..G. .G.....G.. T..C..C..C
Papaver        .....T..A. ..T.G..... T..A..T... ........A. .......C.. ......C...
Asparagus      A.C...G.C. ..T.G..... T..A..T..G .....A..G. .G.....G.. T.....T..C
Illicium       ..A..GG.C. .CT.G..... C..A..T..G .....A..G. .......... T.....T..A
Ephedra        ..A..TG.A. ..T.G..... ......AA.T .....C..C. .G.....T.. T.....T..C

Taxus          ATTCCAAAGC AGATTAATCT TATAAGGCAA TCTGCGTGGC ATGCAGAGTC TGAAACAGGA
Thuja          .......... .A........ .......T.C .....A.... ....T..A.. ..........
Podocarpus     .......... .......... .......T.C ..A..A.... .......... ......C...
Gingko         .......G.. .A..A..... A......T.C .....A.... ..T....... ..........
```

```
Drimys        .......... .......... C.....A... ..C..T.... ..T....AA. ..........
Saruma        .......... ....A..... CC....AACC ..A..A.... .C.....A.. ..........
Liriodendron  .....T.... ....A..C.. C..C..AACC ..G..A.... .CT.G..AG. G.........
Ceratophyllum .....T.... ....A..C.. C.....AACC ..G..A.... .CT.G..AG. G.........
Beta          .....T.... ....A..C.. C.....AACC ..G..A.... .CT.G..AG. G.........
Persea        ..C..G.... ....A..... C..T?.AACC .....A.... .CT....A.. A........T
Pisum         .....T.... ....A..... ...T..AT.T ...AGT.... .CAAT...G. C....GG...
Nicotiana     ..A..G..A. .A..A..... .C..C.TT.T ..C....... ..AAT..T.. A....A....
Papaver       .....G.... ....A..... CT.G..AACC ..A....... ..T.....AA ......T..T
Asparagus     ..A..G.... ....A..C.. .TCC..AGCT ..A..A.... .CT.T..A.. ..........
Illicium      .....C..A. ....A..C.. G......ACC .......... .CT....AAG ......G..G
Ephedra       ........A. .A..A..... CG.....T.C .....C.... ..AAT..AAG ...T.GG...

Taxus         TTTATTACGC CGGGGGATAC AATTGTACGG ATTGAGAAAG GGAGGTTCT TTCGGGCACT
Thuja         ........A. .A........ .TG....... .......... ........A. ...T......
Podocarpus    ...C....AG .T..A..... TG......AT .......... .A........ ...A......
Gingko        .....C.... .A........ CTG...C... .....A.... .A........ ...T......
Drimys        .....C..T. ....A..... CCAA..T... ..A....G.. .....C.... A.GT......
Saruma        .....C..A. .A..A..... TCAA..T... ..A....G.. .A...C.... .G.T......
Liriodendron  .....C..T. ....A..... .CAA..TA.A ..A....G.. .....C.G.. .G.T......
Ceratophyllum .....C..T. .A..A..... .CAA..TA.A ..A....G.. .T...C.G.. .G.T......
Beta          .....C..T. .A..A..... .CAA..TA.A ..A....G.. .T...C.G.. .G.T......
Persea        ........T. .T..A..... .CAA..TA.A ..A..A..G. ....AC.... .G........
Pisum         CC...A..C. .T..A..... G..G..CA.A .....A..G. .....C.... .A.T..T...
Nicotiana     CCC..A..T. .A..A..... TCAA..TA.A ..A....... .....T.G.. AA.T..T...
Papaver       .....A..A. .A..A..... CCAG..TA.A ..A....GG. .....C.... CA.A..T...
Asparagus     .....C..A. .C..A..... GCAA..CA.A ..A....... .A...C.G.. .A.A..A...
Illicium      .....C..T. .A..A..... CCA...TA.. .......... .....C.... ...T.....C
Ephedra       CA.T.G..T. T...A..... TG....CA.A .....A..G. .A..A...A. AA.T..T...

Taxus         CTTTGCAAGA AAACCCTTGG GACATCCTCG GGGAGCCTTA TTCATGTGAT CTGGGAGGAA
Thuja         .......... .......... ...T...... ..A..T..A. .......C.. ..........
Podocarpus    .......... .......... A.....TA.A .....T.... .......... ..........
Gingko        ..G....... ....T..A.. A..G..T..T .....T..G. .......... .........G
Drimys        ..G....... .G........ A.....TA.T ..A..T.... .......T.. ......A..G
Saruma        T.G....... .GG....G.. A..T..TA.C .....T.... .......T.. A.....A..G
Liriodendron  ..C....... .G........ A.....TA.T ..T..T.... .......C.. ......A..G
Ceratophyllum ..C....... .G........ A.....TA.T ..T..T.... .......C.. ......A..G
Beta          ..C....... .G........ A.....TA.T ..T..T.... .......C.. ......A..G
Persea        .......... .......... A.....TA.T ..C..T.... .......T.. ......A..G
Pisum         ........A. .G..A..A.. A..TGG.A.T ..A..T..C. .......CG. T.....A..G
Nicotiana     ........A. .G........ ...G...A.T ..A..T..G. .......T.. A........G
Papaver       ..G....... .GG....... A.....TA.T ..A..T.... .......C.. ......A...
Asparagus     ..G....... .......... A..C..TA.T ..A..T.... .......T.. T...------
Illicium      ..A....... .G..A..C.. ......TA.T ..C..T.... .......C.. .........G
Ephedra       ..C....... ....A..... TG....TAGT ..T..T.... .......... A........G

Taxus         GTTGGTCCAG ATGCTGCTCG TAAGTTTTTA GGTCACACAC AATGGCTTGT AAACTACTGG
Thuja         .......... .......... .......... .......... .G........ ..........
Podocarpus    .....G.... .......A.. ......C... A......... .......G.. T.........
Gingko        ..C..G.... .......... .......... ..G..T.... .G........ C.........
Drimys        ..G..C..T. ....A..C.. G.....CC.G ..A..T.... .G........ T.....T...
Saruma        ..G.....T. .......C.. C......C.C ..A....... .G.....C.. T.........
Liriodendron  .......... ....A..C.. C........G ..C....... .G.....C.. T.....T...
Ceratophyllum .......... .......C.. C..A..C..G ..C....... .G........ T.........
Beta          .......... .......C.. C..A..C..G ..C....... .G........ T.........
Persea        .......... .......... C..A.....G ..A..T..G. .G........ G.....T...
```

```
Pisum         .....C..T. ....A..A.. C..A...C.T ..C..T..T. .G........ ..........
Nicotiana     ..C..G..G. .......... .......C.G ..A..T.... .......G.. T..T..T...
Papaver       ........T. .......... ......C..G ........T. .G........ T.........
Asparagus     ........T. .......... C......C.. ..G..T..C. .G........ G..T..T...
Illicium      ..G....... .......C.. G.....C..G ..A..T..C. .G........ T.........
Ephedra       .......... .......G.. G.....C..G .....T..G. .......... T..T......

Taxus         TTGCTACAAC AGGGTTTCAG TATTGGTATA GGAGACACAA TTGCTGATGC TGCAACCATG
Thuja         .......... .......... .......... .......... .......... ..........
Podocarpus    C......... .......T.. ...G.....T .......... ....C..... ...G..T...
Gingko        C..T....G. .......... .......... .......... .......... ......T...
Drimys        C.TT....GA .T..C..T.. ......A..T ..G....... .......C.. A..T..A...
Saruma        C.TT.....A .T.C...T.. C.....A..T ..G....... ....A..... ......A...
Liriodendron  C.TT....GA .T..A..T.. ......A..T ..G....... ....A..... A.........
Ceratophyllum C.TT.G..GA .T..A..T.. ......A..T ..G....... ....A..... AT....T...
Beta          C.TT.G..GA .T..A..T.. ......A..T ..G....... ....A..... AT....T...
Persea        C.TT.G..GA .T.....T.. ......A..T ..T.....?. ....A..... ......T...
Pisum         C.TT....GA .T.C...T.. C.....A..T ..T..T..C. .......... .T.G......
Nicotiana     C.T..T..GA .T.C...T.. C.....G... .......... .......... .T........
Papaver       C.TT....GA .T.....T.. C.....A..T ........G. .C........ CT.G......
Asparagus     C.TT.....A .T.CC..T.. C........T ..G....... ....A..... A..T..T...
Illicium      C.TT.G..GA .T........ .........T ........G. .......... ......A...
Ephedra       .......... .......... ...G..C..T .....T.... .......... CA....A...

Taxus         GAAGTTATCA ATGAAACAAT TTCAAAAGCA AAACATGAAG TCAAACAACT TATTAAGGCT
Thuja         .....C.... .......G.. ...G...... ..GA...... ....G..... ..........
Podocarpus    ........T. .......... C....G.... ...A...... .......... C..C......
Gingko        ...ACG..T. .C........ .......... ..GA....G. .......... ..........
Drimys        ..CACG..T. GC........ A..T..G... ..GA....T. .G...G.G.. ...C.T....
Saruma        ..GAAG.... ....T..C.. AA....G... ..GA...... .G...G.... ...C......
Liriodendron  ...AAA..T. .......... A..T.....G ..GA...... .G..GG.... ..........
Ceratophyllum ...AAA..T. .......... A..G.....G ..GA...... .G..GG.... ..........
Beta          ...AAA..T. .......... A..G.....G ..GA...... .G..GG.... ..........
Persea        ..GAAA..T. .......G.. A.TG...... ...AG..... .G...G.C.. ..........
Pisum         ..GAC...T. ..C.G..T.. ....C.G..T ..GG..A... .G........ C..CCG..AA
Nicotiana     ..GAAA..T. .......T.. C..G..T... ...AGCA... .....G.... ...C..A...
Papaver       ..GAAA..T. .C..T..... .........G ...A...... .....G.T.. ...C.GAATG
Asparagus     ...AAA..T. .......... C.....G..C ...A....C. .G..GG.T.. .......CTG
Illicium      ...AAA..T. .C........ ...T.....C ..GA...... .....G.... ...C......
Ephedra       ..CACG.... .......... ACA.G.C..T ...AT.A... .GC..G.... G..AG.AAAG

Taxus         GCTCAGGATA AGCAATTAGA GGCAGAACCT GGTCGAACAA TGATGGAGTC TTTTGAAAAC
Thuja         .......... .......G.. ...G...... .......... .......... ..........
Podocarpus    ........A. .......G.. ?......... .....T.... .......... ..........
Gingko        .....A..G. ..GC...G.. .......... ..C..T.... .......... ..........
Drimys        ..C..A..C. ..GCT..... .......... ..G.....T. ...C...A.. A.....G..T
Saruma        ..A..C..A. ....G..G.. ...T...... ..G.....T. .......A.. G.....G...
Liriodendron  ..C..A..C. ....GC.... .......... ..G.....T. .......A.. A.....G...
Ceratophyllum ..C..A..A. ....GC.... .......... ..G.....T. .......A.. A.....G..T
Beta          ..C..A..A. ....GC.... .......... ..G.....T. .......A.. A.....G..T
Persea        ..C..A..G. .ATC...G.. .......... ..G.....T. .......A.. A.........
Pisum         ........A. ..A....G.. ...T...... .....G.... .......T.. A........T
Nicotiana     .....A..G. ....G..... A..T.....A ..GA....T. .......A.. A..C..G...
Papaver       ..C.....G. ..AGT..... AC.T...... ..G.....T. .......A.. G..C..G...
Asparagus     ..C..A..C. .......... .C.T..G... ..CA....T. .......... A.....G..T
Illicium      ..C..A..G. .......G.. .........C ..C.....T. .......... A.....G...
Ephedra       TACAT..CAC .CA.GC.... ACA...G..A ........GC .AT.A..A.. A........T
```

```
Taxus         AGAGTCAATC AGGTGTTAAA TAAGGCCCGT GATGATGCTG GAAGCAGTGC ACAGAAAAGT
Thuja         .......... .......... ......T... ........A. .......... ...A......
Podocarpus    .......... .......G.. C......... ..C.....A. ....T..... C..A......
Gingko        .....G.... .......G.. C.....T... ........G. ....T..... T..ACG...C
Drimys        ..G..G..C. .......G.. ......T... .......... .T..T..... ...A..T..C
Saruma        .....G..C. .....C.G.. C.....T... ........G. .T..T..... ......G..C
Liriodendron  .....G..C. .....C.G.. ......T... .......... .T..T..... ......G...
Ceratophyllum ..G..G..C. .....C.G.. ......T... .......... .G..T..... ......G..C
Beta          ..G..G..C. .....C.G.. ......T... .......... .G..T..... ......G..C
Persea        .....G..C. ....T..G.. .......... .......... .C..T..... ......G...
Pisum         .....T..C. ..ACT..G.. ...A..T... .......... ...AT..... C..A......
Nicotiana     .....G..C. .......G.. ...A..T... ........G. .......... CG....G...
Papaver       .....G.... .......G.. C..A..T... ..C....... ....T..... T.....G...
Asparagus     .....G..C. ....C..G.. ...A..T... .......... .G..T..... ......G...
Illicium      .....T..C. .......G.. C..A..T... .......... .T..T..... C.....G...
Ephedra       CA......C. ....T..... ......T..A ........A. .T.A...... ...A.G....

Taxus         TTATCTGAGA GTAATAATTT GAAGGCTATG GTTACTGCTG GGTCAAAAGG AAGTTTTATT
Thuja         .......... ....C..... .......... .......... ....C..... ......C...
Podocarpus    ..G....... .......... ......A... .?G..G.... .A........ ..........
Gingko        .....A.... .......... ......A... .....A.... .......... ..........
Drimys        ..G..A.... ....C...C. C.....A... .....A..A. .A..C..... .........C
Saruma        ..G....... ....C...C. C......... ..C....... .A..T..... .........C
Liriodendron  .....A..A. ....C...C. .......... ..G..A..A. .A..G..G.. .........C
Ceratophyllum .....A..A. ....C...C. .......... ..G..A..A. .A.....G.. .........C
Beta          .....A..A. ....C...C. .......... ..G..A..A. .A.....G.. .........C
Persea        C....A.... ....C...C. A.....A... ..G..A..G. .A........ ..........
Pisum         C......... .C......C. ...A...... ..C..A.... .T..T..G.. ......C..C
Nicotiana     .....A..A. .C..C...C. T..A...... .......... ....C..G.. ..........
Papaver       ..G....... ........C. C......... .....A..A. .A........ ..........
Asparagus     ........A. ....C..C.. A.....A... ..C....... .A.....G.. ......C..C
Illicium      .....A.... .......... A..A..A... ..C..G..A. ....C..... G.....C..A
Ephedra       .......... .C........ ...A..C... .......... .T........ .........C

Taxus         AACATATCAC AAATGACTGC TTGTGTGGGA CAGCAAAATG TGGAAGGAAA GCGAATTCCC
Thuja         ........T. .......... .......... .......... .A.....G.. .........A
Podocarpus    ........C. .G........ .......... .......... .A.....G.. ...T......
Gingko        .......... .G.....A.. .......... .......... .T........ .........A
Drimys        .....T.... .G........ .......... ..A.....C. ....G..T.. A........T
Saruma        ..T..T.... .G........ .......... ..A..G.... .T..G..C.. A..G..C..G
Liriodendron  .....T.... .G........ .......... ..A..G.... .T..G..T.. A........T
Ceratophyllum .....T.... .G........ .......... ..A..G.... .C..G..T.. A........T
Beta          .....T.... .G........ .......... ..A..G.... .C..G..T.. A........T
Persea        .....C.... .G........ .......... .....G..C. .T..G..G.. A.....C..T
Pisum         ........T. .G........ .........C ..A..G.... .T..G..T.. A........A
Nicotiana     ........T. .......... .........G ..A..G.... .T.....G.. A..T.....T
Papaver       .......... .G........ ......C... .....G.... .A..G..G.. ...G.....T
Asparagus     .......... .G........ .......... ..A..G.... .T..G..C.. A..T..C..T
Illicium      ..T..T..T. .G........ .........C .....G.... .C..G..T.. A..G.....G
Ephedra       ..T....... .G........ .......... .......... .......T.. A........T

Taxus         TATGGATTTG ATGGAAGAAC CCTACCACAT TTTACAAAGG ATGATAAAGG TCCCGAAAGT
Thuja         .......... .......... .......... .......... .C........ ...A......
Podocarpus    .......... .....C.... ...G..T... ..C..C.... ....C..... ...A......
Gingko        .......... ....TC.... T.....C... ..C....... ....C..T.. ...T......
Drimys        .....G..CA T...TC.G.. TT....C... ..C..C..A. ....C..T.. G..T......
```

```
Saruma         .....G..CA T.AACC.... AT....C... ..C...C.A. .......T.. G..A......
Liriodendron   ..C..T..CA T..ATC.G.. A..G...... ..C..C..A. ....CT.T.. G.........
Ceratophyllum  ..C..T..CA T..ATC.G.. G......... ..C..C..A. ....CT.T.. G.........
Beta           ..C..T..CA T..ATC.G.. G......... ..C..C..A. ....CT.T.. G.........
Persea         .....T...A ....TC.G.. AT....C... ..C..T..A. ....CT.T.. G..T...?..
Pisum          .T...G..CA TA.ACC.... A..G..C..C ..C.....A. .....T.T.. G..T......
Nicotiana      .T...G..CA TA.ACC.T.. T.....T... .....C.... .....T.T.. ...A......
Papaver        ..C.....CA T..ATC.... A.....T... .....T.... ....CT.T.. G..T......
Asparagus      .....T..CA TC..TC.... AT....C..C ..C.....A. .....T.T.. A..T......
Illicium       ..C..G...A T..ATC.G.. GT....C... ..C..C..A. ....CT.T.. G..A......
Ephedra        .....T...C .A..CC.G.. A..T...... ...GT..... .......G.. A..T..G...

Taxus          CGAGGCTTTG TGGAGAATTC TTATCTGCGT GGTTTGACTC CGCAGGAGTT CTTTTTCCAT
Thuja          A......... .......... C......... .......... .T........ ......T...
Podocarpus     ..T....... ....A..... .........G .......... .A........ ..........
Gingko         ..T..T.... ....A..C.. .......... ..AC.T.... .C........ ......T..C
Drimys         ..T..A..C. .......C.. A.....C... ..GC.A.... .A.....A.. T..C......
Saruma         ..T....... .......... ......T... ..GC....C. .A..A..A.. ...C..T...
Liriodendron   ..G....... .......C.. A......... ..GC...... .A.....A.. ...C..T..C
Ceratophyllum  ..G....... .......C.. A.....C... ..GC...... .A.....A.. ...C..T..C
Beta           ..G....... .......C.. A.....C... ..GC...... .A.....A.. ...C..T..C
Persea         ..T....... .......C.. A.....A... ..A?....C. .A.....A.. ...C..T...
Pisum          ..T..G.... .......C.. A.....C... ..AC....C. .T..A..... ......T...
Nicotiana      ..T..T.... .......... A.....C..G ..AC....C. .......A.. ...C......
Papaver        ..T..G.... .......... A..CT....G ...C...... .A..A..... ......T..C
Asparagus      ..T....... .C.....C.. C.....C... .......... .A........ ...C..TT..
Illicium       .....A..C. .......C.. C.....T... ...C....C. ....A..A.. T..C..T...
Ephedra        ..T..T.... .A..A..C.. ......T..G .....A.... .T........ ...C..T...

Taxus          GCAATGGGAG GTCGTGAAGG TTTGATAGAT ACTGCAGTGA AAACATCAGA AACAGGCTAC
Thuja          .......... .G........ .......... .......... ....G..... ......A...
Podocarpus     .......... ....A..... .?........ .....T.... .G........ ......T..T
Gingko         .......... ....A..... .........C .....T.... ....C..T.. G.....G...
Drimys         ..T.....T. ..A.G..... .C.......C .....T.... ....T..T.. ...C..T..T
Saruma         ..T....... .AA.G..... ......T... ..A....... .......C.. G..C..G...
Liriodendron   ..C.....T. ..A.A..... .C....T..C .......... .......T.. G..G..G..T
Ceratophyllum  ..T.....T. ..A.G..... .C....T..C .......... .......T.. G..G..G..T
Beta           ..T.....T. ..A.G..... .C....T..C .......... .......T.. G..G..G..T
Persea         ..T.....T. ..A.G..... .C.T..T... .......... ....C..T.. G..T..G...
Pisum          ..C....... ..A.G..... .C.T..T... .......... .G........ ...T..G...
Nicotiana      ..T.....T. ..A.A..... .C.A..T... .....T.... ....T..T.. G..C..G...
Papaver        ..T....... ..A.A..... .C....T... .....T.... ....T..T.. G..G..G..T
Asparagus      .......... ..A.A..... ...A..T... .....T..C. ....C..T.. ...T..A..T
Illicium       ..C.....T. ..A.A..... .C.C..C... .....T.... ....T..T.. ...C..T...
Ephedra        ..T....... .G..A..... ...A...... .....T..C. .G..T..T.. ..........

Taxus          ATACAGCGCC GTCTTGTGAA GGCCATGGAG GACATTATGG TTAAATATGA TGGGACTGTC
Thuja          .......... .......... A..T...... .....C.... ....G..... ..........
Podocarpus     ........G. .A........ ...T...... .......... .G........ ...T......
Gingko         ......A.G. .......A.. ...T...... ..T..C.... ....G..... ......G...
Drimys         ..T..AA.G. .......... ...T...... ..T....... .C..G..... ...T.....T
Saruma         ..C...A.A. .A........ A......... ..T....... .C........ ...T.....T
Liriodendron   ..T...A.G. .......... ...T...... .....C.... .C........ C..T.....T
Ceratophyllum  ..T...A.G. ....C..... ...T.....A .......... .C........ ...T.....T
Beta           ..T...A.G. ....C..... ...T.....A .......... .C........ ...T.....T
Persea         ..C..AA.G. .......... A..T...... .......... .C........ ...T.....G
Pisum          ..T...A.G. .G.....A.. ...T.....A .....C.... .......... ......A..T
```

```
Nicotiana     ..C...A.G. .AT.A..... .........A ..T..C.... .C..G..... ...T.....T
Papaver       ..C...A.G. .......... ...T...... .......... .C........ ...C......
Asparagus     ..T...A.A. .......... ...T...... ..T..C.... .C........ ...C..C..T
Illicium      ..C..AA.G. .C..C..A.. ...T.....A .....C.... .......... ...T..C...
Ephedra       ..T...A.A. ....G..C.. A......... .....C.... .......... ..CA..A..T

Taxus         AGGAATTCGC TTGGCGACGT CATCCAGTTT CTGTATGGGG AGGATGGTAT GGATGCTGTT
Thuja         ........A. .......T.. ...T.....C ........C. .......... ..........
Podocarpus    ........A. ....T..T.. ...T..A... ........T. .A........ ..........
Gingko        .......... .A.....T.. T..T..A... ..C.....T. .A........ ..........
Drimys        .....C..AT .A..T..T.. T..T...... ........A. .A........ ....T.....
Saruma        .....C..TT .G..T..T.. T..T...... T.......A. .A........ ..........
Liriodendron  .....C..AT .A..G..T.. G..T...... T.......A. .......... ..........
Ceratophyllum ........AT .A..G..T.. G..T...... T.......A. .......... ..........
Beta          ........AT .A..G..T.. G..T...... T.......A. .......... ..........
Persea        .....C..T. ....G..... .......... ..T.....A. .A.....C.. ....T.....
Pisum         ........TT .G..A..T.. ...A..A... ..C.....C. .A........ ..........
Nicotiana     ........AT .G..T..T.. T......... .......... .A.....C.. ..........
Papaver       C.......TT .A..A..T.. ...T..A... ..T.....A. .......... ..........
Asparagus     C.......A. .G..A..T.. ...T.....C ..T.....A. .A.....G.. ..........
Illicium      C.......AT .A..G..T.. T.....A... .....C..A. .A........ ..........
Ephedra       .....C..C. ....T..T.. A..A...... ..A..C..A. .A.....C.. .........A

Taxus         TGGATTGAGT CACAGAAGCT TGATTCTTTG AAGATGAAAA AGAAAGAGTT TGAAAATGTT
Thuja         .......... ....A..A.. .......... .......... GA........ ..........
Podocarpus    .......... .G.....A.. .......C.A .......... G......... ..........
Gingko        .......... .T........ G.....A... .......... GA........ ..........
Drimys        ........A. .......AT. G..C..C... ..A....... .AGC...A.. ...C......
Saruma        ........A. .......... G.....C..A ........G. .AGC...... ...C..A...
Liriodendron  ........A. ........T. G......C.. ..A.....G. ..CGT..... C..T......
Ceratophyllum ........A. ........T. G......C.. ..A....... ..CGT..... C..T......
Beta          ........A. ........T. G......C.. ..A....... ..CGT..... C..T......
Persea        ........A. .G......T. A.....C... ..A.....G. .A.G...A.. CCTG......
Pisum         .....A..AA .......... G.....AC.. ..A....... .A.C...T.. ...T.GG.CA
Nicotiana     .........A .G........ G......... ..AGCA..G. .ATC.ACA.. C..TG.CC.G
Papaver       .....C.... .T........ G.....C... ..A.....G. .ACCT..... ...T..G...
Asparagus     .......... ....A...T. G......C.. .......... .......... CA.T......
Illicium      .....C.... .G........ A..C..A..A ..A....... .......... C.........
Ephedra       C....C.... .......A.. A...CA.C.A .......... ....G..A.. ...TG.AAC.

Taxus         TATAGATATG AGATTGATCA AGAAAATTGG AATCCTAGCT ACATGTTACC GGAACATGTG
Thuja         .......... .A..C..... .......... .......... .......... A........A
Podocarpus    ....A..... ..G....... .......... .......T.. .......... A........A
Gingko        ....A...C. ........A. T.....C... ......G.T. ........G. A..T.....T
Drimys        .TCC.G.... .......CA. T..T..C... .....A..T. .T...G.... T..G..C..T
Saruma        .T.C.G..C. ..T.....G. T.CC...... .....G.... .......... T.........
Liriodendron  .TC..G.... .AT....CA. T.CC...... .....A..T. .......... T..G.....T
Ceratophyllum .TC..G.... .AT....CA. T.CC...... .....A..T. .......... T..G.....T
Beta          .TC..G.... .AT....CA. T.CC...... .....A..T. .......... T..G.....T
Persea        .....G.... ....C..CA. T..G..C... .....GG.T. .......... T..G.....T
Pisum         .T...G.... .AT.....G. T.....C... ..G....CT. .....C.TGA A..G.C...A
Nicotiana     ...GC..... ....C...G. TCCT..C... ..C..A..T. .......G.. A...GC.A.C
Papaver       .....G.T.. ..T.....G. T..C...... .C...A.AT. .....CA.T. A..TT....T
Asparagus     .....G..C. .A......G. TA.TC.A... G.C..A..T. .....C.... A..G.....T
Illicium      .TCC.G.... ....A..CA. T..T..C... ..C..G..T. .......... T..G.....C
Ephedra       .T..A.C... ..C...G... .......... C......CA. .......... A...ATG.CA
```

```
Taxus          GATGATCTGA AAACCATTAG AGAATTCCAG AATGTGTTTG ATGCAGAATT GCAGAAACTA
Thuja          .......... .......C.. G..G.....A .....T.... .......... .........?
Podocarpus     ..A.....A. .......C.. G..G..T... .....T.... ........G. C........G
Gingko         ..A...T... .......CC. ...G....GC .....A.... ........G. .........G
Drimys         ..G...T... .......CC. G.....T.G. .......... ....C...G. .........G
Saruma         ..G....... .......... G.....T.G. .G........ ....T...G. T..A...T.G
Liriodendron   ..G...T... .......CC. G..G....GC .....T..C. .C..T...G. T........G
Ceratophyllum  ..G...T... .......CC. G..G....GC .....T.... ....T...G. T........G
Beta           ..G...T... .......CC. G..G....GC .....T.... ....T...G. T........G
Persea         ..G...T... ....T..... G..G....G. .....T.... ....T...G. T.........
Pisum          ..A..CT.A. ....A..... ...G..T.GT .......... .G..G...G. T..A..G..T
Nicotiana      ..A.....A. ....T...C. G..GA.T.GT .G........ .......GG. T........T
Papaver        ..A....... ...AT..... ...G....GT C......... .C......G. TA.....T.G
Asparagus      ..G...T... ....A...CC T......AG. .......... ....C..GG. C......T..
Illicium       A.A...T... ....A...C. G.......GC ..C..A..C. ........G. T........C
Ephedra        ......A.C. ...A...CTC T..G...... G....TC... T...T....A CA....G..G

Taxus          GAGTCAGATA GACGGCAGTT AGGCACAGAG ATTGCTACTA CTGGCGATAA CTCATGGCCT
Thuja          ........?. .......... .......... ......C.?. G...T..?.. ...?......
Podocarpus     .......... ........C. ...T..G... .....AC... G...T..... ...G..T...
Gingko         ..AG.T.... .G......C. .........A ......C... G...A..... ...C......
Drimys         ..AG.C...C ..TAC...C. T..TT..... ........C. .A..T..... ...C......
Saruma         ..AG....CC .GTAC...C. T..A..T..A ...A...... .C..G..... T........G
Liriodendron   ..AG.....C ..TAC...C. T..T..G... .......... .......... ..........
Ceratophyllum  ..AG.....C ..TAC..AC. T..T...... .......... .A..T..... ..........
Beta           ..AG.....C ..TAC..AC. T..T...... .......... .A..T..... ..........
Persea         ..AG.....C ..TAC...C. T..T..C... .......... .G..G..... T..T......
Pisum          ..TG.C.... .GTTT..AC. T.CA..T... ..A..C.... ....T..C.G T..TCT....
Nicotiana      C.AG.T.... ..AAA...C. T..G...... ......GTA. .A..T..C.. T..T.....G
Papaver        .CAGAT...C .G.TC..AC. T..A..T... ..CA....C. .C..T..C.. T..T.....C
Asparagus      ..AG.T..CC ..TTT...C. T..G..T..C .....C..C. ....T..... .A........
Illicium       ..AG.G.... .GTTC..AC. T.....G... .......... .C..G..... ..........
Ephedra        C..CTT.... .G..T...C. T..T.TG..A ..A...C... ....A..G.. .........A

Taxus          ATGCCTGTAA ATCTGAAGAG GCTCATCTGG AATGCCCAGA AGACATTCAA AATTGATCCA
Thuja          .......... .......?.. .......... .......... ....?..... ..........
Podocarpus     .......... .C........ ...A...... ........A. .A........ G........T
Gingko         .......... .......... ...A..A... .......... .......... ......CTTG
Drimys         .....C..C. ....C..... A..T...... .....T..A. .......T.. GG.....TTG
Saruma         .....A..G. ....C..... A..T...... .......... ....C..T.. GG.....TTG
Liriodendron   ........C. .C..C..... ...T...... .....T..A. .......T.. GG....CTTG
Ceratophyllum  ........T. .C..C..... ...T...... .....T..A. .......T.. GG....CTTG
Beta           ........T. .C..C..... ...T...... .....T..A. .......T.. GG....CTTG
Persea         ........T. .C..T..... A..T...... .......... ....G..T.. GG.....TTG
Pisum          C.A.....C. .C..T..... ...T...... .....T.... ....C..T.. GG....CTTC
Nicotiana      C.T.....T. .CA.TC.... ...TG...T. ........A. .A.....T.. G..A...TTT
Papaver        ........G. .CT.A.G... ......T... ..C..T..A. ....C..T.. GG....C..G
Asparagus      ........G. .C..C..... A......... .....A.... .......... .G.C......
Illicium       ........G. .C........ ...T..A... .....G.... .......T.. GG.G..CTTG
Ephedra        T....G..G. ....T..A.. A.....T... .....A.... .A.T...T.. GG.C......

Taxus          AGGAAACGAT CAGATATGCA TCCTATGGAG ATTGTAGAAG CAATTGACAA ACTTCAGGAA
Thuja          .......... .......... .......... .......... .......... G..?......
Podocarpus     .......C.. .......... ...A...... .......... .........G ..........
Gingko         .......CG. .G........ ...C...... .....G.... .TG....... G.........
Drimys         ..A.GG.C.. .T.GC..... ...A...... .....T.... .CG....T.. G.....A...
Saruma         ..A.GG.C.. .T..C..... ...A...... .....T.... ..G....T.. G.........
```

```
Liriodendron  ..A.GG.C.. .T..C..... C......... .....T.... ..G....... G.....A...
Ceratophyllum ..A.GG.C.. .T..C..... C......... .....T.... ..G....... G.........
Beta          ..A.GG.C.. .T..C..... C......... .....T.... ..G....... G.........
Persea        ..A.GG.C.. .T........ ...A...... .....C.... .TG...T... G.....A...
Pisum         C.A.G..CT. .T........ ...A.....A ..A..G.... .T....T... G..C......
Nicotiana     C.A.GG..G. .T........ .......... .....G.... .TG...T... GT.A.....G
Papaver       ..A.GG.CT. .T........ C..C...... .....G.... .T....T... ..........
Asparagus     ..A.G..C.. .T........ C......... ..A....... .TG...T... G.........
Illicium      ....G..CT. .T..C..... C..A...... .....T.... .TG...T... G.....A..G
Ephedra       ......AC.. .......... A..CT..... .....C..CA .......... G.....A..G

Taxus         AGGCTAAAGG TGGTTCCTGG GGACGACCTA ATGAGTATTG AGGCTCAGAA GAATGCCACT
Thuja         .......... .?.....?.. T..?..T.?C ..C....... ....C..... ......?...
Podocarpus    .....T.... .T........ A........T ..A....... .......A.. ..........
Gingko        .....T.... .T........ T..T.....T .......... .......... A.....T...
Drimys        ...T....A. .T........ T..T..T.C. ..T..C..G. .......... A........C
Saruma        ..AT...... .T..C..... T..G..TT.. .....C..G. .A........ A........C
Liriodendron  ...T.G..A. .C....A... C.....TT.T C.C..C..G. .A........ A..C..A..C
Ceratophyllum ...T.G..A. .C....A... T..T..TT.T C....C..G. .A........ A.....A..C
Beta          ...T.G..A. .C....A... T..T..TT.T C....C..G. .A........ A.....A..C
Persea        ..A..?..A. .T........ T..T..TT.T T.......G. .......A.. ..........
Pisum         C.A..T.GA. .A....C... ...T.....T T.A...CAA. .A........ A.........
Nicotiana     .....T.... .T..C..... C..T..TTAC C.A..C..G. .A........ ......T...
Papaver       .....G..A. .T........ T.....TTC. ........G. .......... ......T..C
Asparagus     .A...G.GA. .T..C..A.. T..T..TT.T C.......G. .......... A.........
Illicium      C.......A. .T........ T..T..TT.T C.......G. .A........ A.........
Ephedra       ..A....... .T........ T..T..T.CT C.T..CT... .......... ......A..G

Taxus         CTCTTTTTCA ATATTTTATT GCGCAGTACA TTTGCGAGTA AACGTGTACT GAAGGAATAC
Thuja         ..T..C.... .....C..C. .......... .....A.... .......... ?.....G...
Podocarpus    ..A.....T. .....C..C. T..G...... .....T.... .......G.. A.....G..T
Gingko        ..T....... .C..AC..C. C......... .....C..C. .......G.. ..........
Drimys        ..A.....T. .C..CC.TC. T..A...... .....C..C. .G.....GT. ..........
Saruma        ..T..C..T. .C...C.CC. C..G..C... ...T.A..C. .......GT. ..........
Liriodendron  ..T..C.... .C...C.GC. T..T..C... .....C.... .G.....GT. ...A..G...
Ceratophyllum ..T..C.... .C...C.GC. T..T.CC... .....C.... .G.....GT. ...A..G...
Beta          ..T..C.... .C...C.GC. T..T.CC... .....C.... .G.....GT. ...A..G...
Persea        ..T..C.... .C...C.GC. T..T..C... .....C.... .G.....TT. ...A......
Pisum         ...C.G.... .......GC. T..T..C..T .....C.... ..A.G..C.. .G....G...
Nicotiana     ..T..C.... ....C..GC. C..T...G.C ..G..T.... ..A.G..T.. ......G..T
Papaver       ..G..C.... .C.....GC. ...T..C..T .......... .GA.G..GT. ...A...C..
Asparagus     T.G..G.... .C..A...C. C..G...... .....T.... .GA.A..... ..........
Illicium      .....C.... .C...C.TC. C..A..C... .......... .......G.. ...A.....T
Ephedra       ..G..C..T. .....C.TC. A......GT. .....T.... .G..A..... ...A..G..T

Taxus         CATCTTACAA AGGAAGCATT CGAATGGGTA GTTGGTGAGA TTGAATCCCG TTTCCTTCAG
Thuja         ..C....... .......... T......... .......... .?.....T.. ..........
Podocarpus    ........T. .......... T..G...... ........A. .......T.. ......C...
Gingko        AGA..A..T. GA.....T.. T........G A......... ....G..T.. ......G...
Drimys        .GG..A..TC GT........ .........T A......... .A........ ...TT.A..A
Saruma        AGG..C..TC G......T.. T..G.....T A......... .......T.. G..TT.G..A
Liriodendron  AGG..A..TC GT........ T......... A.C....... .......T.. A...T.A..A
Ceratophyllum AGG..A..TC GT........ T......... A.C....... .A.....T.. A...T.A..A
Beta          AGG..A..TC GT........ T......... A.C....... .A.....T.. A...T.A..A
Persea        AGG.....TC GC........ T........T A.A....... .A.....T.. C..TT.A...
Pisum         AGG...T.TC GT..G..... .......... ..A..G..A. .A.....AA. G.....G..A
Nicotiana     AGA....GTC G...G..T.. T..C...... A......... .A..G..T.. C...?.G..A
```

```
Papaver        ..C..G...C GT........ T..T.....T A.......AG ....G..T.. ...TT.A...
Asparagus      .GA.....CC GT........ T........T A.......A. .A.....T.. C..TT.G...
Illicium       AGG..C..TC GT.....C.. T........T ..A....... .A.....T.. C...T.G..A
Ephedra        ..C....... G...G..... T..T...... A........G .....AAT.. G..T......

Taxus          TCTCTAGTAG CCCCAGGAGA GATGATTGGA TGTGTTGCAG CACAGTCTAT T
Thuja          ...?...... .?..G..G.. .......... .....?.... .......... .
Podocarpus     ..CT.G.... .......G.. .........C .......... .G........ .
Gingko         ...T....G. .A.....T.. .........T .......... .......C.. .
Drimys         ..AT.G.... .A.....T.. A........C .....G..T. .......C.. A
Saruma         ..AT...... .A.....T.. A........T .....G..T. .T..A..C.. A
Liriodendron   ..AT.G.... .T..G..T.. A........C .....G..T. .C..A..C.. A
Ceratophyllum  ..AT.G..T. .A.....T.. A........C .....G..T. .C..A..C.. A
Beta           ..AT.G..T. .A.....T.. A........C .....G..T. .C..A..C.. A
Persea         ..AT.G.... .A..T..T.. A........T ..?..G..T. .C.....C.. A
Pisum          ..A..T..G. ..T.T..G.. A........T .....G..T. ....A..A.. C
Nicotiana      ..C..T..G. .A........ .........C .....C..T. .......A.. .
Papaver        ..AT....T. .T..C..T.. A........T .......... .......... .
Asparagus      .....G.... .A..G..T.. A........C .....A..T. .T........ .
Illicium       .....T..C. .A.....G.. .......... ..C.....T. ....A..... A
Ephedra        .....T..T. ....T..T.. .........T .......... .C.....A.. A
```